

Estimación de parámetros de modelos a priori para segmentación contextual de imágenes.

por Javier Alejandro Gimenez Romero

Presentado ante la Facultad de Matemática, Astronomía y Física como parte de los requerimientos para la obtención del grado de Doctor en Matemática de la

UNIVERSIDAD NACIONAL DE CÓRDOBA

Febrero de 2014

©FAMAF-UNC 2014

Directora: Dra. Ana Georgina Flesia

Resumen

En esta tesis se trabaja en el problema de la estimación del parámetro de una familia exponencial de distribuciones de Gibbs, y su relación con el proceso de segmentación contextual de imágenes vía el algoritmo Iterated Conditional Modes (ICM), trabajando bajo el supuesto que el mapa de clases de la imagen sigue un modelo de Potts isotrópico, y que cada clase emite datos radiométricos Gaussianos multivariados.

En una primera etapa se estudió la consistencia asintótica del estimador de Pseudo-Máxima Verosimilitud (PMV), logrando una prueba más general para la clase de estimadores de PMV correspondiente a densidades de Gibbs con especificaciones no invariantes por traslaciones, que cumplen propiedades específicas detalladas en la tesis.

En una segunda etapa, dentro del mencionado contexto de segmentación, se define un nuevo estimador del parámetro de suavidad del modelo de Potts isotrópico, en el que no solo se tiene en cuenta la información del mapa de clases, sino también la verosimilitud de la información radiométrica proveniente de la imagen original. Este nuevo estimador es el estimador de PMV correspondiente al modelo no invariante a posteriori del mapa de clases, y su consistencia se probó bajo determinadas condiciones sobre el modelo de emisión.

En una tercera etapa, se estudió mediante simulación el desempeño del estimador a muestra finita y bajo condiciones de contaminación usuales en la práctica.

Abstract

In this thesis we study the problem of parameter estimation of an exponential family of Gibbs distributions, and its relation to the process of image contextual segmentation via Iterated Conditional Modes (ICM), under the assumption that the image map class is an isotropic Potts model realization, and each class issues multivariate Gaussian radiometric data.

In a first step, the asymptotic consistency of the estimator of Pseudo-Maximum Likelihood (PML) was studied, obtaining a more general outcome for the class of PML estimators corresponding to Gibbs densities with non-invariant specifications, under specific assumptions detailed in the thesis.

In a second step, within the aforementioned context of segmentation, a new smoothness parameter estimator of the isotropic Potts model was defined. This estimator not only takes into account class map information, but also the radiometric observation likelihood. This new estimator is the PML estimator for the non-invariant posterior model of class map, and its consistency is proved under certain conditions on the emission model.

In a third step, the finite sample performance of our estimator was studied by simulation, and its sensibility assessed under usual conditions seen on the practice.

Palabras claves: modelo de Potts, medidas de Gibbs, segmentación, imágenes, estimación, pseudo-máxima verosimilitud, consistencia.

2010 Mathematics subject Classification: 62M40, 62M05, 60J05.

Agradecimientos

Deseo agradecer a la Universidad Nacional de Córdoba, y en particular, a la comunidad de la Facultad de Matemática, Astronomía y Física, por haberme albergado durante mi periodo doctoral, y por haberme brindado los recursos más que suficientes para trabajar.

Deseo agradecer a CONICET por brindarme la oportunidad de trabajar de lo que amé toda mi vida.

Deseo agradecer a mi directora, Dra. Ana Georgina Flesia por haberme guiado en este difícil camino de la iniciación en la investigación, y por haberme apoyado en todo momento.

Deseo agradecer al Dr. Jorge Adrover porque en él encontré más que un colega, un amigo con quien conversar desinteresadamente de la carrera y de la vida.

Deseo agradecer al Dr. Oscar Bustos, y a través de él, a todo el grupo de Probabilidad y Estadística de la Facultad, por las innumerables charlas de pasillo que me hicieron crecer y aprender.

Pero sobre todo, deseo agradecer a mi familia, y en especial a mi gran amor y amiga Marisol, por el apoyo incondicional, por las palabras justas y por entender que me iba de mi San Juan querido en busca de mi sueño de niño.

Índice general

1. Introducción	7
2. Procesamiento estadístico de imágenes	11
2.1. Introducción	11
2.2. Etiquetado	13
2.3. Funciones de optimización	14
2.4. Modelos	15
3. Medidas de Gibbs	21
3.1. Introducción	21
3.2. Definiciones.	23
3.3. Esperanza Condicional.	28
3.4. Invarianza por Traslaciones.	29
3.5. Modelo de Potts	30
3.6. Ley de los Grandes Números.	33
3.7. Estimación	36
3.8. Consistencia.	37
3.8.1. Modelo Invariante Asociado.	37
3.8.2. Caso Invariante por Traslaciones.	40
3.8.3. Caso General.	50
3.9. Simulación	53
3.9.1. Gibbs Sampler (GS).	54
3.9.2. Potts Sampler (PS).	55
3.9.3. Swendsen-Wang (SW).	55
4. Estimación en el Modelo de Potts	61
4.1. Introducción	61
4.2. Pseudo-Máxima Verosimilitud en el Modelo de Potts	63
4.2.1. A priori	64
4.2.2. A posteriori	65
4.3. Consistencia	66
4.4. Propiedades	67
4.4.1. Gráficos	68
4.4.2. Derivadas	69
4.4.3. Raíces	73
4.5. ECM, Sesgo y Varianza bajo simulación	74

5. Segmentaciones	81
5.1. Criterio MAP: Máximo a Posteriori	82
5.2. Simulated Annealing (SA)	83
5.3. Iterated Conditional Modes (ICM)	84
5.4. ICM y el modelo de Potts.	85
5.5. Sensibilidad	88
5.5.1. Sensibilidad de los Estimadores.	89
5.5.2. Sensibilidad de las Segmentaciones	96
6. Apéndice	99
6.1. Demostraciones de resultados del Capítulo 3	99
6.2. Especificaciones con simetrías.	105
6.3. Consistencia de $\hat{\beta}_{\text{post}}$	107
6.4. Independencia entre h_s y H_s en el modelo de Potts	109

Capítulo 1

Introducción

Las segmentaciones contextuales de imágenes digitales y la estimación eficiente en modelos Markovianos ha tenido un gran auge en la comunidad científica por su amplio campo de aplicaciones. Ante la dificultad y costo computacional de los estimadores clásicos de los parámetros de los modelos Markovianos, Besag [1] introdujo la estimación por Pseudo-Máxima Verosimilitud (PMV), la cual tuvo una gran repercusión por su sencillez y por sus cualidades teóricas asintóticas y a muestra finita. Este trabajo hizo estimables en tiempo finito una gran diversidad de modelos Markovianos, entre ellos, el modelo de Potts o modelo de Potts-Strauss [2], que es una generalización a más de dos estados del clásico modelo de Ising. Las propiedades teóricas de los estimadores de PMV han sido estudiadas con diversas notaciones y desde diversas perspectivas. Janzura en su trabajo [3] muestra un estudio completo de tales propiedades bajo la hipótesis de invarianza por traslaciones de las especificaciones o potenciales que definen el modelo a estimar. En cuanto a la consistencia fuerte del estimador de PMV sobre modelos exponenciales con potenciales invariantes, Geman y Graffigne [4] y Comets [5] desarrollaron los trabajos más sobresalientes. Estos modelos exponenciales son importantes porque contienen a los Automodelos de Besag [1] y todos los modelos propuestos por Geman y Geman [6] en reconstrucción de imágenes. Posteriormente, Guyon en sus libros [7] y [8] estudió la consistencia débil y fuerte para la clase de estimadores de Mínimo Contraste, que contiene a los ya clásicos estimadores de Máxima Verosimilitud y de Pseudo-Máxima Verosimilitud, sin pedir invarianza de las especificaciones. Sin embargo, la generalidad con la que se trabajó produjo teoremas cuyas hipótesis son dificultosas de verificar, sobre todo por modelos generados por especificaciones no invariantes.

En una primera etapa de la tesis se mostrará que si se debilita la condición de invarianza en modelos markovianos exponenciales, y se incorporan determinadas hipótesis que hagan al modelo no invariante equivalente microscópicamente a un modelo invariante asociado, entonces se puede probar aún la consistencia fuerte del estimador de PMV.

Geman y Geman en [6] le dieron un gran énfasis a la aplicación de modelos Markovianos en la comunidad de procesamiento de imágenes, produciendo un antes y un después en el área por la versatilidad de los mismos. Desde allí y hasta la fecha se han desarrollado diversos modelos de segmentación, detección de bordes, análisis de textura, comparación de imágenes, etc., basados en tales modelos, de los cuales sólo nos referiremos a segmentación o clasificación de imágenes. La segmentación de imágenes y la estimación del modelo Markoviano supuesto sobre el mapa de clases son dos procedimientos que se retroalimentan. Por ejemplo, se han diseñado algoritmos específicos de Markov chain Monte Carlo (MCMC) para estimar los parámetros de un campo Markoviano en los trabajos de Descombes et al. [9] y de Murray et al. [10], y más recientemente en los trabajos de Risser et al. [11] y [12]. Bajo el supuesto de que el mapa de clases ajusta a un modelo de Potts, en un trabajo reciente, Pereyra et al. [13] propone un procedimiento que itera entre un ajuste del modelo y una reclasificación de la imagen, evitando el costoso cálculo de la función de partición. Para modelos en

general, Besag [14] desarrolló un segmentador iterativo de imágenes denominado Iterated Conditional Modes (ICM), que permitió optimizar los tiempos de convergencia que presentaba el algoritmo Simulated Annealing (SA) (Geman y Geman [6]). El clasificador ICM requiere una estimación del modelo para poder funcionar, y es aquí donde la estimación a bajo costo que propone la PMV juega un rol importante. Bajo el supuesto de que el mapa de clases de un conjunto de datos radiométricos es a priori modelado por un modelo de Potts, Frery [15] particularizó la estimación por PMV a tal modelo, reduciendo la cantidad de términos que posee la ecuación cuya solución es la estimación buscada, y mostró simulaciones en las que segmentó imágenes reales por medio de ICM mediante un procedimiento que itera entre estimar y segmentar. Posteriormente, Levada [16] optimizó la expresión de la mencionada ecuación, permitiendo considerar vecindades de mayor rango en el proceso de estimación. Ambos estimadores son equivalentes y necesitan del mapa de clases de la imagen como entrada para la estimación. En la práctica, tal mapa no se conoce y debe ser estimado en base a la evidencia radiométrica de la imagen, por medio de una segmentación de la misma. De este modo, la estimación también requiere de un procedimiento de segmentación.

Estimar mirando una segmentación de los datos radiométricos, y no incorporar tales datos como entrada directa en la segmentación, produce una reducción de información que debe analizarse. Este efecto está presente en los estimadores a priori desarrollados en el entorno Potts, debido a que estiman suponiendo que la segmentación de la imagen es una realización del modelo a priori del mapa de clases. Sin embargo, dados los datos radiométricos, la distribución a posteriori de los posibles mapas de clases es diferente a la distribución a priori, y contiene explícitamente la verosimilitud de los datos en su expresión, incorporándolos directamente al proceso de estimación. En una segunda etapa de la tesis, propondremos un nuevo estimador de PMV del modelo de Potts basado en el modelo a posteriori del mapa de clases. Este estimador, a diferencia del estimador basado en el modelo a priori, trabaja sobre modelos con especificaciones no invariantes por traslaciones. Luego, su consistencia no puede ser probada en base a las pruebas de consistencia existentes. Sin embargo, veremos que bajo determinadas condiciones, el estimador propuesto mantiene la propiedad de consistencia.

ICM es un procedimiento iterativo que homogeneiza una segmentación inicial hasta converger y que requiere una estimación del modelo para poder funcionar. En el contexto Potts, Frery et al. [15] propusieron iniciar ICM con una segmentación por MV y reestimar el modelo por PMV iteración tras iteración logrando segmentaciones muy satisfactorias. En una tercera etapa de la tesis inspeccionamos bajo simulación la interrelación existente entre los procesos de segmentación y estimación. Se observaron subestimaciones del modelo cuando se utiliza la segmentación no contextual que ofrece MV, lo cual muestra la necesidad de reestimar el modelo, para de este modo ajustar parámetros que permitan obtener segmentaciones menos ruidosas. Por otro lado, se estimó el modelo utilizando el mapa de clases de convergencia dado por ICM, y se observaron sobrestimaciones del modelo. Como resultado final se observó que la incorporación de la verosimilitud de la información radiométrica al proceso de estimación produce excelentes resultados bajo modelo, pero que al trabajar con segmentaciones en lugar de realizaciones del modelo, las estimaciones son más sensibles al ajuste de la imagen al modelo.

La tesis se organiza del siguiente modo: en el Capítulo 2 se explica que es una imagen desde el punto de vista matemático, y que se entiende como mapa de clases y datos radiométricos de la misma. Además, se define lo que se entiende por clasificación o segmentación de imágenes, remarcando la necesidad de incorporar información contextual de un modo tratable por medio de suposiciones Markovianas que permitan modelar las leyes subyacentes de los mapas de clases por medio de medidas de Gibbs. En el Capítulo 3 se introduce el tecnicismo matemático necesario y suficiente para lograr el formalismo que requiere el trabajo. Se define el estimador de PMV y se prueba la consistencia del mismo para modelos particulares con especificaciones no invariantes. Como herramienta para tal prueba, se prueba una adaptación de una versión clásica de la ley de los grandes números para variables independientes no idénticamente distribuidas al contexto teórico que requiere este problema. Este resultado permite afrontar las diversas pruebas sin necesidad de suponer la ergodicidad de

la medida de Gibbs subyacente. Se finaliza tratando el tema de la simulación de los modelos, mostrando simulaciones del modelo de Potts por medio de Swendsen-Wang y la correspondiente particularización de Gibbs Sampler. En el Capítulo 4 se estudia la estimación en el modelo de Potts, la particularización del estimador de PMV al contexto Potts, se propone un nuevo estimador de PMV basado en el modelo a posteriori del mapa de clases y se prueba la consistencia del mismo. Finalizamos el Capítulo mostrando bajo simulación que la estimación por PMV brinda excelentes resultados a muestra finita. Por último, en el Capítulo 5 se explica por qué la segmentación contextual se reduce a encontrar una moda de una distribución de Gibbs, y como los algoritmos SA e ICM resuelven este problema. Finalizando el Capítulo con un análisis bajo simulación de la relación que existe entre los procesos de estimación y segmentación.

Capítulo 2

Procesamiento estadístico de imágenes

2.1. Introducción

El procesamiento estadístico de imágenes es un campo de notable interés debido a sus múltiples aplicaciones en distintas áreas del conocimiento, tales como medicina, biología, agricultura, ecología, etc. Hay distintos tipos de imágenes, dependiendo de la técnica utilizada para obtenerlas y de su área de aplicación. El sistema de adquisición de imágenes se basa en la interacción energética entre un objeto observado y un sistema sensor, el cual debe ser complementado con una etapa de procesamiento que permita adecuar la información adquirida al contexto requerido para su utilización. El trabajo de tesis se centra en esta última etapa.

En el libro de Gonzalez y Woods [17] y en las notas de clase realizadas por Alvarado [18], entre otros, se presenta un análisis completo de los factores que intervienen en el proceso de adquisición de imágenes. En la presente Sección se incluye un resumen de esta información con el fin de entender la naturaleza de las imágenes con las que se trabajará.

El sistema de adquisición de imágenes incluye los siguientes elementos:

1. una o varias fuentes de energía, que pueden ser de luz visible, rayos X, u otros tipos de ondas electromagnéticas, haces de partículas atómicas o subatómicas, ultrasonido, etc.
2. el objeto (u objetos) a ser capturado en la imagen. Estos objetos interactúan con la energía emitida, ya sea por reflexión, transmisión, refracción, absorción, difracción, etc.
3. la cámara o sensor es un sistema óptico o de otra índole que colecta la energía recibida y la proyecta al sistema de captura de la imagen.
4. el sistema de captura de la imagen transforma la señal proyectada a una representación apropiada al contexto; por ejemplo, la retina en el ojo humano transforma la proyección a impulsos neuronales, un chip CCD transforma la imagen a señales eléctricas aptas para su procesamiento en computador, o una película fotosensible captura la imagen de modo analógico.

El flujo energético que relaciona al sensor con el objeto constituye una forma de radiación electromagnética, cuyas propiedades se han explicado históricamente a través de dos teorías:

1. la que la concibe como un haz ondulatorio (Huygens, Maxwell), y
2. la que la considera como una sucesión de unidades discretas de energía, fotones o cuantos, con masa igual a cero (Planck, Einstein).

Actualmente se ha demostrado que la luz puede comportarse de acuerdo a ambas teorías. El flujo energético conforme a esta teoría puede describirse por dos elementos: longitud de onda y frecuencia. La primera hace referencia a la distancia entre dos picos sucesivos de una onda, mientras que la frecuencia designa el número de ciclos pasando por un punto fijo en una unidad de tiempo. Aunque la sucesión de valores de longitud de onda es continua, suelen establecerse una serie de bandas en donde la radiación electromagnética manifiesta un comportamiento similar. La Figura 2.1 ilustra los nombres brindados a las distintas regiones espectrales.

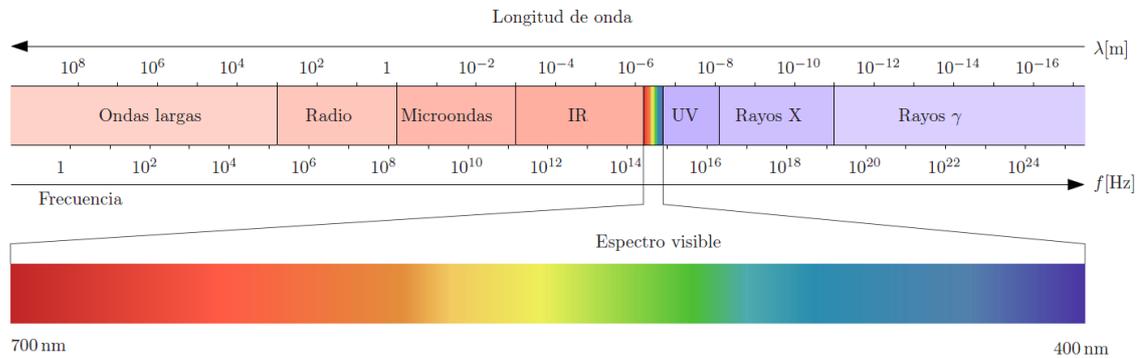


Figura 2.1: Espectro electromagnético.

Cada una de estas bandas posee propiedades que la caracterizan y que son explotadas por las tecnologías actuales para fines particulares. Por ejemplo, en la Teledetección Espacial, la banda del Infrarrojo próximo (0.7 a 1.3 mm) es de especial importancia por su capacidad para discriminar masas vegetales y concentraciones de humedad; mientras que en Medicina, los rayos X son capaces de atravesar cuerpos opacos y son especialmente útiles para captar estructuras óseas, aunque también se utilizan para diagnosticar enfermedades de los tejidos blandos.

En este trabajo las imágenes de interés son obtenidas por sensores que miran una o más bandas simultáneamente del espectro electromagnético, las cuales serán vistas como vectores de \mathbb{R}^B . Esta información se denomina información radiométrica de la imagen y es discretizada en su dominio y en su intensidad para poder ser digitalizada.

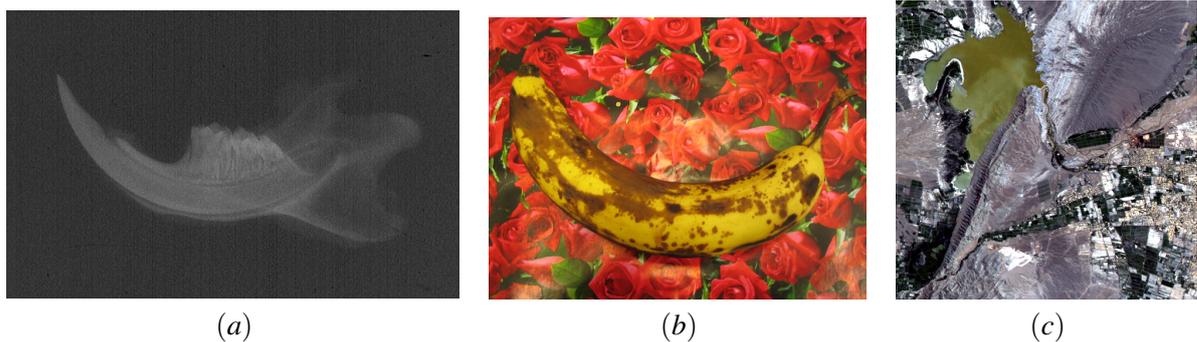


Figura 2.2: Ejemplos de datos radiométricos: (a) radiografía digitalizada inversamente, (b) fotografía RGB, (c) imagen compuesta a partir de datos multibanda Landsat.

En la Figura 2.2 aparecen ejemplos de datos radiométricos. En (a) se muestra una imagen de rayos X de la mandíbula de una ratona que se utilizará a lo largo de la tesis, la cual fue gentilmente cedida por el Odontólogo

J.G. Flesia [19, 20], de la Sección de Procesamiento de Imágenes del Departamento de Odontología de la Universidad Nacional de Córdoba, quien ayudó juzgando y discriminando resultados de los algoritmos trabajados. En (b) aparecen las tres bandas principales del espectro visible de una imagen color. En (c) aparecen 3 de las 6 bandas con las que se trabajará, de un recorte de una imagen Landsat tomada el 11 de Abril de 2005. En el recorte se puede observar la ciudad de San Juan - Argentina con montañas y el dique de Ullum a su lado.

2.2. Etiquetado

Una gran cantidad de problemas en análisis de imágenes pueden proponerse como problemas de etiquetado en los cuales la solución del problema es un conjunto de etiquetas asignadas a puntos de una imagen o características de la imagen.

Un problema de etiquetado se especifica en términos de un conjunto Λ de sitios y un conjunto Ξ de etiquetas, y se resuelve definiendo una asignación o función de etiquetado de Λ en Ξ . Un sitio representa usualmente un punto o región en el espacio Euclideo, como un pixel de una imagen, un punto de esquina, un segmento de línea o un trozo de superficie. Pueden ser regulares o no, por ejemplo, sitios que corresponden a posiciones igualmente espaciadas de una imagen son considerados espacialmente regulares. Las inter-relaciones entre los sitios se moderan usando sistemas de entornos. Las etiquetas son un evento que ocurre en un sitio, las cuales pueden ser discretas o continuas, ordenadas o no, y estas características dependen del problema a modelar. En el problema de detección de bordes, por ejemplo, las etiquetas son: “borde” y “no borde”; y Λ es el conjunto de pixeles de la imagen.

En la terminología de campos aleatorios, un etiquetado se llama configuración. En visión, una configuración o etiquetado puede corresponder a una imagen mejorada, a un mapa de bordes, a la interpretación de características de la imagen en términos de características de un objeto, a un mapa de segmentación en regiones, etc.

En términos de regularidad y continuidad, se pueden clasificar los problemas de etiquetado en visión en uno de las siguientes cuatro categorías:

- LP1) sitios regulares y etiquetas continuas
- LP2) sitios regulares y etiquetas discretas
- LP3) sitios irregulares con etiquetas discretas
- LP4) sitios irregulares con etiquetas continuas

Las primeras dos categorías caracterizan el procesamiento de bajo nivel realizado en imágenes observadas y los otros dos son procesamiento de alto nivel en características extraídas de la imagen. En la tesis se trabajará con etiquetados LP2. La asignación de valores radiométricos a cada sitio de una grilla rectangular finita Λ , y la restauración o suavizado de imágenes con pixeles de valores continuos son etiquetados de tipo LP1. La segmentación de regiones es un problema LP2 y consiste en particionar una imagen de datos radiométricos en regiones mutuamente excluyentes, de modo tal que cada una de ellas posea pixeles con propiedades uniformes y homogéneas, cuyos valores son diferentes a los de los pixeles de las particiones vecinas. Esta propiedad puede ser: tono de gris, color o textura, entre otras. La detección de bordes es también un problema LP2. Cada sitio localizado entre dos pixeles vecinos tiene asignada una etiqueta en el conjunto {borde, no borde} de acuerdo a si hay o no una diferencia significativa entre los dos pixeles. En la Figura 2.3 se muestran ejemplos de etiquetados LP1 y LP2.

Para un problema de etiquetado discreto de $\#\Lambda$ sitios en $\#\Xi$ etiquetas, hay un conjunto de $\#\Xi^{\#\Lambda}$ posibles etiquetados. Para un problema de etiquetado continuo, existe un número infinito de etiquetados. Sin embargo,

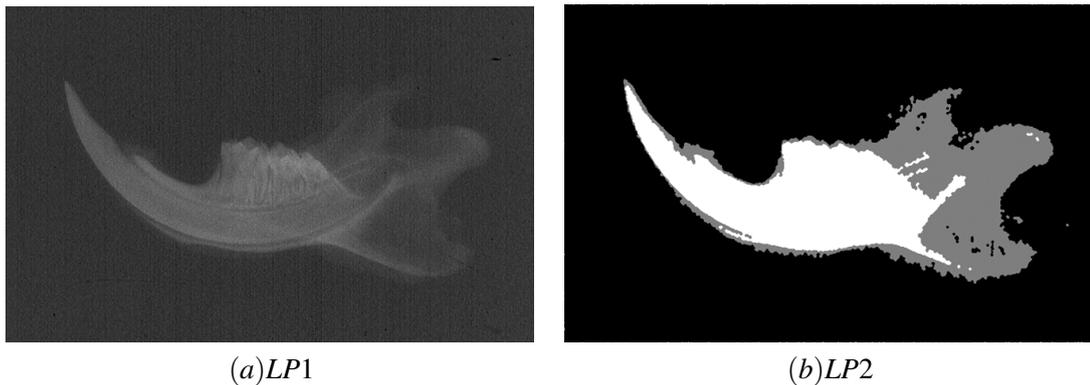


Figura 2.3: Ejemplos de etiquetados LP: (a) es una imagen que requiere un suavizado o reducción de ruido, y (b) es una segmentación de una imagen.

dentro de todos los etiquetados posibles, hay solo un número finito de ellos que son buenas soluciones, y solo muy pocos que son óptimos con respecto a algún criterio. Como definir la solución óptima para un problema y como encontrarla son dos tópicos muy importantes en el cálculo de etiquetados en visión, vía optimización.

2.3. Funciones de optimización

La optimización juega un rol muy importante en el análisis de imágenes. En esta área, un problema de visión es formulado como la optimización de algún criterio, explícito o implícito. El uso extenso de los principios de optimización en visión se debe a las incertezas de los procesos de visión y generado de imágenes, como ruido, oclusión y ambigüedades en la interpretación visual. Soluciones exactas o perfectas prácticamente no existen, pero soluciones óptimas en algún sentido pueden encontrarse, si se restringen suficientemente las condiciones de búsqueda.

Hay tres etapas básicas en el problema de optimización en visión: representación del problema, función objetivo y algoritmo de optimización. El primer problema, el de representación, concierne a la búsqueda de características representativas del fin último del análisis, que sean computables en forma eficiente y con baja incerteza. La representación de la solución refiere a la elección de los sitios y de las etiquetas para un problema de etiquetado particular. Por ejemplo, en segmentación de imágenes, se puede usar una cadena de posiciones de contorno para representar una solución, o también se puede generar un mapa de regiones para el mismo fin. El segundo problema es como formular la función objetivo para la optimización. La función objetivo mapea la solución con un número real que mide la calidad de la solución en función de alguna función de bondad o de costo. La formulación determina como se codifican en la función las distintas restricciones, que pueden ser propiedades de los pixeles como intensidad o color, y/o contexto, como relaciones entre los pixeles o características de objetos. La formulación define la solución óptima. El tercer problema es como optimizar la función objetivo, es decir, como buscar una solución óptima en el espacio admisible. Los mayores puntos a tratar son: a) el problema de mínimos locales que existe en funciones no convexas b) la eficiencia de los algoritmos en espacio y tiempo. Los tres problemas usualmente se estudian al mismo tiempo. La representación de la imagen es creada de tal forma de que esté relacionada con una función objetivo con buenas propiedades de optimización.

En esta tesis trataremos con problemas de minimización donde la función objetivo tiene la forma de una función de energía. En modelos formales, en oposición a los heurísticos, la función de energía se formula en base a criterios establecidos. Pero como la creación de las imágenes involucra incertidumbre inevitables, las funciones de energía son creadas a partir de principios de la estadística, probabilidad y teoría de la información.

Cuando se conoce (o puede suponerse) una distribución de probabilidad para la imagen a observar, pero no hay conocimiento de una distribución a priori, se utiliza el criterio de máxima verosimilitud,

$$x^* = \arg \max_x p(I|x),$$

donde x representa las posibles soluciones e I son los datos. Un ejemplo de este tipo de función de optimización es la que utiliza el clásico segmentador de Máxima verosimilitud que será definido en la Sección 2.4.

Este criterio no es el único, ya que cuando sólo hay información a priori disponible, puede usarse el criterio de máxima entropía:

$$x^* = \arg \max_x \left\{ - \sum_i p(x_i) \ln(p(x_i)) \right\}.$$

Este criterio considera que las configuraciones con mayor entropía son más probables pues la naturaleza puede generarlas de muchas formas.

Por último, cuando se cuenta con las distribuciones a priori y la función de verosimilitud, el mejor resultado se encuentra maximizando el criterio de Bayes,

$$x^* = \arg \max_x p(I|x)p(x), \quad (2.3.1)$$

obteniendo la llamada solución MAP (máximo a posteriori). Este criterio será el central en la tesis, utilizando modelos Markovianos para las distribuciones a priori, y modelos paramétricos para los datos radiométricos.

2.4. Modelos

Formalmente vamos a considerar una imagen como un arreglo de la forma $I = [I_{ij} : (i, j) \in \mathcal{S}]$, siendo \mathcal{S} el conjunto de los sitios físicos de la imagen que se denomina *soporte de la imagen*. En las cuestiones teóricas que plantea la tesis se trabajará con $\mathcal{S} = \mathbb{Z}^d$, $d \in \mathbb{N}$, pero las particularizaciones de las mismas al contexto del procesamiento de imágenes y las simulaciones que se realizan suponen que $d = 2$. Una observación de una imagen no se puede lograr sobre \mathcal{S} , pero sí sobre un subconjunto finito Λ del mismo. Luego una observación de una imagen es un arreglo de la forma $I = [I_{ij} : (i, j) \in \Lambda]$, donde Λ se considerará un cuadrado de la forma $([-n, n] \cap \mathbb{Z})^2$.

Consideraremos que cada entrada de una imagen I_{ij} toma valores en un conjunto de estados Ξ , que puede ser un *conjunto vectorial* como en el caso de imágenes radiométricas, o *categorico* como en el caso de imágenes de clases. En este contexto una determinada imagen será una observación o realización de este arreglo. En la Figura 2.4 se muestran ejemplos de conjuntos Ξ , y como una imagen en tono de grises (*a*) se convierte en una imagen a color (*b*) con sólo aumentar la dimensión de Ξ . Por otro lado, las segmentaciones (*c*) y los mapas de bordes (*d*) presentan un conjunto Ξ categorico. En la presente tesis se trabajará con mapas de clases, en los que Ξ es un conjunto finito de etiquetas, que denotaremos más precisamente por $E = \{\ell_1, \ell_2, \dots, \ell_L\}$, y con imágenes radiométricas multiespectrales con $\Xi = \mathbb{R}^B$, siendo B la cantidad de bandas espectrales.

Cada clase $\ell \in E$, emite información radiométrica de acuerdo a una ley de distribución f_ℓ sobre \mathbb{R}^B . En ambas imágenes de la Figura 2.5 aparecen: a la izquierda la imagen original de la mandíbula de una rata con un recuadro rojo; a la derecha y abajo un zoom del recuadro rojo; y a la derecha y arriba el histograma de los datos radiométricos dentro del recuadro rojo. Si se observan detalladamente los histogramas, se puede notar que cada clase presenta información radiométrica diferente, y que dada la clase, los histogramas tienden a presentar una determinada estructura.

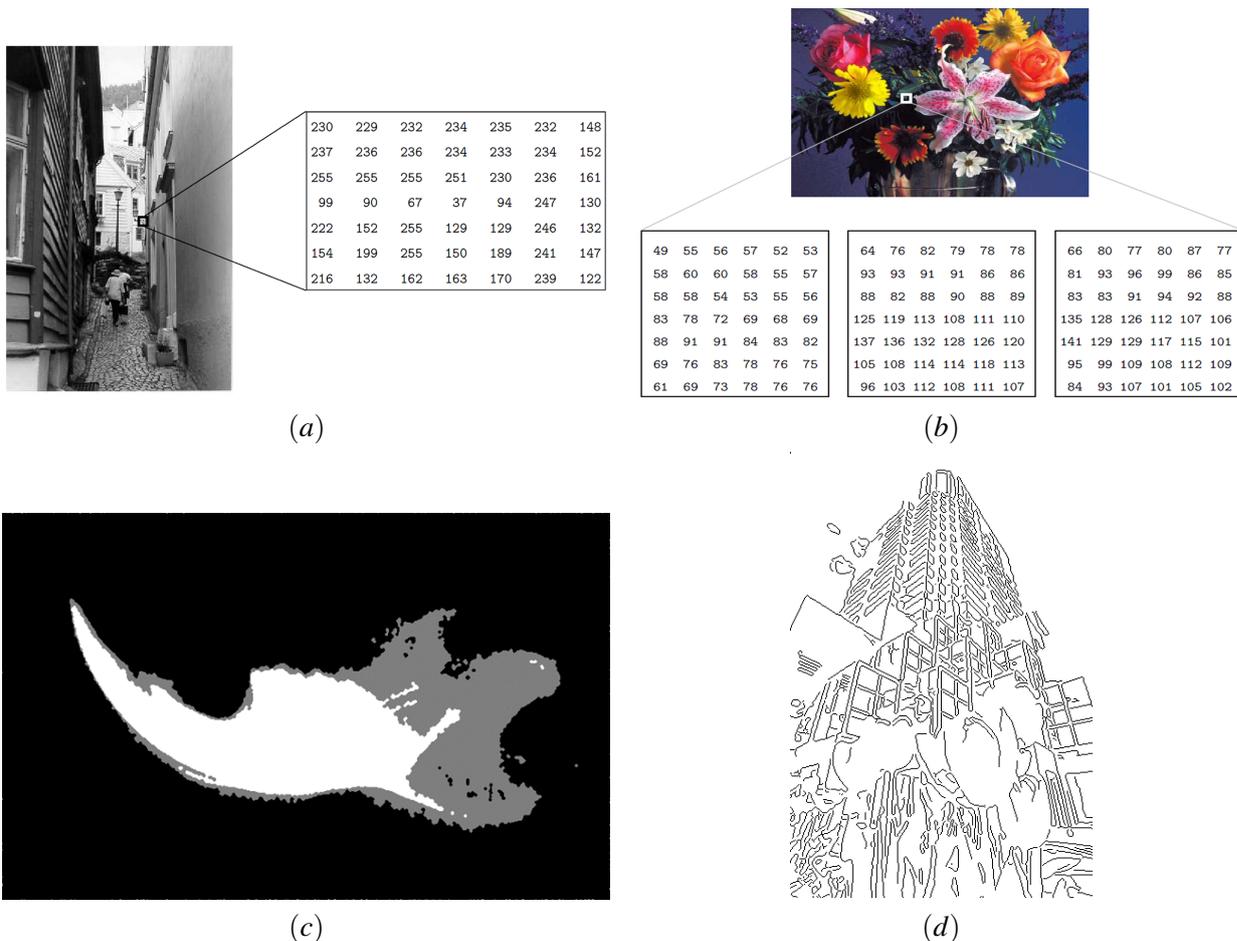


Figura 2.4: Ejemplos de imágenes: (a) $\Xi = \{1,2,\dots,255\}$; (b) $\Xi = \{1,2,\dots,255\}^3$; (c) $\Xi = \{\text{diente, cartílago, fondo}\}$; (d) $\Xi = \{\text{borde, no borde}\}$.

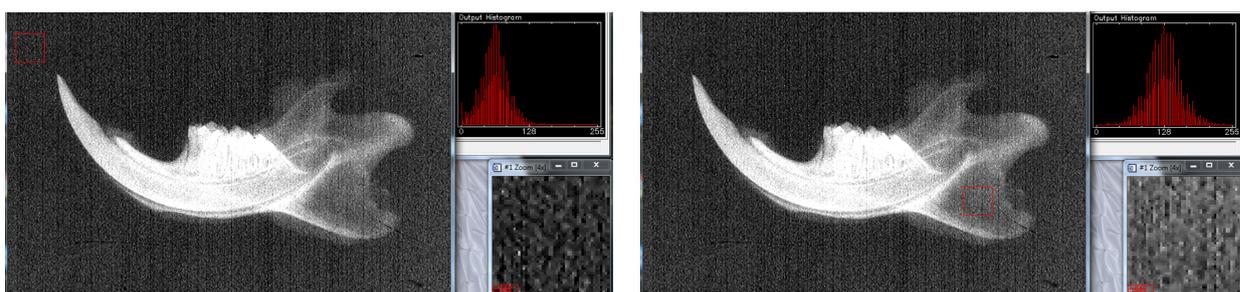


Figura 2.5: Histograma de los datos radiométricos dada la clase. A la izquierda, vemos un recorte del fondo de la imagen y su correspondiente histograma y a la derecha un recorte de la parte de cartílago de la imagen, y su correspondiente histograma. Al aumentar los recortes podemos observar el ruido de adquisición que la imagen tiene, que se asemeja a una textura.

Formalmente, dotemos al conjunto de etiquetas E con la σ -álgebra $\mathcal{E} = \mathcal{P}(E)$ de las partes de E y de la medida que cuenta λ . Consideremos además el espacio producto $(E^{\mathcal{S}}, \mathcal{E}^{\mathcal{S}}, \lambda^{\mathcal{S}})$. Dado un nuevo Espacio de Probabilidad (Ω, \mathcal{A}, P) , si $X : \Omega \rightarrow E^{\mathcal{S}}$ es tal que

$$\{\omega \in \Omega : X(\omega) \in B\} \in \mathcal{A} \quad \forall B \in \mathcal{E}^{\mathcal{S}},$$

y si existe una probabilidad π sobre $(E^{\mathcal{S}}, \mathcal{E}^{\mathcal{S}})$ tal que

$$P(\{\omega \in \Omega : X(\omega) \in B\}) = \pi(B) \quad \forall B \in \mathcal{E}^{\mathcal{S}},$$

entonces diremos que el campo aleatorio X tiene distribución π . Supongamos que X es el campo aleatorio que asigna una etiqueta a cada pixel, y nombremos a π como el modelo a priori del mapa de clases. Además, supongamos que cada clase $\ell \in E$ emite observaciones radiométricas en \mathbb{R}^B , $B \in \mathbb{N}$ fijo, de acuerdo a una densidad f_{ℓ} . Esto es, para cada $\ell \in E$, existe $f_{\ell} : \mathbb{R}^B \rightarrow (0, \infty)$, tal que $\int f_{\ell}(\zeta) d\zeta = 1$. Para cada $s \in \mathcal{S}$, sea $\mathbf{I}_s : \Omega \rightarrow \mathbb{R}^B$ (dato radiométrico sobre s) tal que la densidad de \mathbf{I}_s , dado que $X_s = \ell$, viene dada por f_{ℓ} . Por simplicidad notaremos esto con

$$P(\mathbf{I}_s = I_s | X_s = \ell) = f_{\ell}(I_s) \quad \forall \ell \in E.$$

Sea $\mathbf{I} : \Omega \rightarrow (\mathbb{R}^B)^{\mathcal{S}}$ el campo aleatorio dado por $\mathbf{I}(\omega) = (\mathbf{I}_s(\omega))_{s \in \mathcal{S}}$ que llamaremos datos radiométricos. Ahora, dados los datos radiométricos $\mathbf{I} = I$, la asignación de probabilidades a cada posible mapa de clases x es diferente y la representaremos con Π , esto es,

$$\Pi(x) := P(X = x | \mathbf{I} = I).$$

Este nuevo modelo sobre $E^{\mathcal{S}}$ será denominado modelo a posteriori del mapa de clases.

Las distribuciones de emisión f_{ℓ} seleccionadas dependen de las características del sensor y de las clases de imágenes consideradas. En los ejemplos de la tesis se trabajará exclusivamente con distribuciones Gaussianas Multivariadas, aunque los análisis se pueden generalizar a cualquier clase de distribuciones. Una vez seleccionada la clase de distribuciones paramétricas con parámetros $\theta \in \Theta$ con las que se trabajará, debemos identificar un parámetro θ_{ℓ} para cada clase $\ell \in E$. Esto se puede hacer de un modo supervisado o de un modo no supervisado. Estos procedimientos se diferencian en que el método supervisado necesita de una muestra de cada clase, que le permita estimar cada $\hat{\theta}_{\ell}$. En cambio, el no supervisado, bajo determinadas hipótesis, es capaz de estimar sin contar con muestras de entrenamiento. Las estimaciones de máxima verosimilitud basadas en las muestras consideradas, son un ejemplo de estimación supervisada. Con respecto al segundo, una particularización del algoritmo Expectation Maximization (EM) [21, 22] para estimar por máxima verosimilitud una mezcla de distribuciones puede ser utilizada para tal fin. En la tesis supondremos que los modelos son conocidos o que ya fueron estimados.

De este modo, tenemos un modelo de emisión para cada clase, con los cuales, si contamos con una imagen observada sobre Λ , se puede asignar una clase a cada pixel $s \in \Lambda$, de acuerdo a la verosimilitud con la cual este pixel pertenece a cada clase. Matemáticamente, este procedimiento se puede modelar por medio de una función

$$C : \Lambda \rightarrow E,$$

y en la literatura suele denominarse como clasificación o segmentación de una imagen. En realidad ambos términos no significan lo mismo, pero son equivalentes en el contexto de la tesis. Segmentar significa particionar un objeto de acuerdo a un determinado criterio, en cambio, clasificar significa etiquetar objetos de acuerdo a sus características. Por ejemplo, la Figura 2.3(b) muestra una segmentación de la imagen de la Figura 2.3(a). Por otro lado, un ejemplo de clasificación puede ser darse cuenta que la Figura 2.3(a) es la mandíbula de una rata, y que la Figura 2.4(b) es un recipiente con flores. Se considera que ambos términos son equivalentes en la tesis,

porque lo que se considera como segmentación en nuestro contexto, puede ser visto como una clasificación de cada pixel de la imagen, con las etiquetas dadas en E .

En la literatura existe una gran diversidad de segmentadores de imágenes, y cada uno de ellos es óptimo o subóptimo de acuerdo a algún criterio. Chin-Wei et al. en su trabajo [23] hacen un detallado estudio del uso de métodos de optimización en procedimientos de clasificación de imágenes. Al igual que con la selección de los modelos de emisión, estas reglas de segmentación pueden ser supervisadas o no supervisadas, dependiendo de si requieren o no de muestras de entrenamiento para reconocer determinadas características de cada clase, y con dicha información ser capaz de clasificar un pixel en una clase u otra de un modo eficiente. En el caso de las reglas no supervisadas, no existe tal muestra de entrenamiento, pero su correspondiente algoritmo incluye de algún modo un procedimiento que le permite distinguir las clases.

Dentro de la clase de segmentadores supervisados, se encuentra el clasificador de Máxima Verosimilitud (MV). El mismo consiste en asignarle a cada pixel $s \in \mathcal{S}$, del cual se posea una observación $I_s \in \mathbb{R}^B$, la clase

$$x_s = \arg \max_{\ell \in E} f_{\ell}(\mathbf{I}_s),$$

que maximiza la verosimilitud de las observaciones. Este clasificador minimiza el riesgo de Bayes bajo la hipótesis de independencia entre los pixeles. En la Figura 2.6 se muestra en (b) un ejemplo de clasificación por MV. La imagen presenta una gran cantidad de ruido, provocando que los modelos de emisión se entrecrucen, y por ende, la salida de MV es una segmentación heterogénea.

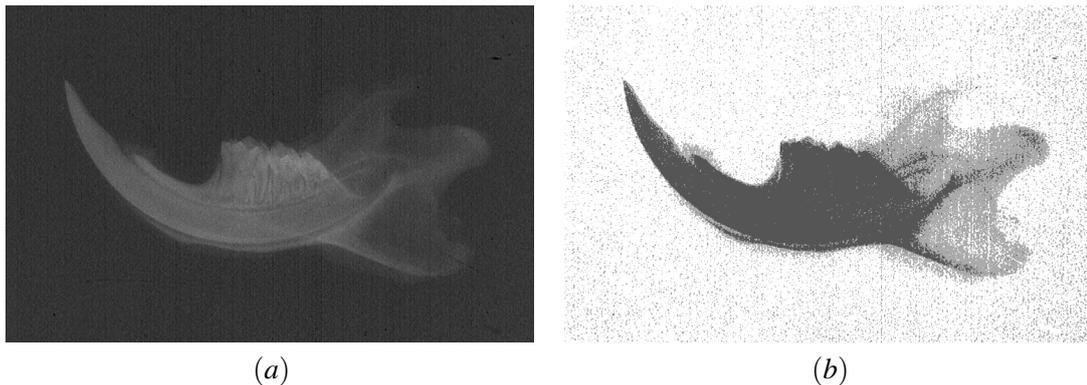


Figura 2.6: Ejemplo de segmentación por MV (máxima verosimilitud): (a) imagen original; (b) clasificación MV de (a).

El algoritmo EM, permite incorporar un clasificador a su procedimiento de estimación no supervisado. La idea básica del mismo es, inicializando con una segmentación arbitraria, iterar entre estimar por MV los parámetros del modelo de emisión, y re-clasificar la imagen por MV. Este segmentador iterativo es no supervisado, debido a que clasifica sin entrenar, y elimina los efectos que puede producir una muestra de entrenamiento errónea, pero lo hace menos robusto a la incorrecta selección del modelo de emisión y más dependiente de la segmentación inicial. La Figura 2.7 muestra un ejemplo de clasificación por medio de EM con modelos de emisión Gaussianos.

Las Figuras 2.6 y 2.7 son una clara muestra de que clasificar cada pixel, con o sin supervisión, independientemente de los demás pixeles, provoca segmentaciones ruidosas cuando la imagen a segmentar no es de calidad o contiene modelos de emisión similares. En la práctica uno busca que la clase asignada a un determinado pixel esté vinculada a la de sus vecinos, evitando la existencia de pixeles clasificados aisladamente a una determinada clase, debido a que los mismos suelen provenir de datos radiométricos erróneos.

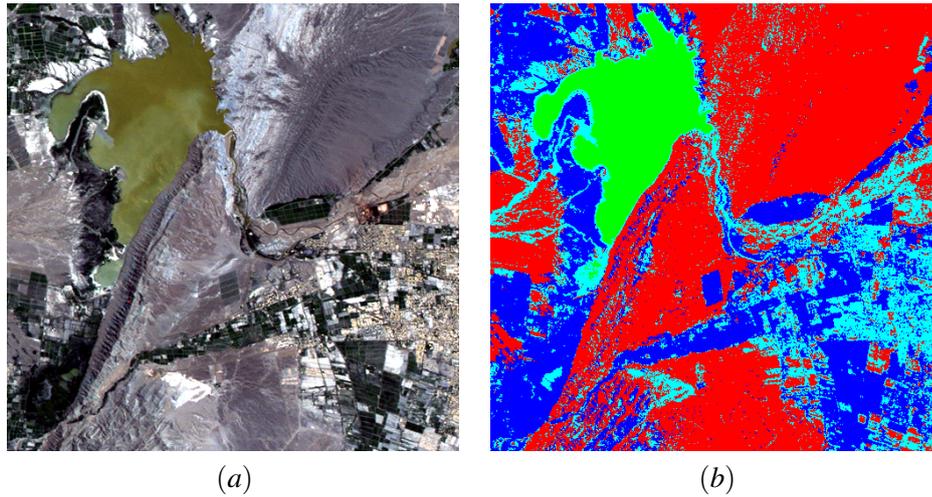


Figura 2.7: Ejemplo de segmentación por EM (expectation-maximization): (a) imagen Landsat de la ciudad de San Juan; (b) EM con modelo de emisión Gaussiano para cada clase.

Los Campos Markovianos brindan un marco teórico adecuado para agregar información contextual al proceso de segmentación de imágenes digitales, esto es, asignarle una clase a cada pixel $s \in \mathcal{S}$, teniendo en cuenta no sólo la observación I_s , sino también, la clase asignada a los pixeles de una vecindad $\partial s \subset \mathcal{S}$. Esta teoría puede dividirse en dos ramas principales: la causal y la no causal. En la causal se modela una secuencia de sucesos aleatorios que se distribuyen en el espacio tiempo, y que cada uno de ellos depende solamente de lo ocurrido en el pasado, e influyen en los sucesos futuros. Por otro lado, en la no causal, el sistema de vecindades que regula la dependencia estadística entre los sitios del soporte no puede ser vinculado con una relación de pasado-futuro.

Los Hidden Markov Model (HMM) unidimensionales han probado ser una herramienta probabilística muy útil para modelar etiquetas no observables suponiendo una dependencia causal en el tiempo. Las adaptaciones 2D de los mismos aplicables a procedimientos de segmentación presentan espacios paramétricos con gran dimensionalidad, lo cual les permite modelar una gran diversidad de clases de imágenes, pero también se encuentran afectados por la llamada maldición de la dimensionalidad. La estimación de dichos parámetros mediante el algoritmo de Baum-Welch (caso particular del algoritmo EM para HMM bidimensional) no puede calcularse en tiempo finito y se han propuesto diversas aproximaciones, como la propuesta de extensión de PCVT (Path Constrained Viterbi Algorithm) de Li et al. [24] y la nuestra, publicada en Baumgartner et al. [25], llamada CEP. La aproximación propuesta por Li, PCVT, fue estudiada, implementada y su desempeño comparado con segmentación contextual mediante procesos no causales Gibbsianos en Flesia et al. (2013) [26], usando imágenes radiográficas cedidas gentilmente por el Od. Flesia. El estudio teórico de dichos procesos no causales son el centro de esta tesis.

Los Campos Aleatorios Markovianos (MRF) brindan un marco teórico adecuado para modelar la información contextual de un mapa de clases en un modo no causal, basándose en la siguiente suposición Markoviana:

$$P(X_s = x_s | X_t = x_t, t \neq s) = P(X_s = x_s | X_t = x_t, t \in \partial s) \quad \forall s \in \mathcal{S}.$$

Si \mathcal{S} fuese finito, entonces existiría una única ley conjunta global sobre el espacio de mapas de clases $E^{\mathcal{S}}$, cuyas condicionales coincidirían con las distribuciones que verifican esta suposición, pero como \mathcal{S} es infinito, se necesitan hipótesis adicionales para asegurar esta existencia y unicidad. Estas leyes conjuntas se denominan distribuciones de Gibbs y serán tratadas con rigurosidad en el Capítulo 3. En el Capítulo 5 veremos que si el

modelo a priori es una distribución de Gibbs, entonces el modelo a posteriori, bajo algunos supuestos, es también una distribución de Gibbs, y que por lo tanto hallar la segmentación óptima se reducirá a hallar máximos globales de una distribución de Gibbs. Allí veremos que las segmentaciones que se obtienen son más homogéneas si se incorpora información contextual. En la práctica, debemos seleccionar la distribución de Gibbs que se utilizará para modelar la información a priori del mapa de clases. El modelo de Potts (ver Sección 3.5) es una subfamilia de los Campos Markovianos, y es el modelo elegido en la tesis para modelar la interrelación existente entre las etiquetas del mapa, por su simplicidad matemática y por su bajo costo computacional. Este modelo posee solamente un parámetro, y las segmentaciones por medio de ICM poseen propiedades teóricas de convergencia, a diferencia de las mencionadas aproximaciones para los HMM, Flesia et al. [26].

Capítulo 3

Medidas de Gibbs

En el presente Capítulo se detallarán las cuestiones teóricas necesarias y suficientes sobre medidas de Gibbs para abordar el problema teórico principal de la presente tesis, el cual es probar la consistencia fuerte del estimador de Pseudo-Máxima Verosimilitud (PMV) para una clase particular de distribuciones de Gibbs generadas por familias de especificaciones no invariantes por traslaciones definidas sobre el retículo regular \mathbb{Z}^d , $d \in \mathbb{N}$. Para tal fin, se reemplazará la hipótesis de invarianza de las especificaciones por una serie de suposiciones que harán que el modelo no invariante sea equivalente microscópicamente a un modelo invariante asociado. El trabajo será autocontenido y seguirá la notación y el trasfondo teórico presentado en el libro de Georgii [27]. Como resultado adicional se mostrará que la prueba dada por Geman y Graffigne en [4] puede ser extendida a una familia exponencial invariante más grande.

3.1. Introducción

En el presente capítulo consideraremos que $\mathcal{S} = \mathbb{Z}^d$, $d \in \mathbb{N}$, E es un conjunto finito, y X es un campo aleatorio que toma valores en $E^{\mathcal{S}}$. Supongamos que la distribución de X es una determinada probabilidad π^ϑ con parámetro $\vartheta \in \text{Int}(\Upsilon)$. Supongamos además que $\pi_\Lambda^\vartheta : E^\Lambda \times E^{\mathcal{S} \setminus \Lambda} \rightarrow [0, 1]$ es la probabilidad condicional con la que X asigna estados a una región finita $\Lambda \subset \mathcal{S}$ dada la configuración del resto de la grilla. Estas probabilidades condicionales tienen una relación muy estrecha con un concepto fundamental en la teoría de medidas de Gibbs, denominado especificaciones. Cuando π_Λ^ϑ y $\pi_{\Lambda+s}^\vartheta$ son iguales salvo traslación para cada $s \in \mathcal{S}$ y cada $\Lambda \subseteq \mathcal{S}$ finito, se dice que las especificaciones del modelo son invariantes, y en caso contrario, se dice que son no invariantes.

Hablando informalmente a modo de introducción, supongamos que contamos con una observación del modelo $x_\Lambda \in E^\Lambda$ sobre Λ y que nos planteamos estimar ϑ . El estimador de Máxima Verosimilitud (MV) de ϑ viene dado por

$$\widehat{\vartheta}_\Lambda^{MV}(x_\Lambda) = \arg \max_{\vartheta} U_\Lambda^{MV}(x_\Lambda, \vartheta),$$

con

$$U_\Lambda^{MV}(x_\Lambda, \vartheta) := \ln \pi_{\Lambda^\circ}^\vartheta(x_{\Lambda^\circ} | x_{\Lambda \setminus \Lambda^\circ}),$$

donde Λ° es el interior de Λ , el cual será definido en la Sección 3.2. Debido a que estimadores clásicos, como MV, son costosos computacionalmente, Besag [28] propuso un estimador alternativo llamado estimador Coding, el cual consiste en seleccionar una subfamilia de sitios de la grilla independientes condicionalmente para luego maximizar un producto de densidades condicionales. Posteriormente, debido a las limitaciones que observó en este estimador, Besag [1] lo mejoró definiendo el estimador de Pseudo-Máxima Verosimilitud (PMV),

el cual viene dado por

$$\widehat{\vartheta}_{\Lambda}^{PMV}(x_{\Lambda}) = \arg \max_{\vartheta} U_{\Lambda}^{PMV}(x_{\Lambda}, \vartheta),$$

con

$$U_{\Lambda}^{PMV}(x_{\Lambda}, \vartheta) := \frac{1}{\#\Lambda^{\circ}} \sum_{s \in \Lambda^{\circ}} \ln \pi_s^{\vartheta}(x_s | x^s).$$

Estos estimadores forman parte de una clase más grande de estimadores denominados Estimadores de Mínimo Contraste $\widehat{\vartheta}_{\Lambda}(x_{\Lambda})$, los cuales son el argumento que minimiza un funcional $U_{\Lambda}(x_{\Lambda}, \vartheta)$ bajo determinadas condiciones.

Si consideramos la sucesión creciente $\{\Lambda_n : n \in \mathbb{N}\}$ de subconjuntos de \mathcal{S} dados por $\Lambda_n = ([-n, n] \cap \mathbb{Z})^d$, la consistencia fuerte del estimador consiste en ver que

$$\pi^{\vartheta} \left(\left\{ x \in E^{\mathcal{S}} : \lim_{n \rightarrow \infty} \|\widehat{\vartheta}_{\Lambda_n}(x_{\Lambda_n}) - \vartheta\| = 0 \right\} \right) = 1.$$

En los libros de Guyon [7, 8], se estudia la consistencia fuerte y débil de los Estimadores de Mínimo Contraste, dando condiciones sobre U_{Λ} para que se verifiquen ambos tipos de consistencias. En cuanto a la consistencia fuerte, dentro de la lista de condiciones que piden el Teorema 3.4.4 de [7] y la Proposición C.1 de [8], se encuentran:

(i) $U_{\Lambda_n}(\cdot, x_{\Lambda_n})$ es convexa $\forall n \in \mathbb{N}$;

(ii) Existe una función $K : \Upsilon \rightarrow [0, \infty)$ con un único mínimo en ϑ tal que

$$U_{\Lambda_n}(\alpha, \cdot) - U_{\Lambda_n}(\vartheta, \cdot) \rightarrow K(\alpha) \text{ en } \pi^{\vartheta}\text{-probabilidad } \forall \alpha \in \Upsilon.$$

Por otro lado, el Teorema 3.4.3 de [7] afirma que la consistencia fuerte vale bajo condiciones tales como:

(iii) si para cada $n \in \mathbb{N}$, $\eta > 0$ y $x \in E^{\mathcal{S}}$, se define

$$W_n(\eta, x) := \sup(\{|U_{\Lambda_n}(\alpha, x_{\Lambda_n}) - U_{\Lambda_n}(\beta, x_{\Lambda_n})| : \alpha, \beta \in \Upsilon, \|\alpha - \beta\| \leq \eta\});$$

entonces existe una sucesión de números reales positivos $(\varepsilon_k)_{k \in \mathbb{N}}$ tal que $\varepsilon_k \rightarrow 0$ y

$$\pi^{\vartheta} \left(\left\{ x \in E^{\mathcal{S}} : \limsup_{n \rightarrow \infty} W_n(1/k, x) > \varepsilon_k \right\} \right) = 0 \quad \forall k \in \mathbb{N};$$

(iv) existe una función $K : \Upsilon \rightarrow [0, \infty)$ con un único mínimo en ϑ tal que

$$\lim_{n \rightarrow \infty} \pi^{\vartheta} \left(\left\{ x \in E^{\mathcal{S}} : U_{\Lambda_n}(\alpha, x_{\Lambda_n}) - U_{\Lambda_n}(\vartheta, x_{\Lambda_n}) \geq K(\alpha) - \varepsilon \right\} \right) = 1 \quad \forall \varepsilon > 0 \forall \alpha \in \Upsilon.$$

A pesar de que estos resultados incluyen al caso no invariante, la prueba de estas condiciones no son triviales y es importante buscar condiciones más razonables que se adapten a los estimadores de PMV.

Geman y Graffigne [4] probaron la consistencia fuerte del estimador de PMV definido sobre familias exponenciales [29] para el caso en el que las especificaciones sean invariantes por traslaciones. La esencia de la prueba consistió en verificar que la función $U_{\Lambda_n}^{PMV}(\cdot, x_{\Lambda_n})$ es estrictamente convexa para cada $n \gg 1$ (lo suficientemente grande) y que el límite de estas funciones es una función con único máximo en ϑ . En esencia lo que se probó fueron las condiciones (i) y (ii) mencionadas anteriormente dentro del caso invariante. Gidas en [30] realizó una prueba alternativa. Posteriormente, aún dentro del caso invariante, Comets [5] generalizó este

resultado prescindiendo de la finitud de E , probando además que la velocidad de convergencia es exponencial. Comets y Gidas [31] probaron la consistencia de la estimación basada en datos observados parcialmente. Janzura [3] presentó un completo resumen de las propiedades asintóticas que satisface el estimador de PMV para modelos invariantes. Pickard [32] brindó un completo análisis de la problemática de estimar modelos Markovianos, ejemplificando con el simple, pero para nada trivial, modelo de Ising.

En el contexto de segmentaciones de imágenes, como veremos en las Secciones 3.5 y 5.1, si el modelo a priori del mapa de clases es representado por un modelo gibbsiano, entonces dada la evidencia radiométrica, el modelo a posteriori no es invariante por traslaciones. Luego, la consistencia del estimador de PMV basado en tal modelo [33] debe ser probada desde un nuevo punto de vista.

En el presente trabajo se adaptará la prueba de Geman y Graffigne [4] a una familia exponencial invariante más grande de la allí considerada. Posteriormente, y como resultado principal del presente trabajo, se mostrará que, bajo determinadas condiciones, la consistencia fuerte vale también si se trabaja con especificaciones o potenciales no invariantes por traslaciones. Este resultado se basa en el hecho que bajo determinadas condiciones la clase de especificaciones no invariantes con la que trabajaremos es equivalente microscópicamente a un modelo asociado con especificaciones invariantes. Además, como resultado extra, se desarrollará una nueva versión de la ley fuerte de los grandes números en el contexto de probabilidades definidas sobre espacios productos arbitrarios, que permitirá afrontar resultados asintóticos sin necesidad de acudir a la ergodicidad de las medidas de Gibbs involucradas.

El Capítulo se organizará como sigue: en la Sección 3.2 se darán las definiciones y los resultados básicos necesarios para el trabajo siguiendo la notación utilizada en el libro de Georgii [27]. Esto permitirá transcribir resultados probados allí para luego utilizarlos. En la Sección 3.3 se definirá esperanza condicional, y se verá por qué las especificaciones pueden ser vistas como las probabilidades condicionales π_Λ^θ . En la Sección 3.4 se darán condiciones necesarias y suficientes para que las especificaciones sean invariantes, dentro de la clase de modelos exponenciales con la que se trabajará. En la Sección 3.5 se definirán los modelos de Potts y se mostrará un ejemplo de modelo con especificaciones invariantes por traslaciones y otro no invariante. Ambos ejemplos se contextualizan dentro del proceso de segmentación de imágenes, y fueron el origen del presente trabajo, debido a que se deseaba probar la consistencia del estimador se introdujo en el trabajo Gimenez et al. [33], y que se desarrolla detalladamente en la Sección 4.2. En la Sección 3.6 se dará una nueva versión de la Ley fuerte de los grandes números para variables independientes no idénticamente distribuidas que se utilizará en la prueba de los Teoremas principales de consistencia fuerte. En la Sección 3.7 se explicará en que consiste una muestra de un modelo definido sobre un espacio infinito y se detalla la significancia de analizar las propiedades asintóticas del estimador de PMV. En la Sección 3.8 se desarrollarán las pruebas de consistencia fuerte mencionadas. Por último, en la Sección 3.9 abordamos el problema de obtener realizaciones de un determinado modelo, centrándonos particularmente en el modelo de Potts isotrópico sin campo externo, por medio de los algoritmos Gibbs Sampler y Swendsen-Wang.

3.2. Definiciones.

Consideremos entonces la grilla $\mathcal{S} = \mathbb{Z}^d$ y supongamos que cada sitio $s \in \mathcal{S}$ toma un único elemento en un conjunto de estados E con

$$(H.1) \quad \#E < \infty.$$

Luego $E^{\mathcal{S}}$ contiene todas las posibles configuraciones que se pueden encontrar sobre \mathcal{S} . Además, dado $x \in E^{\mathcal{S}}$ y $\Lambda \subset \mathcal{S}$, notaremos con $x_\Lambda := \{x_s\}_{s \in \Lambda}$ y con $x^\Lambda := \{x_s\}_{s \notin \Lambda}$ a los respectivos subarreglos. En caso de que $\Lambda = \{s\}$, simplemente escribiremos x_s y x^s . Sobre E consideremos la σ -álgebra $\mathcal{E} = \mathcal{P}(E)$ y la medida que cuenta λ , y

sobre $E^{\mathcal{S}}$ la σ -álgebra producto $\mathcal{F} := \mathcal{E}^{\mathcal{S}}$ y la correspondiente medida producto. Para cada $\Lambda \subset \mathcal{S}$ consideremos las σ -álgebras

$$\mathcal{F}_{\Lambda} := \{\sigma_{\Lambda}^{-1}(A) : A \in \mathcal{E}^{\Lambda}\} \quad \text{y} \quad \mathcal{J}_{\Lambda} := \mathcal{F}_{\mathcal{S} \setminus \Lambda},$$

donde σ_{Λ} es la función proyección de $E^{\mathcal{S}}$ en E^{Λ} dada por $\sigma_{\Lambda}(x) := x_{\Lambda}$. Además, consideremos la σ -álgebra

$$\mathcal{J} = \bigcap_{\Lambda \in \mathfrak{G}} \mathcal{J}_{\Lambda},$$

y la familia

$$\mathcal{F}_0 = \bigcup_{\Lambda \in \mathfrak{G}} \mathcal{F}_{\Lambda},$$

que por definición de espacio producto, verifica que \mathcal{F} es la σ -álgebra generada por \mathcal{F}_0 .

Proposición 3.2.1 *Dados $\Lambda, \Delta \subseteq \mathcal{S}$ y una función $f : E^{\mathcal{S}} \rightarrow \mathbb{R}^m$, $m \in \mathbb{N}$, se satisfacen:*

- (i) $\Lambda \subseteq \Delta \Leftrightarrow \mathcal{F}_{\Lambda} \subseteq \mathcal{F}_{\Delta} \Leftrightarrow \mathcal{J}_{\Delta} \subseteq \mathcal{J}_{\Lambda}$;
- (ii) f es \mathcal{F}_{Λ} -medible $\Leftrightarrow f(x_1) = f(x_2) \quad \forall x_1, x_2 \in E^{\mathcal{S}} : \sigma_{\Lambda}(x_1) = \sigma_{\Lambda}(x_2)$;
- (iii) $\Delta \subseteq \Lambda$ y f es \mathcal{F}_{Δ} -medible $\Rightarrow f$ es \mathcal{F}_{Λ} -medible.
- (iv) f es \mathcal{F}_{Δ} -medible \Rightarrow la función $x \mapsto \int_{E^{\Lambda}} f(\xi x^{\Lambda}) \lambda^{\Lambda}(d\xi)$ es $\mathcal{F}_{\Delta \setminus \Lambda}$ -medible.

Dem. Ver Sección 6.1 del Apéndice. □

Para trabajar con modelos en los que los estados de distintos sitios de \mathcal{S} estén vinculados, se define una vecindad para cada sitio, que regule tal vínculo. Para ello consideremos un sistema de vecindades $\partial = \{\partial s\}_{s \in \mathcal{S}}$ sobre \mathcal{S} , esto es, una familia de subconjuntos de \mathcal{S} que verifica:

$$(H.2) \quad s \notin \partial s \quad \forall s \in \mathcal{S};$$

$$(H.3) \quad s \in \partial t \Leftrightarrow t \in \partial s \quad \forall s, t \in \mathcal{S};$$

$$(H.4) \quad \#\partial s < \infty \quad \forall s \in \mathcal{S};$$

$$(H.5) \quad t + \partial s = \partial(s + t) \quad \forall s, t \in \mathcal{S}.$$

Notemos con $\mathfrak{G} := \{\Lambda \subset \mathcal{S} : 0 < \#\Lambda < \infty\}$ a la familia de subconjuntos no vacíos finitos de \mathcal{S} , y con $\mathcal{C} := \{\Lambda \subset \mathcal{S} : s, t \in \Lambda, s \neq t \Rightarrow t \in \partial s\}$ al conjunto de cliques. Dado $\Lambda \subset \mathcal{S}$, definimos su frontera $\partial\Lambda := \{s \in \mathcal{S} \setminus \Lambda : \partial s \cap \Lambda \neq \emptyset\}$, su interior $\Lambda^{\circ} := \{s \in \Lambda : \partial s \subseteq \Lambda\}$, y $\Lambda^* := \Lambda \cup \partial\Lambda$.

Definición 3.2.1 *Diremos que una familia $\Phi = \{\Phi_{\Lambda}\}_{\Lambda \in \mathfrak{G}}$ es un potencial, si cada Φ_{Λ} es una función de $E^{\mathcal{S}}$ en \mathbb{R} , que es \mathcal{F}_{Λ} -medible.*

Supongamos que contamos con una familia paramétrica de potenciales $\{\Phi^{\gamma}\}_{\gamma \in \Upsilon}$, con $\Upsilon \subseteq \mathbb{R}^m$ el espacio de parámetros que verifica:

$$(H.6) \quad \Upsilon \subseteq \mathbb{R}^m \text{ es convexo};$$

(H.7) $\text{Int}(\Upsilon) \neq \emptyset$.

Asumamos que estos potenciales satisfacen las siguientes condiciones:

(H.8) Φ^ϑ es sumable $\forall \vartheta \in \Upsilon$, esto es:

$$\sum_{\Lambda \in \mathfrak{S}: s \in \Lambda} \|\Phi_\Lambda^\vartheta\|_\infty < \infty \quad \forall s \in \mathcal{S} \quad \forall \vartheta \in \Upsilon;$$

(H.9) $\Phi_\Lambda^\vartheta \equiv 0 \quad \forall \Lambda \notin \mathcal{C} \quad \forall \vartheta \in \Upsilon$.Llamamos función de energía en $\Lambda \in \mathfrak{S}$, asociada al potencial Φ^ϑ , a la función $H_\Lambda^\vartheta : E^{\mathcal{S}} \rightarrow \mathbb{R}$ dada por

$$H_\Lambda^\vartheta(x) := \sum_{\Delta \in \mathfrak{S}(\Lambda)} \Phi_\Delta^\vartheta(x), \quad (3.2.1)$$

donde $\mathfrak{S}(\Lambda) := \{\Delta \in \mathfrak{S} : \Lambda \cap \Delta \neq \emptyset\}$.**Proposición 3.2.2** H_Λ^ϑ es \mathcal{F}_{Λ^*} -medible $\forall \vartheta \in \Upsilon \quad \forall \Lambda \in \mathfrak{S}$.**Dem.** Ver Sección 6.1 del Apéndice. □

Trabajaremos con una familia exponencial que generaliza a la familia considerada por Geman y Graffigne en [4] y Comets en [5], esto es, supondremos que estas funciones tienen la forma

(H.10) $H_\Lambda^\vartheta(x) = -H_{\Lambda,0}(x) - \vartheta^T H_\Lambda(x) \quad \forall x \in E^{\mathcal{S}} \quad \forall \Lambda \in \mathfrak{S} \quad \forall \vartheta \in \Upsilon$,donde el vector de energía $H_\Lambda = (H_{\Lambda,1}, H_{\Lambda,2}, \dots, H_{\Lambda,m})$ es una función \mathcal{F}_{Λ^*} -medible definida de $E^{\mathcal{S}}$ en \mathbb{R}^m que no depende de ϑ ; y $H_{\Lambda,0}$ es una función \mathcal{F}_{Λ^*} -medible definida de $E^{\mathcal{S}}$ en \mathbb{R} que tampoco depende de ϑ . Se dice que generaliza a la familia considerada en [4, 5], porque en estos trabajos $H_{\Lambda,0} \equiv 0 \quad \forall \Lambda \in \mathfrak{S}$.Para cada $\Lambda \in \mathfrak{S}$ definimos las funciones densidades $\rho_\Lambda^\vartheta : E^{\mathcal{S}} \rightarrow (0, \infty)$ dadas por

$$\rho_\Lambda^\vartheta(x) := \frac{\exp\{-H_\Lambda^\vartheta(x)\}}{Z_\Lambda^\vartheta(x)} = \frac{h_\Lambda(x) \exp\{\vartheta^T H_\Lambda(x)\}}{Z_\Lambda^\vartheta(x)},$$

donde $h_\Lambda := \exp\{H_{\Lambda,0}\}$ y

$$Z_\Lambda^\vartheta(x) := \int_{E^\Lambda} \exp\{-H_\Lambda^\vartheta(\xi x^\Lambda)\} \lambda^\Lambda(d\xi).$$

Proposición 3.2.3 Dados $\vartheta \in \Upsilon$ y $\Lambda \in \mathfrak{S}$, se satisfacen:

- (i) Z_Λ^ϑ es $\mathcal{F}_{\partial\Lambda}$ -medible.
- (ii) ρ_Λ^ϑ es \mathcal{F}_{Λ^*} -medible;

Dem. Ver Sección 6.1 del Apéndice. □

Definición 3.2.2 Una especificación es una familia $\gamma = \{\gamma_\Lambda\}_{\Lambda \in \mathfrak{S}}$ de funciones $\gamma_\Lambda : \mathcal{F} \times E^{\mathcal{S}} \rightarrow [0, 1]$ tales que para cada $\Lambda \in \mathfrak{S}$ se verifican

- (i) Para cada $x \in E^{\mathcal{S}}$, la función $\gamma_\Lambda(\cdot|x)$ es una probabilidad sobre $(E^{\mathcal{S}}, \mathcal{F})$;
- (ii) Para cada $A \in \mathcal{F}$, la función $\gamma_\Lambda(A|\cdot)$ es una función \mathcal{J}_Λ -medible;
- (iii) $B \in \mathcal{J}_\Lambda, x \in E^{\mathcal{S}} \Rightarrow \gamma_\Lambda(B|x) = 1_B(x)$;
- (iv) $\Delta \subseteq \Lambda, A \in \mathcal{F} \Rightarrow \gamma_\Lambda(A|\cdot) = \gamma_\Lambda \gamma_\Delta(A|\cdot) := \int \gamma_\Delta(A|w) \gamma_\Lambda(dw|\cdot)$.

Para cada $\vartheta \in \Upsilon$ consideremos la familia $\gamma^\vartheta = \{\gamma_\Lambda^\vartheta\}_{\Lambda \in \mathfrak{S}}$, donde para cada $\Lambda \in \mathfrak{S}$, γ_Λ^ϑ es una función de $\mathcal{F} \times E^{\mathcal{S}}$ en $[0, 1]$ definida por

$$\gamma_\Lambda^\vartheta(A|x) := \int_{E^\Lambda} 1_A(\xi x^\Lambda) \rho_\Lambda^\vartheta(\xi x^\Lambda) \lambda^\Lambda(d\xi). \quad (3.2.2)$$

Proposición 3.2.4 Para cada $\vartheta \in \Upsilon$, la familia $\gamma^\vartheta = \{\gamma_\Lambda^\vartheta\}_{\Lambda \in \mathfrak{S}}$ es una especificación.

Dem. Ver Sección 6.1 del Apéndice. □

Esta familia se la conoce como la especificación asociada al potencial Φ^ϑ . Además, se dice que ρ^ϑ es una λ -modificación y que γ^ϑ es una λ -especificación a la que en [27] se la nota como $\rho^\vartheta \lambda_\bullet$.

Por (H.1) y por ser λ la medida que cuenta, vale la siguiente propiedad de gran utilidad:

Proposición 3.2.5 $\rho_\Lambda^\vartheta(x) = \gamma_\Lambda^\vartheta(\sigma_\Lambda^{-1}(\{x_\Lambda\})|x) \quad \forall x \in E^{\mathcal{S}} \quad \forall \Lambda \in \mathfrak{S} \quad \forall \vartheta \in \Upsilon$.

Dem. Corresponde a la Observación 1.28(5) de [27]. □

Definición 3.2.3 Dada una función $f : E^{\mathcal{S}} \rightarrow \mathbb{R}^k$, $k \in \mathbb{N}$, \mathcal{F} -medible, se define la función $\gamma_\Lambda^\vartheta(f|\cdot) : E^{\mathcal{S}} \rightarrow \mathbb{R}^k$ dada por

$$\gamma_\Lambda^\vartheta(f|x) := \int f(w) \gamma_\Lambda^\vartheta(dw|x).$$

Dada una probabilidad π sobre el espacio medible $(E^{\mathcal{S}}, \mathcal{F})$, definimos la medida

$$\pi \gamma_\Lambda^\vartheta(A) := \int \gamma_\Lambda^\vartheta(A|x) \pi(dx).$$

Además, usaremos la siguiente notación

$$\pi(f) := \int f d\pi,$$

que representa lo que se conoce como esperanza matemática.

Proposición 3.2.6 Dada $\pi \in \mathcal{P}(E^{\mathcal{S}}, \mathcal{F})$ y $\Lambda, \Delta \subseteq \mathcal{S}$, se satisfacen:

- (i) $f \in \mathcal{L}^1(E^{\mathcal{S}}, \mathcal{F}_\Lambda, \pi, \mathbb{R})$ y $g \in \mathcal{L}^1(E^{\mathcal{S}}, \mathcal{F}_\Delta, \pi, \mathbb{R})$, con $\Lambda \cap \Delta = \emptyset \Rightarrow \pi(fg) = \pi(f)\pi(g)$.
- (ii) f y g \mathcal{F}_Λ -medibles con $\Lambda \in \mathfrak{S}$ y $\pi(\{x \in E^{\mathcal{S}} : f(x) = g(x)\}) = 1 \Rightarrow f = g$.

Dem. Ver Sección 6.1 del Apéndice. □

Si en la condición (i) se agrega la hipótesis de que $\Lambda, \Delta \subseteq \mathfrak{S}$, entonces por (H.1) la integrabilidad se reduce a probar que f y g son \mathcal{F}_Λ y \mathcal{F}_Δ -medibles respectivamente.

Proposición 3.2.7 *Dados $\vartheta \in \Upsilon$, $\Lambda \in \mathfrak{S}$ y $\Delta \subseteq \mathcal{S}$ se satisfacen:*

(i) $A \in \mathcal{F}_\Delta \Rightarrow \gamma_\Lambda^\vartheta(A|\cdot)$ es $\mathcal{F}_{\partial\Lambda \cup (\Delta \setminus \Lambda)}$ -medible;

(ii) $f \in \mathcal{L}^1(E^{\mathcal{S}}, \mathcal{F}, \pi^\vartheta, \mathbb{R}) \Rightarrow \gamma_\Lambda^\vartheta(f|x) = \int_{E^\Lambda} f(\xi x^\Lambda) \rho_\Lambda^\vartheta(\xi x^\Lambda) \lambda^\Lambda(d\xi)$;

(iii) f una función \mathcal{F}_Δ -medible $\Rightarrow \gamma_\Lambda^\vartheta(f|\cdot)$ es $\mathcal{F}_{\partial\Lambda \cup (\Delta \setminus \Lambda)}$ -medible;

(iv) $\pi \in \mathcal{P}(E^{\mathcal{S}}, \mathcal{F}) \Rightarrow \pi \gamma_\Lambda^\vartheta \in \mathcal{P}(E^{\mathcal{S}}, \mathcal{F})$.

Dem. Ver Sección 6.1 del Apéndice. □

Definición 3.2.4 *Dada una especificación γ , definimos el conjunto de las medidas de Gibbs asociadas a γ como*

$$\mathcal{G}(\gamma) := \{\pi \in \mathcal{P}(E^{\mathcal{S}}, \mathcal{F}) : \pi \gamma_\Lambda = \pi \forall \Lambda \in \mathfrak{S}\},$$

Bajo las hipótesis hechas verifica lo siguiente:

Proposición 3.2.8 *Para cada $\vartheta \in \Upsilon$, $\mathcal{G}(\gamma^\vartheta)$ es convexo y no vacío.*

Dem. Ver Sección 6.1 del Apéndice. □

Luego para cada $\vartheta \in \Upsilon$, tomemos arbitrariamente

$$(H.11) \quad \pi^\vartheta \in \text{ex}(\mathcal{G}(\gamma^\vartheta)) := \{\pi \in \mathcal{G}(\gamma^\vartheta) : \pi \neq \alpha\pi_1 + (1-\alpha)\pi_2 \forall \alpha \in (0,1) \forall \pi_1, \pi_2 \in \mathcal{G}(\gamma^\vartheta)\},$$

esto es, π^ϑ es un elemento extremal de $\mathcal{G}(\gamma^\vartheta)$.

Una propiedad importante de analizar es la “cercanía” entre $\gamma_\Lambda^\vartheta(\cdot|x)$ y π^ϑ cuando Λ es lo suficientemente grande. De igual modo, interesa analizar la relación asintótica entre $\gamma_\Lambda^\vartheta(f|x)$ y $\pi^\vartheta(f)$, para una determinada función f . Para ello se necesita del siguiente concepto

Definición 3.2.5 *Una familia $\mathfrak{C} \subseteq \mathfrak{S}$ se dice cofinal, si para cada $\Lambda \in \mathfrak{S}$, existe $\Delta \in \mathfrak{C}$, tal que $\Lambda \subseteq \Delta$.*

Consideremos la familia cofinal de subconjunto de \mathcal{S} dada por

$$(H.12) \quad \Lambda_n = ([-n, n] \cap \mathbb{Z})^d \quad \forall n \in \mathbb{N}.$$

Bajo el supuesto (H.11) se verifica que

Teorema 3.2.1 *Se verifican*

$$(i) f \in \mathcal{L}^\infty(E^S, \mathcal{F}, \pi^\vartheta, \mathbb{R}) \Rightarrow \pi^\vartheta \left(\left\{ x \in E^S : \lim_{n \rightarrow \infty} \gamma_\Lambda^\vartheta(f|x) = \pi^\vartheta(f) \right\} \right) = 1;$$

$$(ii) A \in \mathcal{F} \Rightarrow \pi^\vartheta \left(\left\{ x \in E^S : \lim_{n \rightarrow \infty} \gamma_{\Lambda_n}^\vartheta(A|x) = \pi^\vartheta(A) \right\} \right) = 1.$$

Dem. Ver Sección 6.1 del Apéndice. □

3.3. Esperanza Condicional.

En la presente sección se mostrará el vínculo que existe entre las especificaciones y las probabilidades condicionales.

Definición 3.3.1 *Dados un espacio de probabilidad $(\Omega, \mathcal{F}, \pi)$, una σ -álgebra $\mathcal{A} \subset \mathcal{F}$ y $f \in \mathcal{L}^1(\Omega, \mathcal{F}, \pi, \mathbb{R})$, llamaremos esperanza condicional de f dada \mathcal{A} , al conjunto*

$$\pi(f|\mathcal{A}) := \left\{ \tilde{f} \in \mathcal{L}^1(\Omega, \mathcal{F}, \pi, \mathbb{R}) : \tilde{f} \text{ es } \mathcal{A}\text{-medible y } \int_A \tilde{f} d\pi = \int_A f d\pi \forall A \in \mathcal{A} \right\}.$$

Por el Teorema de Radon-Nikodym la esperanza condicional nunca es vacía y está bien definida. Además, dadas dos versiones $\tilde{f}_1, \tilde{f}_2 \in \pi(f|\mathcal{A})$ de la esperanza condicional, se verifica que $\tilde{f}_1 = \tilde{f}_2$ π -p.p.

Proposición 3.3.1 *Si $f \in \mathcal{L}^1(E^S, \mathcal{F}_\Lambda, \pi^\vartheta, \mathbb{R})$, $\Lambda \in \mathfrak{S}$ y $\vartheta \in \Upsilon$. Entonces $\gamma_\Lambda^\vartheta(f|\cdot) \in \pi^\vartheta(f|\mathcal{F}_{\partial\Lambda})$.*

Dem. Ver Sección 6.1 del Apéndice. □

Ahora veamos la relación que existe entre las especificaciones y las probabilidades condicionales. Teniendo en cuenta la Proposición 3.3.1, definamos la función $\pi_\Lambda^\vartheta : \mathcal{E}^\Lambda \times E^{S \setminus \Lambda} \rightarrow [0, 1]$ dada por

$$\pi_\Lambda^\vartheta(A|y) := \gamma_\Lambda^\vartheta(\sigma_\Lambda^{-1}(A)|x),$$

siendo $x \in E^S$ tal que $\sigma_{S \setminus \Lambda}(x) = y$. Esta función es la probabilidad condicional de observar A en \mathcal{E}^Λ , dado que se observó $y = x^\Lambda$ sobre $E^{S \setminus \Lambda}$. Si $A = \{z\}$, entonces simplemente se notará con $\pi_\Lambda^\vartheta(z|y)$, y representa la densidad condicional de observar z sobre E^Λ , dado que se observó y sobre $E^{S \setminus \Lambda}$.

Proposición 3.3.2 $\pi_\Lambda^\vartheta(x_\Lambda|x^\Lambda) = \rho_\Lambda^\vartheta(x) \forall x \in E^S$.

Dem. Ver Sección 6.1 del Apéndice. □

3.4. Invarianza por Traslaciones.

En la presente sección analizaremos bajo qué condiciones en el modelo exponencial (H.10) las especificaciones son invariantes por traslaciones. Este concepto no es equivalente al concepto de estacionaridad o invarianza de la medida de Gibbs, ya que una medida de Gibbs puede no ser invariante a pesar de que la especificación que la genere sea invariante. Como veremos en la Sección 6.2 del Apéndice, en caso de que $\#\mathcal{G}(\gamma^\vartheta) = 1$, se tiene que π^ϑ es estacionaria. En el presente trabajo, cuando hablemos de invarianza lo haremos respecto de las especificaciones, y para su correspondiente medida de Gibbs sólo supondremos (H.11).

Definición 3.4.1 Dado $s \in \mathcal{S}$, consideremos las funciones de traslación $\theta_s : \mathcal{S} \rightarrow \mathcal{S}$ dada por $\theta_s(t) = t - s$; y $\theta_s : E^{\mathcal{S}} \rightarrow E^{\mathcal{S}}$ dada por $\theta_s(x) = x \circ \theta_s$. Además, dada una especificación $\gamma = \{\gamma_\Lambda\}_{\Lambda \in \mathcal{G}}$ se define una nueva especificación $\theta_s(\gamma)$ dada por

$$\theta_s(\gamma)_\Lambda(A|x) = \gamma_{\Lambda-s}(\theta_s^{-1}(A)|\theta_s^{-1}(x)) \quad \forall x \in E^{\mathcal{S}} \quad \forall A \in \mathcal{F} \quad \forall \Lambda \in \mathcal{G}.$$

Por último, dada una familia de funciones $f = \{f_\Lambda\}_{\Lambda \in \mathcal{G}}$ de $E^{\mathcal{S}}$ en \mathbb{R} , se nota con $\theta_s(f) = \{\theta_s(f)_\Lambda\}_{\Lambda \in \mathcal{G}}$ a la familia de funciones dadas por

$$\theta_s(f)_\Lambda(x) = (f_{\Lambda-s} \circ \theta_s^{-1})(x) = f_{\Lambda-s}(\theta_s^{-1}(x)) \quad \forall x \in E^{\mathcal{S}} \quad \forall \Lambda \in \mathcal{G}.$$

Las traslaciones son un caso particular de las simetrías que serán tratadas en la Sección 6.2 del Apéndice.

Definición 3.4.2 Se dice que una familia de funciones $f = \{f_\Lambda\}_{\Lambda \in \mathcal{G}}$ es invariante por traslaciones o simplemente invariante si

$$\theta_s(f) = f \quad \forall s \in \mathcal{S},$$

y que una especificación $\gamma = \{\gamma_\Lambda\}_{\Lambda \in \mathcal{G}}$ es invariante por traslaciones o simplemente invariante si

$$\theta_s(\gamma) = \gamma \quad \forall s \in \mathcal{S}.$$

Proposición 3.4.1 Para cada $\vartheta \in \Upsilon$, las siguientes condiciones son equivalentes

- (i) $\gamma^\vartheta = \{\gamma_\Lambda^\vartheta\}_{\Lambda \in \mathcal{G}}$ es invariante por traslaciones;
- (ii) $\rho^\vartheta = \{\rho_\Lambda^\vartheta\}_{\Lambda \in \mathcal{G}}$ es invariante por traslaciones.

Dem. Corresponde al Corolario 5.9(a) de [27], teniendo en cuenta que λ es la medida que cuenta, (H.1) y que por notación $\gamma^\vartheta = \rho^\vartheta \lambda_\bullet$. □

El siguiente resultado muestra la relación existente entre el modelo exponencial supuesto en (H.10) y la invarianza de su correspondiente especificación.

Proposición 3.4.2 Dado $\vartheta \in \Upsilon$, las siguientes condiciones son equivalentes:

- (i) γ^ϑ es invariante;
- (ii) se verifican las siguientes condiciones:

$$(H.13) \quad H_{\Lambda+s} \circ \theta_s = H_\Lambda \quad \forall \Lambda \in \mathfrak{S} \forall s \in \mathcal{S};$$

$$(H.14) \quad h_{\Lambda+s} \circ \theta_s = h_\Lambda \quad \forall \Lambda \in \mathfrak{S} \forall s \in \mathcal{S}.$$

Dem. Ver Sección 6.1 del Apéndice. □

3.5. Modelo de Potts

En la presente sección se definirán los modelos de Potts [2], los cuales son una clase particular de modelos Gibbsianos. Bajo el supuesto de que el mapa de clases se rige por dicho modelo, incluyen inteligentemente información contextual a los procedimientos de segmentación. En este contexto, se mostrarán dos ejemplos de modelos que motivaron la tesis, de los cuales uno contiene especificaciones invariantes y el otro no.

Sigamos considerando el conjunto finito de todas las posibles etiquetas $E = \{\ell_1, \ell_2, \dots, \ell_L\}$, y fijemos un sistema de vecindades $\partial = \{\partial_s\}_{s \in \mathcal{S}}$. Un modelo de Potts (o modelos de Potts-Strauss) es un modelo en el que los potenciales tienen la forma

$$\Phi_\Lambda(x) = \begin{cases} -\alpha_\ell & \text{si } \Lambda = \{s\} \text{ y } x_s = \ell; \\ -\beta_{\ell, \tilde{\ell}} & \text{si } \Lambda = \{s, t\} \text{ y } x_s = \ell, x_t = \tilde{\ell}, \text{ con } s \in \partial_t; \\ 0 & \text{en otro caso.} \end{cases}$$

Este modelo es el más general de la familia, y es conocido como modelo de Potts anisotrópico con campo externo. Si $\alpha_\ell = 0$ para cada $\ell \in E$, entonces se dice que el modelo es sin campo externo, y si $\beta_{\ell, \tilde{\ell}} \equiv \beta$ para todo $\ell, \tilde{\ell} \in E$, entonces se dice que el modelo es isotrópico.

Ejemplo 3.5.1 (*Modelo a Priori: modelo de Potts isotrópico y sin campo externo*). Existen diversos modelos Markovianos que permiten modelar a priori el mapa de clases de una imagen. Se llama modelo a priori, porque es el modelo que se le otorga al mapa de clases antes de contar con la evidencia radiométrica. Entre ellos se encuentra el modelo de Potts isotrópico y sin campo externo, en el que el potencial viene dado por las funciones

$$\Phi_\Lambda^\beta(x) = \begin{cases} -\beta & \text{si } \Lambda = \{s, t\} \text{ y } x_s = x_t, \text{ con } s \in \partial_t; \\ 0 & \text{en otro caso.} \end{cases} \quad (3.5.3)$$

donde $\vartheta = \beta \in \Upsilon = \mathbb{R}$ es un parámetro real llamado temperatura inversa que mide el grado de vínculo existente entre las clases correspondientes a dos píxeles vecinos. Luego la función de energía del potencial en $\Lambda \in \mathfrak{S}$ viene dada por

$$H_\Lambda^\beta(x) := \sum_{\{\{s, t\} \in \mathfrak{S}(\Lambda)\}} \Phi_{\{s, t\}}(\{x_s, x_t\}) = -\beta H_\Lambda(x),$$

donde $H_\Lambda(x) = \#\{\Delta \in \mathfrak{S}(\Lambda) : \#\Delta = 2, \Delta = \{s, t\}, x_s = x_t, t \in \partial_s\}$ es la cantidad de pares de vecinos (sin importar el orden) en Λ^* con la misma etiqueta, con al menos uno de los sitios en Λ . Luego la densidad condicional del campo está dada por

$$\pi_\Lambda^\beta(x_\Lambda | x^\Lambda) \propto \exp\{\beta H_\Lambda(x)\}. \quad (3.5.4)$$

Por la Proposición 3.2.2, cada función H_s dada por $H_s(x) = \#\{t \in \partial_s : x_s = x_t\}$ es $\mathcal{F}_{s \cup \partial_s}$ -medible, y por lo tanto $H_s(x)$ depende sólo de $x_{s \cup \partial_s}$. Cuando las funciones $H_s(x)$ aparecen como parte de la definición de las densidades condicionales, esto es

$$\pi_s^\beta(x_s | x^s) \propto \exp\{\beta H_s(x)\},$$

sólo dependen de x_s , pues la configuración fuera de s , esto es $x^s \in E^{\mathcal{S} \setminus \{s\}}$, ya está prefijada. Luego, por simplicidad notacional, y siempre que no dé lugar a confusión, cuando trabajemos con el modelos de Potts, la función H_s se verá como una función de E en $\mathbb{N} \cup \{0\}$. De este modo, las densidades condicionales sobre $s \in \Lambda$ del modelo a priori se reescriben como

$$\pi_s^\beta(x_s|x^s) \propto \exp\{\beta H_s(x_s)\}.$$

En este ejemplo $h_\Lambda \equiv 1$ y $H_\Lambda(x) = \#\{\Delta \in \mathfrak{S}(\Lambda) : \#\Delta = 2, \Delta = \{s, t\}, x_s = x_t, t \in \partial s\}$. De las condiciones de invarianza, claramente vale (H.14), pero debemos verificar (H.13):

$$\begin{aligned} (H_{\Lambda+r} \circ \theta_r)(x) &= \#\{\Delta \in \mathfrak{S}(\Lambda+r) : \#\Delta = 2, \Delta = \{s, t\}, (\theta_r(x))_s = (\theta_r(x))_t, t \in \partial s\} = \\ &= \#\{\Delta - r \in \mathfrak{S}(\Lambda) : \#(\Delta - r) = 2, \Delta - r = \{s-r, t-r\}, x_{s-r} = x_{t-r}, t-r \in \partial s-r\} = \\ &= \#\{\Delta \in \mathfrak{S}(\Lambda) : \#\Delta = 2, \Delta = \{s, t\}, x_s = x_t, t \in \partial s\} = H_\Lambda(x). \end{aligned}$$

Además, las densidades condicionales de observar la configuración de clases $x_\Lambda \in E^\Lambda$ dado que se observó $x_{\partial\Lambda} \in E^{\partial\Lambda}$ vienen dada por $\pi_\Lambda^\beta(x_\Lambda|x_{\partial\Lambda}) \propto \exp\{\beta H_\Lambda(x)\}$.

Ejemplo 3.5.2 (Modelo a Posteriori) Supongamos que contamos con una familia de distribuciones $\{\pi^\beta\}_{\beta \in \mathbb{R}}$ donde π^β es una distribución de Gibbs extremal generada por el potencial Φ^β definido en el Ejemplo anterior.

Dado un nuevo Espacio de Probabilidad (Ω, \mathcal{A}, P) , si $X : \Omega \rightarrow E^{\mathcal{S}}$ es tal que

$$\{\omega \in \Omega : X(\omega) \in B\} \in \mathcal{A} \quad \forall B \in \mathcal{E}^{\mathcal{S}},$$

y si existe una probabilidad π sobre $(E^{\mathcal{S}}, \mathcal{E}^{\mathcal{S}})$ tal que

$$P(\{\omega \in \Omega : X(\omega) \in B\}) = \pi(B) \quad \forall B \in E^{\mathcal{S}},$$

entonces diremos que el campo aleatorio X tiene distribución π .

Sea X es el campo aleatorio que asigna una etiqueta a cada pixel, y supongamos que su distribución es π^β (modelo a priori del mapa de clases). Además, supongamos que cada clase $\ell \in E$ emite observaciones radiométricas en \mathbb{R}^B , $B \in \mathbb{N}$ fijo, de acuerdo a una densidad f_ℓ . Esto es, si para cada $s \in \mathcal{S}$, la variable $I_s : \Omega \rightarrow \mathbb{R}^B$ representa el dato radiométrico que se observa sobre el sitio s , entonces

$$P(I_s = I_s | X_s = \ell) = f_\ell(I_s) \quad \forall \ell \in E,$$

donde $P(I_s = \cdot | X_s = \ell)$ representa por simplicidad notacional a la densidad condicional con la que la clase ℓ emite observaciones. Sea $\mathbf{I} : \Omega \rightarrow (\mathbb{R}^B)^{\mathcal{S}}$ el campo aleatorio dado por $\mathbf{I}(\omega) = (I_s(\omega))_{s \in \mathcal{S}}$ que llamaremos datos radiométricos. Supongamos que las observaciones radiométricas sobre cada $\Lambda \in \mathfrak{S}$ son independientes dado el mapa de clases sobre Λ , esto es

$$P(\mathbf{I}_\Lambda = I_\Lambda | X_\Lambda = x_\Lambda) = \prod_{s \in \Lambda} P(I_s = I_s | X_s = x_s) \quad \forall I_\Lambda \in (\mathbb{R}^B)^\Lambda \quad \forall x_\Lambda \in E^\Lambda \quad \forall \Lambda \in \mathfrak{S},$$

y supongamos que dado el mapa de clases sobre Λ , el mapa de clases sobre $\partial\Lambda$ y las observaciones radiométricas sobre Λ son independientes, esto es,

$$P(X_{\partial\Lambda} = x_{\partial\Lambda}, \mathbf{I}_\Lambda = I_\Lambda | X_\Lambda = x_\Lambda) = P(X_{\partial\Lambda} = x_{\partial\Lambda} | X_\Lambda = x_\Lambda) P(\mathbf{I}_\Lambda = I_\Lambda | X_\Lambda = x_\Lambda) \quad \forall I_\Lambda \in (\mathbb{R}^B)^\Lambda \quad \forall x_\Lambda \in E^\Lambda \quad \forall \Lambda \in \mathfrak{S}.$$

Luego bajo estos supuestos, se prueba que

$$P(\mathbf{I}_\Lambda = I_\Lambda | X_\Lambda = x_\Lambda) = \prod_{s \in \Lambda} P(I_s = I_s | X_\Lambda = x_\Lambda) = \prod_{s \in \Lambda} P(I_s = I_s | X_s = x_s, X_{\Lambda \setminus \{s\}} = x_{\Lambda \setminus \{s\}}) =$$

$$= \prod_{s \in \Lambda} P(I_s = I_s | X_s = x_s) \quad \forall \Lambda \in \mathfrak{S}.$$

Esta identidad sobre el modelo de emisión coincide con la asunción 1 realizada por Besag en su trabajo [14], la cual es una de las hipótesis en las que se basa el clasificador contextual ICM que será tratado con más detalle en la Sección 5.3.

Suponga que se desea calcular la probabilidad de observar la configuración de clases $x_\Lambda \in E^\Lambda$, dado que se observaron los datos radiométricos $I_\Lambda \in (\mathbb{R}^B)^\Lambda$ y la configuración de clases $x_{\partial\Lambda} \in E^{\partial\Lambda}$ (o equivalentemente, por (H.9), sobre $x^\Lambda \in E^{S \setminus \Lambda}$). Esta es la densidad condicional a posteriori del mapa de clases sobre E^Λ , y será denotada con Π_Λ^β . Luego

$$\begin{aligned} \Pi_\Lambda^\beta(x_\Lambda | x_{\partial\Lambda}, I_\Lambda) &:= P(X_\Lambda = x_\Lambda | X_{\partial\Lambda} = x_{\partial\Lambda}, I_\Lambda = I_\Lambda) = \frac{P(X_\Lambda = x_\Lambda, X_{\partial\Lambda} = x_{\partial\Lambda}, I_\Lambda = I_\Lambda)}{P(X_{\partial\Lambda} = x_{\partial\Lambda}, I_\Lambda = I_\Lambda)} = \\ &= \frac{P(I_\Lambda = I_\Lambda | X_\Lambda = x_\Lambda, X_{\partial\Lambda} = x_{\partial\Lambda}) P(X_\Lambda = x_\Lambda | X_{\partial\Lambda} = x_{\partial\Lambda}) P(X_{\partial\Lambda} = x_{\partial\Lambda})}{P(I_\Lambda = I_\Lambda | X_{\partial\Lambda} = x_{\partial\Lambda}) P(X_{\partial\Lambda} = x_{\partial\Lambda})} = \\ &= \frac{P(I_\Lambda = I_\Lambda | X_\Lambda = x_\Lambda) P(X_\Lambda = x_\Lambda | X_{\partial\Lambda} = x_{\partial\Lambda})}{P(I_\Lambda = I_\Lambda | X_{\partial\Lambda} = x_{\partial\Lambda})} \propto P(I_\Lambda = I_\Lambda | X_\Lambda = x_\Lambda) P(X_\Lambda = x_\Lambda | X_{\partial\Lambda} = x_{\partial\Lambda}) = \\ &= \prod_{s \in \Lambda} P(I_s = I_s | X_s = x_s) P(X_\Lambda = x_\Lambda | X_{\partial\Lambda} = x_{\partial\Lambda}) = \prod_{s \in \Lambda} f_{x_s}(I_s) \pi_\Lambda^\beta(x_\Lambda | x_{\partial\Lambda}) = \\ &= \exp \left\{ \sum_{s \in \Lambda} \ln f_{x_s}(I_s) + \beta H_\Lambda(x) \right\}, \end{aligned}$$

la cual corresponde a la densidad condicional generada por el potencial $\Phi^{\beta, I}$ conformado por las funciones

$$\Phi_\Lambda^{\beta, I}(x) := \begin{cases} -\ln f_{x_s}(I_s) & \text{si } \Lambda = \{s\}; \\ -\beta & \text{si } \Lambda = \{s, t\}, x_s = x_t, t \in \partial s; \\ 0 & \text{caso contrario.} \end{cases}$$

Este es un modelo de Potts isotrópico con campo externo dado por $(\ln f_{x_s}(I_s))_{s \in S}$. Además, las funciones de energía tienen la forma

$$-H_\Lambda^{\beta, I}(x) = \ln h_\Lambda(x) + \beta H_\Lambda(x),$$

donde

$$h_\Lambda(x) = \prod_{s \in \Lambda} f_{x_s}(I_s)$$

y

$$H_\Lambda(x) = \#\{\Delta \in \mathfrak{S}(\Lambda) : \#\Delta = 2, \Delta = \{s, t\}, x_s = x_t, t \in \partial s\}.$$

Luego se verifica la condición de invarianza (H.13) tal y como se probó en el Ejemplo anterior, pero veamos que no se verifica (H.14). En particular $h_s(x) = f_{x_s}(I_s)$, y por ende

$$(h_{s+t} \circ \mathbf{0}_t)(x) = h_{s+t}(\mathbf{0}_t(x)) = f_{\mathbf{0}_t(x)_{s+t}}(I_{s+t}) = f_{x_s}(I_{s+t}).$$

Luego para que se verifique (H.14) es necesario que $f_{x_s}(I_s) = f_{x_s}(I_{s+t}) \forall s, t \in S$, lo cual vale sólo si

$$f_\ell(I_s) = f_\ell(I_t) \quad \forall s, t \in S \quad \forall \ell \in E,$$

lo cual no vale en general.

3.6. Ley de los Grandes Números.

Dados $\pi \in \mathcal{P}(E^{\mathcal{S}}, \mathcal{F})$, Λ y Δ dos subconjuntos disjuntos de \mathcal{S} , $f \in \mathcal{L}^1(E^{\mathcal{S}}, \mathcal{F}_{\Lambda}, \pi, \mathbb{R})$ y $g \in \mathcal{L}^1(E^{\mathcal{S}}, \mathcal{F}_{\Delta}, \pi, \mathbb{R})$, en la Proposición 3.2.6(i) se mostró que $\pi(fg) = \pi(f)\pi(g)$. Hablando informalmente, esto puede ser visto como una no correlación entre f y g . Sin embargo, $\Lambda \cap \Delta = \emptyset$ implica también, usando la Proposición 3.2.1(ii), que conocer el valor de f no significa un cambio en la probabilidad de que g tome algún valor. Por tal motivo intuitivo adaptaremos al contexto del problema analizado y con la notación del trabajo, la prueba general de la Ley de los Grandes Números para variables independientes no idénticamente distribuidas detallada en las notas de clase de Yohai [34]. Este resultado permitirá extender la clase de modelos exponenciales invariantes para los cuales vale la prueba de consistencia dada por Geman y Graffigne en [4]. Además, es fundamental en la prueba de la consistencia del estimador de PMV sobre modelos con especificaciones no invariantes.

Comencemos probando los siguientes dos Lemas auxiliares, los cuales son análogos a los resultados auxiliares probados en [34].

Lema 3.6.1 *Sea $\pi \in \mathcal{P}(E^{\mathcal{S}}, \mathcal{F})$, sea $\{f_n\}_{n \in \mathbb{N}}$ una sucesión de funciones \mathcal{F} -medibles y sea f una función \mathcal{F} -medible. Entonces las siguientes condiciones son equivalentes:*

$$(i) \quad \pi(\{x \in E^{\mathcal{S}} : f_n(x) \rightarrow f(x)\}) = 1;$$

$$(ii) \quad \lim_{m \rightarrow \infty} \pi\left(\bigcup_{n=m}^{\infty} B_{n,\varepsilon}^c\right) = 0 \quad \forall \varepsilon > 0, \text{ donde } B_{n,\varepsilon} := \{x \in E^{\mathcal{S}} : |f_n(x) - f(x)| < \varepsilon\}.$$

Dem. Ver Sección 6.1 del Apéndice. □

Lema 3.6.2 *Sea $\pi \in \mathcal{P}(E^{\mathcal{S}}, \mathcal{F})$, sea $\{f_n\}_{n \in \mathbb{N}}$ una sucesión de funciones \mathcal{F} -medibles y sea f una función \mathcal{F} -medible, tales que para cada $\varepsilon > 0$, existe una sucesión creciente $\{r_n\}_{n \in \mathbb{N}}$ de números reales positivos que verifican*

$$\sum_{i=1}^{\infty} \pi\left(\bigcup_{n=r_i}^{r_{i+1}-1} B_{n,\varepsilon}^c\right) < \infty,$$

donde $B_{n,\varepsilon} := \{x \in E^{\mathcal{S}} : |f_n(x) - f(x)| < \varepsilon\}$. Entonces

$$\pi(\{x \in E^{\mathcal{S}} : f_n(x) \rightarrow f(x)\}) = 1.$$

Dem. Ver Sección 6.1 del Apéndice. □

La siguiente desigualdad es una nueva versión de la Desigualdad de Kolmogorov. En la misma se reemplaza la condición de independencia por la de medibilidad sobre conjuntos de sitios disjuntos.

Lema 3.6.3 *Sean $\pi \in \mathcal{P}(E^{\mathcal{S}}, \mathcal{F})$, $\varepsilon > 0$, $n \in \mathbb{N}$, $\Delta_1, \dots, \Delta_n$ subconjuntos disjuntos de \mathcal{S} , y para cada $1 \leq j \leq n$, sea f_j una función real \mathcal{F}_{Δ_j} -medible con $\pi^{\emptyset}(f_j) = 0$ y $\pi^{\emptyset}(f_j^2) < \infty$. Sea también para cada $1 \leq i \leq n$ la función $S_i = \sum_{j=1}^i f_j(x)$. Entonces*

$$\pi\left(\left\{x \in E^{\mathcal{S}} : \max_{1 \leq i \leq n} |S_i(x)| \geq \varepsilon\right\}\right) \leq \frac{1}{\varepsilon^2} \sum_{i=1}^n \pi(f_i^2).$$

Dem. Ver Sección 6.1 del Apéndice. □

La siguiente Ley de los grandes números es un nuevo punto de vista de una versión clásica de la misma. Nuevamente se reemplaza la condición de independencia por la de medibilidad sobre conjuntos de sitios disjuntos.

Teorema 3.6.1 Sean $\pi \in \mathcal{P}(E^{\mathcal{S}}, \mathcal{F})$, $\{\Delta_n\}_{n \in \mathbb{N}}$ una sucesión de subconjuntos disjuntos de \mathcal{S} , y para cada $n \in \mathbb{N}$, sea f_n una función real \mathcal{F}_{Δ_n} -medible con $\pi(f_n) = 0$. Si

$$\sum_{n=1}^{\infty} \frac{\pi(f_n^2)}{n^2} < \infty. \quad (3.6.5)$$

Entonces

$$\pi \left(\left\{ x \in E^{\mathcal{S}} : \frac{1}{n} \sum_{i=1}^n f_i(x) \rightarrow 0 \right\} \right) = 1.$$

Dem. Sean $\varepsilon > 0$, $r_i = 2^{i-1}$ y $a_i = \pi \left(\bigcup_{n=2^{i-1}}^{2^i-1} B_{n,\varepsilon}^c \right)$, donde $B_{n,\varepsilon} = \{x \in E^{\mathcal{S}} : |\bar{f}_n(x)| < \varepsilon\}$, $\bar{f}_n(x) := \frac{1}{n} S_n(x)$ y $S_n(x) := \sum_{i=1}^n f_i(x)$ para cada $n \in \mathbb{N}$. Luego por el Lema 3.6.3, sólo se debe probar que $\sum_{i=1}^{\infty} a_i < \infty$.

$$\begin{aligned} a_i &= \pi \left(\bigcup_{n=2^{i-1}}^{2^i-1} \{x \in E^{\mathcal{S}} : |\bar{f}_n(x)| \geq \varepsilon\} \right) = \pi \left(\bigcup_{n=2^{i-1}}^{2^i-1} \{x \in E^{\mathcal{S}} : |S_n(x)| \geq n\varepsilon\} \right) \leq \\ &\leq \pi \left(\bigcup_{n=2^{i-1}}^{2^i-1} \{x \in E^{\mathcal{S}} : |S_n(x)| \geq 2^{i-1}\varepsilon\} \right) \leq \pi \left(\bigcup_{n=1}^{2^i-1} \{x \in E^{\mathcal{S}} : |S_n(x)| \geq 2^{i-1}\varepsilon\} \right) = \\ &= \pi \left(\left\{ x \in E^{\mathcal{S}} : \max_{1 \leq n \leq 2^{i-1}} |S_n(x)| \geq 2^{i-1}\varepsilon \right\} \right). \end{aligned}$$

Luego por el Lema 3.6.3,

$$\sum_{i=1}^{\infty} a_i \leq \sum_{i=1}^{\infty} \left(\frac{1}{4^{i-1}\varepsilon^2} \sum_{j=1}^{2^{i-1}} \pi(f_j^2) \right) = \frac{1}{\varepsilon^2} \sum_{j=1}^{\infty} \left(\pi(f_j^2) \sum_{\{i: 2^{i-1} \geq j\}} \frac{1}{4^{i-1}} \right) = \frac{1}{\varepsilon^2} \sum_{j=1}^{\infty} \left(\pi(f_j^2) \sum_{i \geq \log_2(j+1)} \frac{1}{4^{i-1}} \right). \quad (3.6.6)$$

Si a_0 es el primer término de la serie geométrica $\sum_{i \geq \log_2(j+1)} \frac{1}{4^i}$, esto es, $a_0 = 4^{-\lceil \log_2(j+1) \rceil}$, donde $\lceil \log_2(j+1) \rceil$ representa al entero más chico que es mayor o igual que $\log_2(j+1)$. Entonces

$$\sum_{i \geq \log_2(j+1)} \frac{1}{4^{i-1}} = 4a_0 \frac{1}{1 - \frac{1}{4}} = \frac{16}{3} a_0.$$

Luego como, $i \geq \log_2(j+1) \geq \log_2(j) \Rightarrow 2^i \geq j \Rightarrow \frac{1}{4^i} \leq \frac{1}{j^2} \Rightarrow a_0 \leq \frac{1}{j^2}$, se tiene que por (3.6.6),

$$\sum_{i=1}^{\infty} a_i \leq \frac{16}{3\varepsilon^2} \sum_{j=1}^{\infty} \frac{\pi(f_j^2)}{j^2} < \infty,$$

lo cual finaliza la prueba. □

En muchos resultados se necesita de una Ley de los Grandes Números para funciones medibles sobre regiones no disjuntas de \mathcal{S} . En el siguiente resultado veremos que la ley continua valiendo si las regiones de medibilidad de las funciones que intervienen pueden subdividirse en una colección finita de familias infinitas de subconjuntos disjuntos de \mathcal{S} .

Teorema 3.6.2 Sean $\pi \in \mathcal{P}(E^{\mathcal{S}}, \mathcal{F})$, $\{\Delta_n\}_{n \in \mathbb{N}}$ una sucesión de subconjuntos de \mathcal{S} que puede ser particionada en k subfamilias infinitas de conjuntos disjuntos; y para cada $n \in \mathbb{N}$, sea f_n una función real \mathcal{F}_{Δ_n} -medible con $\pi(f_n) = 0$ y $\pi(\{x \in E^{\mathcal{S}} : |f_n(x)| \leq c\}) = 1$, con c una constante real positiva. Entonces

$$\pi \left(\left\{ x \in E^{\mathcal{S}} : \frac{1}{n} \sum_{i=1}^n f_i(x) \rightarrow 0 \right\} \right) = 1.$$

Dem. Sea $\{F_1, \dots, F_k\}$ la partición que existe por hipótesis. Luego, para cada $1 \leq i \leq k$, $F_i = \{\Delta_{i_j}\}_{j \in \mathbb{N}}$ y $\Delta_{i_{j_1}} \cap \Delta_{i_{j_2}} = \emptyset$ si $j_1 \neq j_2$. Como las funciones f_n son acotadas, se verifica (3.6.5) para cualquier subsucesión de $\{f_n\}_{n \in \mathbb{N}}$. Luego por el Teorema 3.6.1

$$\pi(A_i) = 1 \quad \forall 1 \leq i \leq k,$$

donde $A_i := \{x \in E^{\mathcal{S}} : \frac{1}{n} \sum_{j=1}^n f_{i_j}(x) \rightarrow 0\}$. Si $A := \bigcap_{i=1}^k A_i$, entonces $\pi(A) = 1$. Consideremos $x \in A$ y probemos que

$$\frac{1}{n} \sum_{i=1}^n f_i(x) \rightarrow 0.$$

Para cada $n \in \mathbb{N}$ y $1 \leq i \leq k$, sea $B_{n,i} := \{j : i_j \leq n\}$. Luego por una consecuencia del Teorema 3.6.1, se tiene que

$$\frac{1}{\#B_{n,i}} \sum_{j \in B_{n,i}} f_{i_j}(x) \rightarrow 0.$$

Por lo que, dado $\varepsilon > 0$, existen $n_1, \dots, n_k \in \mathbb{N}$ y $n_0 = \max_{1 \leq i \leq k} n_i$, tales que

$$\left| \frac{1}{\#B_{n,i}} \sum_{j \in B_{n,i}} f_{i_j}(x) \right| < \frac{\varepsilon}{k} \quad \forall 1 \leq i \leq k \quad \forall n \geq n_0.$$

Luego

$$\left| \frac{1}{n} \sum_{i=1}^n f_i(x) \right| = \left| \frac{1}{n} \sum_{i=1}^k \sum_{j \in B_{n,i}} f_{i_j}(x) \right| \leq \sum_{i=1}^k \left| \frac{1}{n} \sum_{j \in B_{n,i}} f_{i_j}(x) \right| \leq \sum_{i=1}^k \left| \frac{1}{\#B_{n,i}} \sum_{j \in B_{n,i}} f_{i_j}(x) \right| < \varepsilon \quad \forall n \geq n_0,$$

lo cual finaliza la prueba. □

El anterior Teorema es el Ley de los Grandes Números más general que se probará en la tesis. Sin embargo, la particularizaremos a las hipótesis necesarias y suficientes para afrontar los resultados del trabajo.

Corolario 3.6.1 Sean $\pi \in \mathcal{P}(E^{\mathcal{S}}, \mathcal{F})$ y para cada $s \in \mathcal{S}$, sea f_s una función real $\mathcal{F}_{s \cup \partial s}$ -medible con $\pi(f_s) = 0$ y $\pi(\{x \in E^{\mathcal{S}} : |f_s(x)| \leq c\}) = 1$, con c una constante real positiva. Entonces

$$\pi \left(\left\{ x \in E^{\mathcal{S}} : \frac{1}{\#\Lambda_n^{\circ}} \sum_{s \in \Lambda_n^{\circ}} f_s(x) \rightarrow 0 \right\} \right) = 1.$$

Dem. Sea C el cuadrado más chico que contiene al conjunto $(\{0\}^*)^* = \{0\} \cup \partial(\{0\}) \cup \partial(\{0\} \cup \partial(\{0\}))$, el cual es finito por (H.4). Para cada $n \in \mathbb{N}$, sea $T_n : \mathcal{S} \rightarrow \mathcal{S}$ una función tal que

$$\bigcup_{n=1}^{\infty} T_n(C) = \mathcal{S}, \quad T_n(C) \cap T_m(C) = \emptyset \quad \forall n \neq m, \quad \partial T_n(s) \cap \partial T_m(s) = \emptyset \quad \forall s \in C \quad \forall n \neq m.$$

De este modo \mathcal{S} es particionada en los bloques cuadrados $\{T_n(C) : n \in \mathbb{N}\}$, de modo tal que para cada $s \in C$, los sitios la sucesión $\{T_n(s) : n \in \mathbb{N}\}$ no posee vecinos en común. Luego \mathcal{S} puede particionarse en $\#C$ familias infinitas $\{T_n(s) \cup \partial(T_n(s)) : n \in \mathbb{N}\}$ de subconjuntos disjuntos de \mathcal{S} .

Consideremos una función biyectiva de enumeración $R : \mathcal{S} \rightarrow \mathbb{N}$ tal que

$$s \in \Lambda_n^\circ, t \in \Lambda_m^\circ \setminus \Lambda_n^\circ, n < m \Rightarrow R(s) < R(t).$$

Esta es una enumeración consecutiva de los sitios de \mathcal{S} que le asigna un número a todos los sitios de Λ_n° antes de continuar asignándole un número a los sitios de Λ_{n+1}° , para cada $n \in \mathbb{N}$.

Luego por el Teorema 3.6.2, se tiene que

$$\pi \left(\left\{ x \in E^{\mathcal{S}} : \frac{1}{n} \sum_{i=1}^n f_{R^{-1}(i)}(x) \rightarrow 0 \right\} \right) = 1.$$

Como la sucesión $\frac{1}{\#\Lambda_n^\circ} \sum_{s \in \Lambda_n^\circ} f_s(x)$ es una subsucesión de la sucesión $\frac{1}{n} \sum_{i=1}^n f_{R^{-1}(i)}(x)$, finaliza la prueba. \square

3.7. Estimación

Teniendo en cuenta (H.7), consideremos $\vartheta_0 \in \text{Int}(\Upsilon)$. Obtener una observación de π^{ϑ_0} es imposible debido a que \mathcal{S} es infinito. Sin embargo, si es posible obtener observaciones de las leyes $\gamma_\Lambda^{\vartheta_0}(\cdot|w)$ para $\Lambda \in \mathfrak{G}$ y $w \in E^{\mathcal{S}}$ fijados arbitrariamente. Teniendo en cuenta la Proposición 3.2.5, si se obtiene una observación $x_\Lambda \in E^\Lambda$ con probabilidad proporcional a $\rho_\Lambda^{\vartheta_0}(x_\Lambda w^\Lambda)$, entonces $x_\Lambda w^\Lambda \in E^{\mathcal{S}}$ es una observación de $\gamma_\Lambda^{\vartheta_0}(\cdot|w)$, ya que

$$\begin{aligned} \gamma_\Lambda^{\vartheta_0}(\{x_\Lambda w^\Lambda\}|w) &= \gamma_\Lambda^{\vartheta_0} \left(\sigma_\Lambda^{-1}(\{x_\Lambda\}) \cap \sigma_{\mathcal{S} \setminus \Lambda}^{-1}(\{w^\Lambda\}) | w \right) = \gamma_\Lambda^{\vartheta_0}(\sigma_\Lambda^{-1}(\{x_\Lambda\}) | w) 1_{\sigma_{\mathcal{S} \setminus \Lambda}^{-1}(\{w^\Lambda\})}(w) = \\ &= \gamma_\Lambda^{\vartheta_0}(\sigma_\Lambda^{-1}(\{x_\Lambda\}) | w) = \rho_\Lambda^{\vartheta_0}(x_\Lambda w^\Lambda). \end{aligned}$$

Lo interesante de esta cuestión es que $\gamma_\Lambda^{\vartheta_0}(\cdot|w)$ y π^{ϑ_0} comparten el parámetro, y por ende se puede utilizar la observación de $\gamma_\Lambda^{\vartheta_0}(\cdot|w)$ para estimar el parámetro ϑ_0 de π^{ϑ_0} .

Dada una realización $x_\Lambda w^\Lambda$ de $\gamma_\Lambda^{\vartheta_0}(\cdot|w)$, los estimadores de Pseudo-Máxima Verosimilitud de ϑ_0 son los elementos del conjunto

$$M_\Lambda(x_\Lambda w^\Lambda) := \left\{ \vartheta \in \Upsilon : U_\Lambda(\vartheta, x_\Lambda w^\Lambda) = \max_{\alpha \in \Upsilon} U_\Lambda(\alpha, x_\Lambda w^\Lambda) \right\},$$

donde

$$U_\Lambda(\vartheta, x_\Lambda w^\Lambda) := \frac{1}{\#\Lambda^\circ} \sum_{s \in \Lambda^\circ} \ln \rho_s^\vartheta(x_\Lambda w^\Lambda).$$

Notemos que se ha obviado de la notación el supraíndice *PMV* utilizado en la introducción, lo cual se hará de ahora en adelante para no sobredimensionar la escritura.

Como trabajamos con $s \in \Lambda^\circ$, se tiene que $\{s\} \cup \partial s \subseteq \Lambda$. Por la Proposición 3.2.3, la función ρ_s^ϑ es $\mathcal{F}_{s \cup \partial s}$ -medible. Por lo que, por la Proposición 3.2.1(ii), ρ_s^ϑ sólo depende de las entradas sobre $\{s\} \cup \partial s$, y por ende, la función $U_\Lambda(\vartheta, x_\Lambda w^\Lambda)$ no depende de w , por ser \mathcal{F}_Λ -medible. Por tal motivo ignoraremos de la notación a w , y simplemente escribiremos $M_\Lambda(x)$ y

$$U_\Lambda(\vartheta, x_\Lambda w^\Lambda) = U_\Lambda(\vartheta, x) = \frac{1}{\#\Lambda^\circ} \sum_{s \in \Lambda^\circ} \ln \rho_s^\vartheta(x) = \frac{1}{\#\Lambda^\circ} \sum_{s \in \Lambda^\circ} (H_{s,0}(x) + \vartheta^T H_s(x) - \ln Z_s^\vartheta(x)),$$

donde $x \in E^S$ contiene sobre Λ la observación x_Λ del modelo.

Para cada $n \in \mathbb{N}$, consideremos arbitrariamente un elemento de $M_{\Lambda_n}(x)$, y lo notemos con $\widehat{\vartheta}_n(x)$. En lo que sigue se utilizará la notación U_n en lugar de la notación U_{Λ_n} .

En la Sección 3.8 probaremos que se verifican las siguientes propiedades asintóticas:

(T.1) Unicidad asintótica:

$$\pi^{\vartheta_0}(\{x \in E^S : \#M_{\Lambda_n}(x) \rightarrow 1\}) = 1;$$

(T.2) Consistencia:

$$\pi^{\vartheta_0}(\{x \in E^S : \|\widehat{\vartheta}_n(x) - \vartheta_0\| \rightarrow 0\}) = 1.$$

3.8. Consistencia.

Una propiedad deseable para cualquier estimador es que su estimación tienda al verdadero parámetro cuando la muestra es lo suficientemente grande. En la presente sección analizaremos las hipótesis bajo las cuales valen (T1) y (T2) para el estimador de PMV.

En primera instancia probaremos la consistencia para el caso invariante, basándonos en la prueba realizada por Geman y Graffigne en [4], pero utilizando el Corolario 3.6.1 en lugar de la única Proposición dada en [4]. Teniendo en mente la suposición (H.10), estamos trabajando sobre una familia de modelos exponenciales que contiene a los modelos exponenciales [29] considerados en [4]. Tal trabajo se desarrolló bajo el supuesto de que $H_{\Lambda,0} \equiv 0 \forall \Lambda \in \mathfrak{G}$, y por tal motivo la sección 3.8.2 puede considerarse una extensión del mismo, aunque la prueba es similar.

Posteriormente, generalizaremos este resultado reemplazando la hipótesis de invarianza (H.14) por dos hipótesis: una que supone no correlación entre las funciones h_s y H_s , y la otra que supone una cota para una sucesión que asegura la equivalencia microscópica entre la especificación del modelo no invariante particular con el que se trabajará, y la de un modelo invariante asociado.

Este modelo asociado pertenecerá a la clase de familias exponenciales considerada en [4], y será invariante con sólo suponer válido (H.13), independientemente de (H.14). En la prueba de consistencia para el caso invariante, y sobre todo en la prueba del caso no invariante, este modelo asociado cumple un rol muy importante, y por lo tanto lo definiremos y estudiaremos antes de comenzar con las pruebas de consistencia.

3.8.1. Modelo Invariante Asociado.

Recordemos que

$$\rho_\Lambda^\vartheta(x) = \frac{h_\Lambda(x) \exp\{\vartheta^T H_\Lambda(x)\}}{Z_\Lambda^\vartheta(x)} \quad \forall \Lambda \in \mathfrak{G},$$

y llamaremos función asociada a ρ_Λ^ϑ a la función $\tilde{\rho}_\Lambda^\vartheta : E^S \rightarrow (0, \infty)$ dada por

$$\tilde{\rho}_\Lambda^\vartheta(x) := \frac{\exp\{\vartheta^T H_\Lambda(x)\}}{\tilde{Z}_\Lambda^\vartheta(x)} \quad \forall \Lambda \in \mathfrak{S},$$

donde

$$\tilde{Z}_\Lambda^\vartheta(x) := \int_{E^\Lambda} \exp\{\vartheta^T H_\Lambda(\xi x^\Lambda)\} \lambda^\Lambda(d\xi).$$

Para la cual se puede definir la especificación $\tilde{\gamma}^\vartheta = \{\tilde{\gamma}_\Lambda^\vartheta\}_{\Lambda \in \mathfrak{S}}$, donde para cada $\Lambda \in \mathfrak{S}$, $\tilde{\gamma}_\Lambda^\vartheta$ es una función de $\mathcal{F} \times E^S$ en $[0, 1]$ dada por

$$\tilde{\gamma}_\Lambda^\vartheta(A|x) := \int_{E^\Lambda} 1_A(\xi x^\Lambda) \tilde{\rho}_\Lambda^\vartheta(\xi x^\Lambda) \lambda^\Lambda(d\xi).$$

Si existen dos parámetros que generan la misma distribución, entonces una muestra de una de ellas puede provocar que estimemos la otra. Por este motivo, agreguemos la siguiente suposición que tiene como fin eliminar la posibilidad de que esto suceda.

$$(H.15) \quad s \in \mathcal{S}, \vartheta \neq \vartheta_0 \Rightarrow \tilde{\rho}_s^\vartheta \neq \tilde{\rho}_s^{\vartheta_0},$$

Esta suposición de identificabilidad del modelo asociado generaliza la condición de identificabilidad supuesta por Geman y Graffigne en [4], ya que son equivalentes si $H_{\Lambda,0} \equiv 0 \forall \Lambda \in \mathfrak{S}$.

Analicemos algunas propiedades que serán de gran utilidad en la prueba de consistencia del caso no invariante, y que valen independientemente de las hipótesis de invarianza (H.13) y (H.14).

Lema 3.8.1 *Para cada $\Lambda \in \mathfrak{S}$, se verifican*

$$(i) \quad \frac{Z_\Lambda^\vartheta(x)}{\tilde{Z}_\Lambda^\vartheta(x)} = \tilde{\gamma}_\Lambda^\vartheta(h_\Lambda|x);$$

$$(ii) \quad \frac{\rho_\Lambda^\vartheta(x)}{\tilde{\rho}_\Lambda^\vartheta(x)} = \frac{h_\Lambda(x)}{\tilde{\gamma}_\Lambda^\vartheta(h_\Lambda|x)}.$$

(iii) *Dada una función columna $f : E^S \rightarrow \mathbb{R}^k$, $k \in \mathbb{N}$, \mathcal{F} -medible se verifica que*

$$\gamma_\Lambda^\vartheta(f|\cdot) = \frac{\tilde{\gamma}_\Lambda^\vartheta(fh_\Lambda|\cdot)}{\tilde{\gamma}_\Lambda^\vartheta(h_\Lambda|\cdot)} \quad \forall \Lambda \in \mathfrak{S}.$$

Dem. Veamos (i):

$$\frac{Z_\Lambda^\vartheta(x)}{\tilde{Z}_\Lambda^\vartheta(x)} = \frac{\int_{E^\Lambda} h_\Lambda(\xi x^\Lambda) \exp\{\vartheta^T H_\Lambda(\xi x^\Lambda)\} \lambda^\Lambda(d\xi)}{\int_{E^\Lambda} \exp\{\vartheta^T H_\Lambda(\xi x^\Lambda)\} \lambda^\Lambda(d\xi)} = \int_{E^\Lambda} h_\Lambda(\xi x^\Lambda) \tilde{\rho}_\Lambda^\vartheta(\xi x^\Lambda) \lambda^\Lambda(d\xi) = \tilde{\gamma}_\Lambda^\vartheta(h_\Lambda|x),$$

donde la última igualdad vale por la Proposición 3.2.7(ii).

Veamos (ii): usando (i), se obtiene que

$$\frac{\rho_\Lambda^\vartheta(x)}{\tilde{\rho}_\Lambda^\vartheta(x)} = \frac{\frac{1}{Z_\Lambda^\vartheta(x)} h_\Lambda(x) \exp\{\vartheta^T H_\Lambda(x)\}}{\frac{1}{\tilde{Z}_\Lambda^\vartheta(x)} \exp\{\vartheta^T H_\Lambda(x)\}} = \frac{h_\Lambda(x)}{\tilde{\gamma}_\Lambda^\vartheta(h_\Lambda|x)}.$$

Veamos (iii):

$$\begin{aligned}\gamma_{\Lambda}^{\vartheta}(f|x) &= \int_{E^{\Lambda}} f(\xi x^{\Lambda}) \frac{h_{\Lambda}(\xi x^{\Lambda}) \exp\{\vartheta^T H_{\Lambda}(\xi x^{\Lambda})\}}{Z_{\Lambda}^{\vartheta}(x)} \lambda^{\Lambda}(d\xi) \\ &= \frac{\tilde{Z}_{\Lambda}^{\vartheta}(x)}{Z_{\Lambda}^{\vartheta}(x)} \int_{E^{\Lambda}} f(\xi x^{\Lambda}) h_{\Lambda}(\xi x^{\Lambda}) \frac{\exp\{\vartheta^T H_{\Lambda}(\xi x^{\Lambda})\}}{\tilde{Z}_{\Lambda}^{\vartheta}(x)} \lambda^{\Lambda}(d\xi) = \frac{\tilde{\gamma}_{\Lambda}^{\vartheta}(fh_{\Lambda}|x)}{\tilde{\gamma}_{\Lambda}^{\vartheta}(h_{\Lambda}|x)},\end{aligned}$$

donde en la última igualdad se usa (i). □

Definición 3.8.1 Dos especificaciones γ y $\tilde{\gamma}$ se dicen equivalentes microscópicamente, y se nota con $\gamma \approx \tilde{\gamma}$, si existe una constante $c \in [1, \infty)$ tal que

$$c^{-1} \gamma_{\Lambda}(A|\cdot) \leq \tilde{\gamma}_{\Lambda}(A|\cdot) \leq c \gamma_{\Lambda}(A|\cdot) \quad \forall A \in \mathcal{F} \quad \forall \Lambda \in \mathfrak{S}.$$

Lema 3.8.2 Si para cada $n \in \mathbb{N}$ existe una función $f_n : E^{\mathcal{S}} \rightarrow (0, \infty)$ \mathcal{F} -medible tales que $\sup_{n \in \mathbb{N}} \|\ln f_n\|_{\infty} < \infty$ y

$$\tilde{\gamma}_{\Lambda_n}(A|\cdot) = \int_A f_n(x) \gamma_{\Lambda_n}(dx|\cdot) \quad \forall A \in \mathcal{F}.$$

Entonces $\gamma \approx \tilde{\gamma}$.

Dem. Dados $A \in \mathcal{F}$ y $n \in \mathbb{N}$, por un lado

$$\tilde{\gamma}_{\Lambda_n}(A|\cdot) = \int_A f_n(x) \gamma_{\Lambda_n}(dx|\cdot) \leq \|f_n\|_{\infty} \int_A \gamma_{\Lambda_n}(dx|\cdot) \leq c_1 \gamma_{\Lambda_n}(A|\cdot),$$

donde $c_1 = \sup_n \|f_n\|_{\infty} \in (0, \infty)$. Por el otro lado, usando la Desigualdad de Jensen

$$\ln \tilde{\gamma}_{\Lambda_n}(A|\cdot) \geq \int_A \ln f_n(x) \gamma_{\Lambda_n}(dx|\cdot) \geq -\|\ln f_n(x)\|_{\infty} \geq -\sup_n \|\ln f_n(x)\|_{\infty}.$$

Por lo que

$$\tilde{\gamma}_{\Lambda_n}(A|\cdot) \geq \left(\gamma_{\Lambda_n}(A|\cdot) \exp \left\{ \sup_n \|\ln f_n(x)\|_{\infty} \right\} \right)^{-1} \gamma_{\Lambda_n}(A|\cdot) \geq c_2^{-1} \gamma_{\Lambda_n}(A|\cdot),$$

donde $c_2 = \exp\{\sup_n \|\ln f_n(x)\|_{\infty}\}$. Luego el resultado surge de tomar $c = \max\{c_1, c_2\}$. □

Luego agregando el siguiente supuesto

$$(H.16) \quad \sup_n \left\| \ln \frac{h_{\Lambda_n}}{\tilde{\gamma}_{\Lambda_n}^{\vartheta}(h_{\Lambda_n}|\cdot)} \right\|_{\infty} < \infty \quad \forall \vartheta \in \Upsilon,$$

se puede probar la equivalencia microscópica de los modelos asociados.

Lema 3.8.3 Para cada $\vartheta \in \Upsilon$, $\gamma^{\vartheta} \approx \tilde{\gamma}^{\vartheta}$.

Dem. Consideremos las funciones \mathcal{F} -medibles $f_n := \frac{h_{\Lambda_n}}{\tilde{\gamma}_{\Lambda_n}^\vartheta(h_{\Lambda_n}|\cdot)}$. Luego $\sup_n \|\ln f_n\|_\infty < \infty$ por (H.16). Además por el Lema 3.8.1(ii),

$$\begin{aligned} \tilde{\gamma}_{\Lambda_n}^\vartheta(A|x) &= \int_{E^{\Lambda_n}} 1_A(\xi x^{\Lambda_n}) \tilde{\rho}_{\Lambda_n}^\vartheta(\xi x^{\Lambda_n}) \lambda^{\Lambda_n}(d\xi) = \int_{E^{\Lambda_n}} 1_A(\xi x^{\Lambda_n}) \frac{h_{\Lambda_n}(\xi x^{\Lambda_n})}{\tilde{\gamma}_{\Lambda_n}^\vartheta(h_{\Lambda_n}|x)} \rho_{\Lambda_n}^\vartheta(\xi x^{\Lambda_n}) \lambda^{\Lambda_n}(d\xi) = \\ &= \int_{E^{\Lambda_n}} 1_A(\xi x^{\Lambda_n}) f_n(\xi x^{\Lambda_n}) \rho_{\Lambda_n}^\vartheta(\xi x^{\Lambda_n}) \lambda^{\Lambda_n}(d\xi) = \int_A f_n(y) \gamma_{\Lambda_n}^\vartheta(dy|x), \end{aligned}$$

donde la última igualdad vale por la Proposición 3.2.7(ii). □

Proposición 3.8.1 Sean γ y $\tilde{\gamma}$ dos especificaciones equivalentes. Entonces se verifican

- (i) $\mathcal{G}(\gamma) \neq \emptyset \Leftrightarrow \mathcal{G}(\tilde{\gamma}) \neq \emptyset$;
- (ii) existe una biyección $\pi \leftrightarrow \tilde{\pi}$ entre $\mathcal{G}(\gamma)$ y $\mathcal{G}(\tilde{\gamma})$;
- (iii) $\#ex(\mathcal{G}(\tilde{\gamma})) = \#ex(\mathcal{G}(\gamma))$.

Dem. Corresponde al Teorema 7.33 de [27]. □

Por el Lema 3.8.3 y la Proposición 3.8.1, para cada $\vartheta \in \Upsilon$, existe una probabilidad $\tilde{\pi}^\vartheta \in ex(\mathcal{G}(\tilde{\gamma}^\vartheta))$, la cual verifica el siguiente resultado

Lema 3.8.4 Para cada $\vartheta \in \Upsilon$, $\pi^\vartheta \ll \tilde{\pi}^\vartheta$ y $\tilde{\pi}^\vartheta \ll \pi^\vartheta$.

Dem. Por simetría, sólo probaremos que $\tilde{\pi}^\vartheta \ll \pi^\vartheta$. Sea $A \in \mathcal{F}$, tal que $\pi^\vartheta(A) = 0$. Como $\pi^\vartheta \in ex(\mathcal{G}(\gamma^\vartheta))$ y $\tilde{\pi}^\vartheta \in ex(\mathcal{G}(\tilde{\gamma}^\vartheta))$, entonces por el Teorema 3.2.1(ii), existe $x \in E^S$ tal que $\gamma_{\Lambda_n}^\vartheta(A|x) \rightarrow \pi^\vartheta(A) = 0$ y $\tilde{\gamma}_{\Lambda_n}^\vartheta(A|x) \rightarrow \tilde{\pi}^\vartheta(A)$. Luego por el Lema 3.8.3, existe $c \geq 1$ tal que $0 \leq \tilde{\gamma}_{\Lambda_n}^\vartheta(A|x) \leq c\gamma_{\Lambda_n}^\vartheta(A|x) \rightarrow 0$, lo cual implica que $\tilde{\pi}^\vartheta(A) = 0$, y culmina la prueba. □

3.8.2. Caso Invariante por Traslaciones.

La prueba del caso en el cual las especificaciones son invariante por traslaciones, se hará siguiendo la prueba dada por Geman y Graffigne en [4], modificando y detallando las pruebas, y adaptando notación. En la Proposición 3.4.2 se mostró que, bajo los supuestos (H.1)-(H.11), las especificaciones son invariantes sí, y sólo si, se verifican (H.13) y (H.14). En la presente sección entonces supondremos válidas (H.1)-(H.14). Además asumiremos que vale (H.15) para lograr identificabilidad. No se necesita la validez de (H.16) en el análisis del caso invariante.

Los conjuntos $E^{\partial s}$ son iguales salvo la debida traslación, y existe una biyección entre $E^{\partial s}$ (funciones de ∂s en E) y $E^{\#\partial s}$ ($\#\partial s$ -úplas de elementos de E). Por tal motivo, y con el fin de simplificar la notación, se escribirá

indistintamente $E^{\partial s}$ y $E^{\#\partial s}$. Luego, bajo el supuesto de invarianza, teniendo en cuenta las Proposiciones 3.3.2 y 3.2.5, las densidades condicionales $\pi_s^{\partial_0} : E^{\{s\}} \times E^{\mathcal{S} \setminus \{s\}} \rightarrow [0, 1]$ dadas por

$$\pi_s^{\partial_0}(x_s | x^s) = \rho_s^{\partial_0}(x) = \gamma_s^{\partial_0}(\sigma_s^{-1}(\{x_s\}) | x)$$

son iguales salvo la debida traslación. Teniendo en cuenta la Proposición 3.2.3, las densidades condicionales $\pi_s^{\partial_0}$ pueden ser vistas como funciones de $E \times E^{\#\partial s}$ en $[0, 1]$, que no dependen del sitio s considerado.

Para cada $n \in \mathbb{N}$, usando a ξ y \mathbf{v} como elementos genéricos de E y $E^{\partial s}$ respectivamente, definamos las funciones $N_n(\mathbf{v}) : E^{\mathcal{S}} \rightarrow \mathbb{R}$ y $N_n(\xi, \mathbf{v}) : E^{\mathcal{S}} \rightarrow \mathbb{R}$ dadas por

$$N_n(\mathbf{v})(x) := \#\{s \in \Lambda_n^\circ : x_{\partial s} = \mathbf{v}\} \quad \text{y} \quad N_n(\xi, \mathbf{v})(x) := \#\{s \in \Lambda_n^\circ : x_s = \xi, x_{\partial s} = \mathbf{v}\}.$$

Los siguientes Lemas tienen como fin segmentar la prueba del Teorema 3.8.1.

Lema 3.8.5

$$\liminf_{n \rightarrow \infty} \frac{N_n(\mathbf{v})}{\#\Lambda_n^\circ} > 0 \quad \pi^{\partial_0}\text{-}p.p. \quad \forall \mathbf{v} \in E^{\partial s}$$

Dem. Fijemos $\mathbf{v} \in E^{\partial s}$ y para cada $s \in \mathcal{S}$, consideremos las funciones $Y_s : E^{\mathcal{S}} \rightarrow \mathbb{R}$ dadas por

$$Y_s(x) := \begin{cases} 1 - \pi^{\partial_0}(\sigma_{\partial s}^{-1}(\{\mathbf{v}\})) & \text{si } x_{\partial s} = \mathbf{v} \\ -\pi^{\partial_0}(\sigma_{\partial s}^{-1}(\{\mathbf{v}\})) & \text{en otro caso,} \end{cases}$$

las cuales son $\mathcal{F}_{\partial s}$ -medibles y acotadas. Además,

$$\pi^{\partial_0}(Y_s) = (1 - \pi^{\partial_0}(\sigma_{\partial s}^{-1}(\{\mathbf{v}\})))\pi^{\partial_0}(\sigma_{\partial s}^{-1}(\{\mathbf{v}\})) - \pi^{\partial_0}(\sigma_{\partial s}^{-1}(\{\mathbf{v}\}))(1 - \pi^{\partial_0}(\sigma_{\partial s}^{-1}(\{\mathbf{v}\}))) = 0.$$

Luego por el Corolario 3.6.1 se tiene que

$$\bar{Y}_n := \frac{1}{\#\Lambda_n^\circ} \sum_{s \in \Lambda_n^\circ} Y_s = \frac{1}{\#\Lambda_n^\circ} (N_n(\mathbf{v}) - \#\Lambda_n^\circ \pi^{\partial_0}(\sigma_{\partial s}^{-1}(\{\mathbf{v}\}))) = \frac{N_n(\mathbf{v})}{\#\Lambda_n^\circ} - \pi^{\partial_0}(\sigma_{\partial s}^{-1}(\{\mathbf{v}\})) \rightarrow 0. \quad \pi^{\partial_0}\text{-}p.p.$$

Luego, teniendo en cuenta que π^{∂_0} pertenece a la familia exponencial,

$$\frac{N_n(\mathbf{v})}{\#\Lambda_n^\circ} \rightarrow \pi^{\partial_0}(\sigma_{\partial s}^{-1}(\{\mathbf{v}\})) > 0 \quad \pi^{\partial_0}\text{-}p.p.,$$

lo cual prueba el Lema. □

Lema 3.8.6

$$\lim_{n \rightarrow \infty} \frac{N_n(\xi, \mathbf{v})}{N_n(\mathbf{v})} = \pi_s^{\partial_0}(\xi | \mathbf{v}) = \gamma_s^{\partial_0}(\sigma_s^{-1}(\{\xi\}) | y) \quad \pi^{\partial_0}\text{-}p.p. \quad \forall \xi \in E, \mathbf{v} \in E^{\partial s},$$

donde $y \in E^{\mathcal{S}}$ es tal que $\sigma_{\partial s}(y) = \mathbf{v}$.

Dem. Fijemos $\xi \in E$ y $\nu \in E^{\partial s}$, y para cada $s \in \mathcal{S}$, consideremos las funciones $Y_s : E^{\mathcal{S}} \rightarrow \mathbb{R}$ dadas por

$$Y_s(x) := \begin{cases} 1 - \pi_s^{\vartheta_0}(\xi|\nu) & \text{si } x_s = \xi, x_{\partial s} = \nu \\ 0 & \text{si } x_{\partial s} \neq \nu \\ -\pi_s^{\vartheta_0}(\xi|\nu) & \text{si } x_s \neq \xi, x_{\partial s} = \nu, \end{cases}$$

las cuales son $\mathcal{F}_{s \cup \partial s}$ -medibles y acotadas. Además

$$\begin{aligned} \pi^{\vartheta_0}(Y_s) &= (1 - \pi_s^{\vartheta_0}(\xi|\nu))\pi^{\vartheta_0}(\sigma_{s \cup \partial s}^{-1}(\{\xi\nu\})) - \pi_s^{\vartheta_0}(\xi|\nu) (\pi^{\vartheta_0}(\sigma_{\partial s}^{-1}(\{\nu\})) - \pi^{\vartheta_0}(\sigma_{s \cup \partial s}^{-1}(\{\xi\nu\}))) \\ &= \pi^{\vartheta_0}(\sigma_{s \cup \partial s}^{-1}(\{\xi\nu\})) - \pi_s^{\vartheta_0}(\xi|\nu)\pi^{\vartheta_0}(\sigma_{s \cup \partial s}^{-1}(\{\xi\nu\})) - \pi_s^{\vartheta_0}(\xi|\nu)\pi^{\vartheta_0}(\sigma_{\partial s}^{-1}(\{\nu\})) + \\ &\quad + \pi_s^{\vartheta_0}(\xi|\nu)\pi^{\vartheta_0}(\sigma_{s \cup \partial s}^{-1}(\{\xi\nu\})) \\ &= \pi^{\vartheta_0}(\sigma_{s \cup \partial s}^{-1}(\{\xi\nu\})) - \pi_s^{\vartheta_0}(\xi|\nu)\pi^{\vartheta_0}(\sigma_{\partial s}^{-1}(\{\nu\})) \\ &= 0, \end{aligned}$$

donde la última igualdad vale pues

$$\frac{\pi^{\vartheta_0}(\sigma_{s \cup \partial s}^{-1}(\{\xi\nu\}))}{\pi^{\vartheta_0}(\sigma_{\partial s}^{-1}(\{\nu\}))} = \frac{\pi^{\vartheta_0}(\sigma_s^{-1}(\{\xi\}) \cap \sigma_{\partial s}^{-1}(\{\nu\}))}{\pi^{\vartheta_0}(\sigma_{\partial s}^{-1}(\{\nu\}))} = \pi^{\vartheta_0}(\sigma_s^{-1}(\{\xi\}) | \sigma_{\partial s}^{-1}(\{\nu\})) = \pi_s^{\vartheta_0}(\xi|\nu).$$

Luego por el Corolario 3.6.1 se tiene que

$$\bar{Y}_n := \frac{1}{\#\Lambda_n^\circ} \sum_{s \in \Lambda_n^\circ} Y_s = \frac{1}{\#\Lambda_n^\circ} (N_n(\xi, \nu) - N_n(\nu)\pi_s^{\vartheta_0}(\xi|\nu)) = \frac{N_n(\nu)}{\#\Lambda_n^\circ} \left(\frac{N_n(\xi, \nu)}{N_n(\nu)} - \pi_s^{\vartheta_0}(\xi|\nu) \right) \rightarrow 0 \text{ p.p.}$$

Finalmente, por el Lema 3.8.5

$$\frac{N_n(\xi, \nu)}{N_n(\nu)} \rightarrow \pi_s^{\vartheta_0}(\xi|\nu) \text{ } \pi^{\vartheta_0}\text{-p.p.}$$

□

Lema 3.8.7 Dado $\vartheta \in \text{Int}(\Upsilon)$, se verifican:

- (i) $\nabla \ln Z_\Lambda^\vartheta = \gamma_\Lambda^\vartheta(H_\Lambda | \cdot) \quad \forall \Lambda \in \mathfrak{S}$,
- (ii) $\nabla U_\Lambda(\vartheta, \cdot) = \frac{1}{\#\Lambda^\circ} \sum_{s \in \Lambda^\circ} (H_s - \gamma_s^\vartheta(H_s | \cdot)) \quad \forall \Lambda \in \mathfrak{S}$,
- (iii) $\nabla^2 U_\Lambda(\vartheta, \cdot) = -\frac{1}{\#\Lambda^\circ} \sum_{s \in \Lambda^\circ} \gamma_s^\vartheta \left(H_s (H_s - \gamma_s^\vartheta(H_s | \cdot))^T | \cdot \right) \leq 0 \quad \forall \Lambda \in \mathfrak{S}$,

donde ∇ y ∇^2 representan respectivamente el vector gradiente y la matriz hessiana con respecto al parámetro del modelo.

Dem. Veamos (i): para cada $i = 1, \dots, m$, se tiene que

$$\frac{\partial}{\partial \vartheta_i} \ln Z_\Lambda^\vartheta(x) = \frac{\int_{E^\Lambda} h_\Lambda(\xi x^\Lambda) \frac{\partial}{\partial \vartheta_i} \exp\{\vartheta^T H_\Lambda(\xi x^\Lambda)\} \lambda^\Lambda(d\xi)}{Z_\Lambda^\vartheta(x)} = \frac{\int_{E^\Lambda} h_\Lambda(\xi x^\Lambda) H_{\Lambda,i}(\xi x^\Lambda) \exp\{\vartheta^T H_\Lambda(\xi x^\Lambda)\} \lambda^\Lambda(d\xi)}{Z_\Lambda^\vartheta(x)}$$

$$= \int_{E^\Lambda} H_{\Lambda,i}(\xi x^\Lambda) \rho_\Lambda^\vartheta(\xi x^\Lambda) \lambda^\Lambda(d\xi) = \gamma_\Lambda^\vartheta(H_{\Lambda,i}|x).$$

Escribiendo en forma matricial:

$$\nabla \ln Z_\Lambda^\vartheta(x) = \gamma_\Lambda^\vartheta(H_\Lambda|x).$$

Veamos (ii): usando (i), se tiene que

$$\nabla U_\Lambda(\vartheta, x) = \frac{1}{\#\Lambda^\circ} \sum_{s \in \Lambda^\circ} \nabla(\vartheta^T H_s(x) - \ln Z_s^\vartheta(x)) = \frac{1}{\#\Lambda^\circ} \sum_{s \in \Lambda^\circ} (H_s(x) - \gamma_s^\vartheta(H_s|x))$$

Veamos (iii):

$$\nabla^2 U_\Lambda(\vartheta, x) = -\frac{1}{\#\Lambda^\circ} \sum_{s \in \Lambda^\circ} \nabla \gamma_s^\vartheta(H_s|x) = -\frac{1}{\#\Lambda^\circ} \sum_{s \in \Lambda^\circ} \int_E H_s(\xi x^s) \nabla \rho_s^\vartheta(\xi x^s) \lambda^\Lambda(d\xi) = (*).$$

Como

$$\begin{aligned} \nabla \rho_s^\vartheta(x) &= h_s(x) \nabla \frac{\exp\{\vartheta^T H_s(x)\}}{Z_s^\vartheta(x)} = h_s(x) \exp\{\vartheta^T H_s(x)\} \frac{H_s(x) Z_s^\vartheta(x) - \nabla Z_s^\vartheta(x)}{Z_s^\vartheta(x)^2} = \\ &= \rho_s^\vartheta(x) (H_s(x) - \nabla \ln Z_s^\vartheta(x)) = \rho_s^\vartheta(x) (H_s(x) - \gamma_s^\vartheta(H_s|x)), \end{aligned}$$

donde la última igualdad vale por (i). Luego

$$\begin{aligned} (*) &= -\frac{1}{\#\Lambda^\circ} \sum_{s \in \Lambda^\circ} \int_E H_s(\xi x^s) (H_s(\xi x^s) - \gamma_s^\vartheta(H_s|x))^T \rho_s^\vartheta(\xi x^s) \lambda^\Lambda(d\xi) \\ &= -\frac{1}{\#\Lambda^\circ} \sum_{s \in \Lambda^\circ} \gamma_s^\vartheta \left(H_s (H_s - \gamma_s^\vartheta(H_s|x))^T \mid x \right), \end{aligned}$$

la cual es una cantidad no positiva por la desigualdad de Cauchy-Schwarz. □

Definición 3.8.2 Diremos que una afirmación vale $\forall n \gg 1$, si existe $n_0 \in \mathbb{N}$ tal que la afirmación vale $\forall n \geq n_0$.

Lema 3.8.8 $\pi^{\vartheta_0}(\{x \in E^S : U_n(\cdot, x) \text{ es estrictamente cóncava } \forall n \gg 1\}) = 1$.

Dem. Sea $x \in E^S$ en el conjunto de probabilidad 1, dado en el Lema 3.8.5, de las configuraciones que verifican

$$\liminf_{n \rightarrow \infty} \frac{N_n(\mathfrak{v})(x)}{\#\Lambda_n^\circ} > 0 \quad \forall \mathfrak{v} \in E^{\partial s}. \quad (3.8.7)$$

Por el Lema 3.8.7(iii),

$$\nabla^2 U_n(\vartheta_0, x) = -\frac{1}{\#\Lambda_n^\circ} \sum_{s \in \Lambda_n^\circ} \gamma_s^{\vartheta_0} \left(H_s (H_s - \gamma_s^{\vartheta_0}(H_s|x))^T \mid x \right) \leq 0.$$

Luego, $U_n(\cdot, x)$ es cóncava y

$$\phi^T \nabla^2 U_n(\vartheta_0, x) \phi = -\frac{1}{\#\Lambda_n^\circ} \sum_{s \in \Lambda_n^\circ} \gamma_s^{\vartheta_0} \left(\left(\phi^T H_s - \gamma_s^{\vartheta_0}(\phi^T H_s|x) \right)^2 \mid x \right) \leq 0 \quad \forall \phi \in \mathbb{R}^m.$$

Sea $\vartheta \neq \vartheta_0$ y supongamos que $\phi^T \nabla^2 U_n(\vartheta, x) \phi = 0$ para algún ϕ no nulo. Entonces,

$$\phi^T H_s(\xi x^s) = \gamma_s^\vartheta (\phi^T H_s | x) \quad \forall \xi \in E \quad \forall s \in \mathcal{S}.$$

En particular, $H_s(\cdot x^s)$ es constante $\forall s \in \mathcal{S}$, y esto implica que

$$\tilde{\rho}_s^\vartheta(\xi x^s) = \frac{\exp\{\vartheta^T H_s(\xi x^s)\}}{\int_E \exp\{\vartheta^T H_s(\tilde{\xi} x^s)\} \lambda(d\tilde{\xi})} = \frac{1}{\#E} \quad \forall \xi \in E \quad \forall s \in \mathcal{S}. \quad (3.8.8)$$

Luego, para que (3.8.8) contradiga (H.15), se necesita que

$$E^{\partial 0} = \{(\boldsymbol{\theta}_{-s}(x))_{\partial 0} : s \in \Lambda_n^\circ\} \quad \forall n \gg 1, \quad (3.8.9)$$

ya que si esto sucede, si tomamos n lo suficientemente grande y $\tilde{x} \in E^{\mathcal{S}}$, entonces existe $\tilde{s} \in \Lambda_n^\circ$ tal que $(\boldsymbol{\theta}_{-\tilde{s}}(x))_{\partial 0} = \tilde{x}_{\partial 0}$. Luego

$$\tilde{\rho}_0^\vartheta(\tilde{x}) = \tilde{\rho}_0^\vartheta(\tilde{x}_{\partial 0} \tilde{x}^{\partial 0}) = \tilde{\rho}_0^\vartheta((\boldsymbol{\theta}_{-\tilde{s}}(x))_{\partial 0} \tilde{x}^{\partial 0}).$$

Análogamente para ϑ_0

$$\tilde{\rho}_0^{\vartheta_0}(\tilde{x}) = \tilde{\rho}_0^{\vartheta_0}((\boldsymbol{\theta}_{-\tilde{s}}(x))_{\partial 0} \tilde{x}^{\partial 0}).$$

Como $\tilde{\rho}_0^\vartheta$ una función $\mathcal{F}_{0 \cup \partial 0}$ -medible,

$$\tilde{\rho}_0^\vartheta(\tilde{x}) = \tilde{\rho}_0^\vartheta(\tilde{x}_0 (\boldsymbol{\theta}_{-\tilde{s}}(x))^0).$$

Análogamente para ϑ_0

$$\tilde{\rho}_0^{\vartheta_0}(\tilde{x}) = \tilde{\rho}_0^{\vartheta_0}(\tilde{x}_0 (\boldsymbol{\theta}_{-\tilde{s}}(x))^0).$$

Luego por (3.8.8) y por (H.13), resulta que

$$\tilde{\rho}_0^\vartheta(\tilde{x}) = \tilde{\rho}_0^\vartheta(\boldsymbol{\theta}_{-\tilde{s}}(x)) = \tilde{\rho}_{\tilde{s}}^\vartheta(x) = \tilde{\rho}_{\tilde{s}}^{\vartheta_0}(x) = \tilde{\rho}_0^{\vartheta_0}(\boldsymbol{\theta}_{-\tilde{s}}(x)) = \tilde{\rho}_0^{\vartheta_0}(\tilde{x}_0 (\boldsymbol{\theta}_{-\tilde{s}}(x))^0) = \tilde{\rho}_0^{\vartheta_0}(\tilde{x}).$$

Por lo que $\tilde{\rho}_0^\vartheta = \tilde{\rho}_0^{\vartheta_0}$, lo cual contradice (H.15).

Veamos entonces que vale (3.8.9): supongamos entonces que existe $\zeta_0 \in E^{\partial 0}$ tal que $\zeta_0 \neq (\boldsymbol{\theta}_{-s}(x))_{\partial 0} \forall s \in \mathcal{S}$. Entonces $N_n(\zeta_0)(x) = 0 \forall n \in \mathbb{N}$, y por lo tanto $\liminf_{n \rightarrow \infty} N_n(\zeta_0)(x) / \#\Lambda_n^\circ = 0$, lo cual contradice la cláusula del conjunto al cual pertenece x . Luego para cada $\zeta \in E^{\partial 0}$, existe $s_\zeta \in \mathcal{S}$, tal que $(\boldsymbol{\theta}_{-s_\zeta}(x))_{\partial 0} = \zeta$. Por (H.1), vale (3.8.9) y concluye la prueba. \square

Este Lema nos dice que vale (T.1) en el caso invariante, esto es que $U_n(\cdot, x)$ tiene un único máximo $\widehat{\vartheta}_n(x)$ para todo $n \gg 1$, lo cual no implica que ϑ_0 sea tal máximo, ni que $\widehat{\vartheta}_n(x) \rightarrow \vartheta_0$. Definiremos una nueva función $G_n : \Upsilon \times E^{\mathcal{S}} \rightarrow \mathbb{R}$, para la cual $G_n(\cdot, x)$ tiene un único máximo en ϑ_0 para cada $n \in \mathbb{N}$ y todo $x \in E^{\mathcal{S}}$, y que se aproxime uniformemente a $U_n(\cdot, x)$ en un entorno de ϑ_0 . Para ello observemos primero que por (H.1), (H.13) y (H.14),

$$U_n(\vartheta, \cdot) = \sum_{\mathbf{v}} \frac{N_n(\mathbf{v})}{\#\Lambda_n^\circ} \sum_{\xi} \frac{N_n(\xi, \mathbf{v})}{N_n(\mathbf{v})} \ln \pi_s^\vartheta(\xi | \mathbf{v})$$

y teniendo en cuenta el Lema 3.8.6, probemos el siguiente resultado

Lema 3.8.9 Si para cada $n \in \mathbb{N}$, definimos la función $G_n : \Upsilon \times E^S \rightarrow \mathbb{R}$, dada por

$$G_n(\vartheta, x) := \sum_{\mathfrak{v}} \frac{N_n(\mathfrak{v})(x)}{\#\Lambda_n^\circ} \sum_{\xi} \pi_s^{\vartheta_0}(\xi|\mathfrak{v}) \ln \pi_s^{\vartheta}(\xi|\mathfrak{v}).$$

Entonces

(i) Con π^{ϑ_0} -probabilidad 1, para cada $\varepsilon \in (0, d(\vartheta_0, \partial\Upsilon))$, existe $\delta > 0$, tal que

$$\limsup_{n \rightarrow \infty} \sup_{|\vartheta - \vartheta_0| \leq \varepsilon} \sup_{\phi \in \mathbb{R}^m: \|\phi\|=1} \phi^T \nabla^2 G_n(\vartheta, \cdot) \phi < -\delta,$$

donde $d(\vartheta_0, \partial\Upsilon) = \inf_{\vartheta \in \partial\Upsilon} \|\vartheta - \vartheta_0\|$ y $\partial\Upsilon$ es la frontera del conjunto convexo Υ .

(ii) $G_n(\vartheta, \cdot) \leq G_n(\vartheta_0, \cdot) \quad \forall \vartheta \in \Upsilon \quad \forall n \in \mathbb{N}$.

Dem. Dado $s \in \mathcal{S}$, como H_s es $\mathcal{F}_{s \cup \partial s}$ -medible, por la Proposición 3.2.7(iii), la función

$$\gamma_s^{\vartheta_0} \left(\left(\phi^T H_s - \gamma_s^{\vartheta_0}(\phi^T H_s | \cdot) \right)^2 | \cdot \right)$$

es $\mathcal{F}_{\partial s}$ -medible. Luego por la Proposición 3.2.1(ii), esta función depende sólo de las configuraciones sobre ∂s , y podemos definir la función $g : E^{\partial s} \times \Upsilon \times \mathbb{R}^m \rightarrow \mathbb{R}$ dada por

$$g(\mathfrak{v}, \vartheta, \phi) := \gamma_s^{\vartheta_0} \left(\left(\phi^T H_s - \gamma_s^{\vartheta_0}(\phi^T H_s | y) \right)^2 | y \right),$$

siendo $y \in E^S$ tal que $\sigma_{\partial s}(y) = \mathfrak{v}$.

Consideremos $x \in E^S$ en el conjunto de probabilidad 1 dado en el Lema 3.8.5. Luego

$$\gamma_s^{\vartheta_0} \left(\left(\phi^T H_s - \gamma_s^{\vartheta_0}(\phi^T H_s | x) \right)^2 | x \right) = g(x_{\partial s}, \vartheta, \phi) \quad \forall s \in \mathcal{S} \quad \forall \vartheta \in \Upsilon \quad \forall \phi \in \mathbb{R}^m.$$

Por el Lema 3.8.5 y por (H.1), con π^{ϑ_0} -probabilidad 1, existe una constante $\zeta_1 > 0$ tal que $\inf_{\mathfrak{v}} \frac{N_n(\mathfrak{v})(x)}{\#\Lambda_n^\circ} > \zeta_1 \quad \forall n \gg 1, \mathfrak{v} \in E^{\partial s}$. Luego, por un razonamiento análogo al hecho en el Lema 3.8.8,

$$\phi^T \nabla^2 G_n(\vartheta, x) \phi = - \sum_{\mathfrak{v}} \frac{N_n(\mathfrak{v})(x)}{\#\Lambda_n^\circ} g(\mathfrak{v}, \vartheta, \phi) \leq -\zeta_1 \sum_{\mathfrak{v}} g(\mathfrak{v}, \vartheta, \phi) \quad \forall n \gg 1. \quad (3.8.10)$$

Como por (H.10), para cada \mathfrak{v} , la función $g(\mathfrak{v}, \cdot, \cdot)$ es continua, y como el conjunto $K = \{(\vartheta, \phi) : \|\vartheta - \vartheta_0\| \leq \varepsilon, \|\phi\| = 1\}$ es compacto, existe una constante real ζ_2 tal que

$$\zeta_2 \leq \sum_{\mathfrak{v}} g(\mathfrak{v}, \vartheta, \phi) \quad \forall (\vartheta, \phi) \in K.$$

Por el mismo argumento usado en el Lema 3.8.8, $G_n(\cdot, x)$ es estrictamente cóncava. Luego puede suponerse $\zeta_2 > 0$, y por (3.8.10), tomando $\delta = \zeta_1 \zeta_2 > 0$, se tiene que,

$$\limsup_{n \rightarrow \infty} \sup_{|\vartheta - \vartheta_0| \leq \varepsilon} \sup_{\phi \in \mathbb{R}^m: \|\phi\|=1} \phi^T \nabla^2 G_n(\vartheta, x) \phi < -\delta,$$

y el ítem (i) queda probado. Para la parte (ii), por la concavidad del \ln ,

$$\begin{aligned} \sum_{\xi} \pi_s^{\vartheta_0}(\xi|\nu) \ln \pi_s^{\vartheta}(\xi|\nu) - \sum_{\xi} \pi_s^{\vartheta_0}(\xi|\nu) \ln \pi_s^{\vartheta_0}(\xi|\nu) &= \sum_{\xi} \pi_s^{\vartheta_0}(\xi|\nu) \ln \frac{\pi_s^{\vartheta}(\xi|\nu)}{\pi_s^{\vartheta_0}(\xi|\nu)} \leq \\ &\leq \ln \sum_{\xi} \pi_s^{\vartheta_0}(\xi|\nu) \frac{\pi_s^{\vartheta}(\xi|\nu)}{\pi_s^{\vartheta_0}(\xi|\nu)} = \ln \sum_{\xi} \pi_s^{\vartheta}(\xi|\nu) = \ln 1 = 0, \end{aligned}$$

y por lo tanto vale (ii). □

Lema 3.8.10

$$\lim_{n \rightarrow \infty} \sup_{\|\vartheta - \vartheta_0\| \leq \varepsilon} |U_n(\vartheta, \cdot) - G_n(\vartheta, \cdot)| = 0 \quad \pi^{\vartheta_0}\text{-}p.p. \quad \forall \varepsilon \in (0, d(\vartheta_0, \partial Y))$$

Dem. Sea $\vartheta \in Y$ tal que $\|\vartheta - \vartheta_0\| \leq \varepsilon$. Luego

$$\begin{aligned} |U_n(\vartheta, \cdot) - G_n(\vartheta, \cdot)| &= \left| \sum_{\nu} \frac{N_n(\nu)}{\#\Lambda_n^{\circ}} \sum_{\xi} \left(\frac{N_n(\xi, \nu)}{N_n(\nu)} - \pi_s^{\vartheta_0}(\xi|\nu) \right) \ln \frac{\pi_s^{\vartheta}(\xi|\nu)}{\pi_s^{\vartheta_0}(\xi|\nu)} \right| \\ &\leq \sum_{\nu} \frac{N_n(\nu)}{\#\Lambda_n^{\circ}} \sum_{\xi} \left| \frac{N_n(\xi, \nu)}{N_n(\nu)} - \pi_s^{\vartheta_0}(\xi|\nu) \right| \left| \ln \frac{\pi_s^{\vartheta}(\xi|\nu)}{\pi_s^{\vartheta_0}(\xi|\nu)} \right| \\ &\leq \#E \sup_{\xi, \nu, \|\vartheta - \vartheta_0\| < \varepsilon} \left| \ln \frac{\pi_s^{\vartheta}(\xi|\nu)}{\pi_s^{\vartheta_0}(\xi|\nu)} \right| \sup_{\xi, \nu} \left| \frac{N_n(\xi, \nu)}{N_n(\nu)} - \pi_s^{\vartheta_0}(\xi|\nu) \right|. \end{aligned} \quad (3.8.11)$$

Como $\pi_s^{\vartheta}(\xi|\nu) > 0 \quad \forall \xi \in E \quad \forall \nu \in E^{\#\partial s} \quad \forall \vartheta \in Y$, como la cantidad de posibles configuraciones (ξ, ν) es finita, y como por (H.10) la función

$$\vartheta \mapsto \ln \frac{\pi_s^{\vartheta}(\xi|\nu)}{\pi_s^{\vartheta_0}(\xi|\nu)}$$

es continua sobre el conjunto $\{\vartheta : \|\vartheta - \vartheta_0\| < \varepsilon\}$, que está contenido en un compacto, se tiene que existe una constante $C > 0$ tal que

$$\sup_{\xi, \nu, \|\vartheta - \vartheta_0\| < \varepsilon} \left| \ln \frac{\pi_s^{\vartheta}(\xi|\nu)}{\pi_s^{\vartheta_0}(\xi|\nu)} \right| \leq C.$$

Luego, por (3.8.11), se tiene que

$$\limsup_{n \rightarrow \infty} \sup_{\|\vartheta - \vartheta_0\| \leq \varepsilon} |U_n(\vartheta, \cdot) - G_n(\vartheta, \cdot)| \leq C \#E \limsup_{n \rightarrow \infty} \sup_{\xi, \nu} \left| \frac{N_n(\xi, \nu)}{N_n(\nu)} - \pi_s^{\vartheta_0}(\xi|\nu) \right|.$$

Por el Lema 3.8.6,

$$\limsup_{n \rightarrow \infty} \sup_{\xi, \nu} \left| \frac{N_n(\xi, \nu)}{N_n(\nu)} - \pi_s^{\vartheta_0}(\xi|\nu) \right| \rightarrow 0 \quad \pi^{\vartheta_0}\text{-}p.p.$$

y por lo tanto queda demostrado el Lema. □

Lema 3.8.11

$$\liminf_{n \rightarrow \infty} \inf_{\|\vartheta - \vartheta_0\| = \varepsilon} (G_n(\vartheta_0, \cdot) - G_n(\vartheta, \cdot)) > 0 \quad \pi^{\vartheta_0}\text{-p.p.} \quad \forall \varepsilon \in (0, d(\vartheta_0, \partial Y))$$

Dem. Por el Lema 3.8.9, $G_n(\vartheta_0, \cdot) > G_n(\vartheta, \cdot) \forall \vartheta \neq \vartheta_0$ con probabilidad 1. Luego sólo hay que probar que con probabilidad 0, fijado arbitrariamente $\varepsilon > 0$, se verifica

$$\liminf_{n \rightarrow \infty} \inf_{\|\vartheta - \vartheta_0\| = \varepsilon} (G_n(\vartheta_0, \cdot) - G_n(\vartheta, \cdot)) = 0. \quad (3.8.12)$$

Fijemos $\varepsilon > 0$ y tomemos una realización $x \in E^S$ para el cual vale (3.8.12), entonces

$$\inf_{n \geq k} \inf_{\|\vartheta - \vartheta_0\| = \varepsilon} (G_n(\vartheta_0, x) - G_n(\vartheta, x)) = 0 \quad \forall k. \quad (3.8.13)$$

Fijemos arbitrariamente k , y consideremos $\eta = \frac{\delta \varepsilon^2}{2} > 0$, donde δ es la constante que aparece en la cota de la parte (i) del Lema 3.8.9. Luego por (3.8.13), existen $n \geq k$ y $\tilde{\vartheta} \in \{\vartheta \in \mathbb{R}^d : \|\vartheta - \vartheta_0\| = \varepsilon\}$, tales que

$$G_n(\vartheta_0, x) - G_n(\tilde{\vartheta}, x) < \eta. \quad (3.8.14)$$

Sea $A \subseteq \mathbb{R}$ un conjunto abierto tal que $[0, 1] \subseteq A$ y tal que $\zeta \tilde{\vartheta} + (1 - \zeta)\vartheta_0 \in Y \forall \zeta \in A$, y consideremos la función continua $f : \mathbb{R} \rightarrow \mathbb{R}$ dada por $f(\zeta) = G_n(\zeta \tilde{\vartheta} + (1 - \zeta)\vartheta_0, x)$. Luego,

$$\begin{aligned} f(0) &= G_n(\vartheta_0, x), \quad f(1) = G_n(\tilde{\vartheta}, x), \quad f'(0) = (\tilde{\vartheta} - \vartheta_0)^T \nabla G_n(\vartheta_0, x) = 0, \\ f''(\zeta) &= (\tilde{\vartheta} - \vartheta_0)^T \nabla^2 G_n(\zeta \tilde{\vartheta} + (1 - \zeta)\vartheta_0, x) (\tilde{\vartheta} - \vartheta_0) \quad \forall \zeta \in [0, 1]. \end{aligned}$$

Aquí pueden darse dos posibilidades: x pertenece al conjunto de probabilidad 1 para el cual vale la cota del apartado (i) del Lema 3.8.9, o no. Supongamos que pertenece a tal conjunto, entonces,

$$f''(\zeta) < -\delta \|\tilde{\vartheta} - \vartheta_0\|^2 = -\delta \varepsilon^2 = -2\eta \quad \forall \zeta.$$

Luego

$$f'(\zeta) = f'(0) + \int_0^\zeta f''(u) du = \int_0^\zeta f''(u) du < -2\eta \zeta \quad \forall \zeta > 0$$

Por lo que

$$f(\zeta) = f(0) + \int_0^\zeta f'(u) du = G_n(\vartheta_0, x) + \int_0^\zeta f'(u) du < G_n(\vartheta_0, x) - \eta \zeta^2 \quad \forall \zeta > 0.$$

En particular

$$G_n(\tilde{\vartheta}, x) = f(1) < G_n(\vartheta_0, x) - \eta,$$

lo cual contradice (3.8.14). Esto provino de suponer que x pertenecía al conjunto de probabilidad 1 para el cual vale la cota del apartado (i) del Lema 3.8.9. Por lo tanto, la probabilidad de que una realización del campo aleatorio en cuestión verifique (3.8.12) es 0. □

Lema 3.8.12

$$\liminf_{n \rightarrow \infty} \inf_{\|\vartheta - \vartheta_0\| = \varepsilon} (U_n(\vartheta_0, \cdot) - U_n(\vartheta, \cdot)) > 0 \quad \pi^{\vartheta_0}\text{-p.p.} \quad \forall \varepsilon \in (0, d(\vartheta_0, \partial Y)).$$

Dem. Sea x en los conjuntos de probabilidad 1 dados en los Lema 3.8.10 y 3.8.11. Dado $\varepsilon > 0$, probemos que

$$\liminf_{n \rightarrow \infty} \inf_{\|\vartheta - \vartheta_0\| = \varepsilon} (U_n(\vartheta_0, x) - U_n(\vartheta, x)) = \liminf_{n \rightarrow \infty} \inf_{\|\vartheta - \vartheta_0\| = \varepsilon} (G_n(\vartheta_0, x) - G_n(\vartheta, x)) =: \ell.$$

Sea $\eta > 0$ y $K = \{\vartheta \in Y : \|\vartheta - \vartheta_0\| \leq \varepsilon\}$. Luego por el Lema 3.8.11,

$$\exists k_1 \forall n \geq k_1 \forall \vartheta \in \partial K, \quad \ell - \frac{\eta}{3} < G_n(\vartheta_0, x) - G_n(\vartheta, x).$$

Por el Lema 3.8.10,

$$\exists k_2 \forall n \geq k_2 \forall \vartheta \in K, \quad |U_n(\vartheta, x) - G_n(\vartheta, x)| < \frac{\eta}{3}.$$

Luego, considerando $k_0 = \max\{k_1, k_2\}$, se tiene que

$$\begin{aligned} \forall n \geq k_0 \forall \vartheta \in \partial K, \quad U_n(\vartheta_0, x) - U_n(\vartheta, x) &= U_n(\vartheta_0, x) - G_n(\vartheta_0, x) + G_n(\vartheta_0, x) - G_n(\vartheta, x) + G_n(\vartheta, x) - U_n(\vartheta, x) \\ &> -\frac{\eta}{3} + \ell - \frac{\eta}{3} - \frac{\eta}{3} = \ell - \eta \end{aligned}$$

y el Lema queda probado. □

Teorema 3.8.1 (*Consistencia del Estimador de PMV*) Sean $\mathcal{S} = \mathbb{Z}^d$, $d \in \mathbb{N}$; ∂ un sistema de vecindades que satisface (H.2)-(H.5); $\{\Lambda_n\}_{n \in \mathbb{N}}$ como se definió en (H.12); E un conjunto de estados que verifica (H.1); $Y \subseteq \mathbb{R}^d$, $d \in \mathbb{N}$, un espacio de parámetros que verifica (H.6)-(H.7); $\vartheta_0 \in \text{Int}(Y)$; $\{\Phi^\vartheta\}_{\vartheta \in Y}$ una familia de potenciales tales que se verifican (H.8)-(H.10) y (H.13)-(H.14); $\pi^{\vartheta_0} \in \mathcal{G}(Y^{\vartheta_0})$, cuya existencia vale por (H.11), y supongamos que se verifica (H.15). Para cada $n \in \mathbb{N}$ y $x \in E^{\mathcal{S}}$, consideremos $\hat{\vartheta}_n(x) \in M_{\Lambda_n}(x)$. Entonces se verifican

$$(i) \quad \pi^{\vartheta_0}(\{x \in E^{\mathcal{S}} : \#M_{\Lambda_n}(x) = 1 \forall n \gg 1\}) = 1;$$

$$(ii) \quad \pi^{\vartheta_0}(\{x \in E^{\mathcal{S}} : \|\hat{\vartheta}_n(x) - \vartheta_0\| \rightarrow 0\}) = 1.$$

Dem. Consideremos $x \in E^{\mathcal{S}}$ que pertenece a los conjuntos de probabilidad 1 dados en los Lemas 3.8.5-3.8.12. Por el Lema 3.8.8, vale (i). Veamos ahora (ii): elijamos para cada n , $\hat{\vartheta}_n(x) \in M_{\Lambda_n}(x)$, y supongamos que

$$\limsup_{n \rightarrow \infty} \|\hat{\vartheta}_n(x) - \vartheta_0\| = \infty.$$

Luego existe una subsucesión $\{\hat{\vartheta}_{n_k}(x)\}_k$ tal que $\|\hat{\vartheta}_{n_k}(x) - \vartheta_0\| \uparrow \infty$, y por ende, dado $\varepsilon > 0$,

$$\exists k_1 \forall k \geq k_1, \quad \|\hat{\vartheta}_{n_k}(x) - \vartheta_0\| > \varepsilon.$$

Para cada k , sea

$$\vartheta_{k,\varepsilon} := \left(1 - \frac{\varepsilon}{\|\hat{\vartheta}_{n_k}(x) - \vartheta_0\|}\right) \vartheta_0 + \frac{\varepsilon}{\|\hat{\vartheta}_{n_k}(x) - \vartheta_0\|} \hat{\vartheta}_{n_k}(x),$$

que pertenece a Υ por (H.6), y verifica $\|\vartheta_{k,\varepsilon} - \vartheta_0\| = \varepsilon \forall k$. Luego,

$$U_{n_k}(\vartheta_0, x) - U_{n_k}(\vartheta_{k,\varepsilon}, x) \geq \inf_{\|\vartheta - \vartheta_0\| = \varepsilon} (U_{n_k}(\vartheta_0, x) - U_{n_k}(\vartheta, x)) \quad \forall k. \quad (3.8.15)$$

Por el Lema 3.8.8,

$$U_{n_k}(\vartheta_{k,\varepsilon}, x) > \left(1 - \frac{\varepsilon}{\|\widehat{\vartheta}_{n_k}(x) - \vartheta_0\|}\right) U_{n_k}(\vartheta_0, x) + \frac{\varepsilon}{\|\widehat{\vartheta}_{n_k}(x) - \vartheta_0\|} U_{n_k}(\widehat{\vartheta}_{n_k}(x), x)$$

Luego, por (3.8.15),

$$\begin{aligned} U_{n_k}(\vartheta_0, x) - \inf_{\|\vartheta - \vartheta_0\| = \varepsilon} (U_{n_k}(\vartheta_0, x) - U_{n_k}(\vartheta, x)) &> \\ &> \left(1 - \frac{\varepsilon}{\|\widehat{\vartheta}_{n_k}(x) - \vartheta_0\|}\right) U_{n_k}(\vartheta_0, x) + \frac{\varepsilon}{\|\widehat{\vartheta}_{n_k}(x) - \vartheta_0\|} U_{n_k}(\widehat{\vartheta}_{n_k}(x), x) \end{aligned}$$

lo cual implica que

$$\begin{aligned} \inf_{\|\vartheta - \vartheta_0\| = \varepsilon} (U_{n_k}(\vartheta_0, x) - U_{n_k}(\vartheta, x)) &< \frac{\varepsilon}{\|\widehat{\vartheta}_{n_k}(x) - \vartheta_0\|} (U_{n_k}(\vartheta_0, x) - U_{n_k}(\widehat{\vartheta}_{n_k}(x), x)) \\ &< U_{n_k}(\vartheta_0, x) - U_{n_k}(\widehat{\vartheta}_{n_k}(x), x) \quad \forall k \geq k_1. \end{aligned}$$

Como $\widehat{\vartheta}_{n_k}(x)$ maximiza a la función U_{n_k} , se tiene que

$$\inf_{\|\vartheta - \vartheta_0\| = \varepsilon} (U_{n_k}(\vartheta_0, x) - U_{n_k}(\vartheta, x)) < 0 \quad \forall k \geq k_1,$$

lo cual contradice al Lema 3.8.12, y por lo tanto la sucesión $\{\widehat{\vartheta}_n(x)\}$ es acotada.

Supongamos ahora que

$$\limsup_{n \rightarrow \infty} \|\widehat{\vartheta}_n(x) - \vartheta_0\| = \varepsilon > 0.$$

Luego, dado $\eta \in (0, \frac{\varepsilon}{2})$, $\exists n_0 \forall n \geq n_0$, $\|\widehat{\vartheta}_n(x) - \vartheta_0\| < \varepsilon + \eta$, y por ende

$$|U_n(\widehat{\vartheta}_n(x), x) - G_n(\widehat{\vartheta}_n(x), x)| \leq \sup_{\|\vartheta - \vartheta_0\| \leq \varepsilon + \eta} |U_n(\vartheta, x) - G_n(\vartheta, x)| \quad \forall n \geq n_0.$$

Luego por el Lema 3.8.10, $|U_n(\widehat{\vartheta}_n(x), x) - G_n(\widehat{\vartheta}_n(x), x)| \rightarrow 0$ y $|U_n(\vartheta_0, x) - G_n(\vartheta_0, x)| \rightarrow 0$. Como $U_n(\vartheta_0, x) \leq U_n(\widehat{\vartheta}_n(x), x)$, se tiene que

$$\begin{aligned} G_n(\vartheta_0, x) - G_n(\widehat{\vartheta}_n(x), x) &= G_n(\vartheta_0, x) - U_n(\widehat{\vartheta}_n(x), x) + U_n(\widehat{\vartheta}_n(x), x) - G_n(\widehat{\vartheta}_n(x), x) \\ &\leq G_n(\vartheta_0, x) - U_n(\vartheta_0, x) + U_n(\widehat{\vartheta}_n(x), x) - G_n(\widehat{\vartheta}_n(x), x) \end{aligned}$$

y por lo tanto

$$|G_n(\vartheta_0, x) - G_n(\widehat{\vartheta}_n(x), x)| \rightarrow 0. \quad (3.8.16)$$

Como ε es un punto de acumulación de la sucesión $\|\widehat{\vartheta}_n(x) - \vartheta_0\|$, existe una subsucesión $\|\widehat{\vartheta}_{n_k}(x) - \vartheta_0\|$ tal que

$$\varepsilon - \eta < \|\widehat{\vartheta}_{n_k}(x) - \vartheta_0\| < \varepsilon + \eta \quad \forall k.$$

Luego por el Lema 3.8.9(i), haciendo un razonamiento análogo al hecho en la prueba del Lema 3.8.11, se tiene que

$$|G_n(\vartheta_0, x) - G_n(\widehat{\vartheta}_{n_k}(x), x)| \geq \delta \|\widehat{\vartheta}_{n_k}(x) - \vartheta_0\|^2 > \delta(\varepsilon - \eta)^2 > \frac{\delta\varepsilon^2}{4} \quad \forall k,$$

lo cual contradice (3.8.16) y culmina la prueba del Teorema. \square

3.8.3. Caso General.

La prueba para el caso invariante se basa en el hecho de que una observación de $\gamma_\Lambda^{\vartheta_0}(\cdot|w)$ puede ser vista como $\#\Lambda$ observaciones de $\gamma_s^{\vartheta_0}(\cdot|w)$. Luego parece ser necesario pedir invarianza por traslaciones, para que de este modo las $\#\Lambda$ muestras sean de la misma distribución condicional. En este trabajo se pretende agregar hipótesis que suplanten la invarianza por traslaciones de las especificaciones, y generalicen los resultados. Debilitaremos la hipótesis de invarianza, eliminando solamente la suposición (H.14). Luego, en la presente sección, además de suponer válidas (H.1)-(H.13) y (H.15), consideraremos válidas (H.16) y más adelante agregaremos la condición de no correlación (H.17) entre h_s y H_s .

Una consecuencia inmediata del Lema 3.8.4 es que si

$$\tilde{\vartheta}_n(x) \in \tilde{M}_{\Lambda_n}(x) := \left\{ \vartheta \in \Upsilon : \tilde{U}_n(\vartheta, x) = \max_{\alpha \in \Upsilon} \tilde{U}_n(\alpha, x) \right\},$$

con

$$\tilde{U}_\Lambda(\vartheta, x) := \frac{1}{\#\Lambda^\circ} \sum_{s \in \Lambda^\circ} \ln \tilde{\rho}_s^\vartheta(x) = \frac{1}{\#\Lambda^\circ} \sum_{s \in \Lambda^\circ} (\vartheta^T H_s(x) - \ln \tilde{Z}_s^\vartheta(x)),$$

es un Estimador de PMV del modelo asociado, entonces se verifica el siguiente resultado

Lema 3.8.13 $\pi^{\vartheta_0}(\{x \in E^S : \|\tilde{\vartheta}_n(x) - \vartheta_0\| \rightarrow 0\}) = 1$.

Dem. Vale por el Lema 3.8.4 y el Teorema 3.8.1(ii). \square

Lema 3.8.14 *Se verifican:*

$$(i) \quad \nabla \ln Z_\Lambda^\vartheta = \gamma_\Lambda^\vartheta(H_\Lambda | \cdot) = \frac{\tilde{\gamma}_\Lambda^\vartheta(h_\Lambda H_\Lambda | \cdot)}{\tilde{\gamma}_\Lambda^\vartheta(h_\Lambda | \cdot)} \quad \forall \Lambda \in \mathfrak{S};$$

$$(ii) \quad \nabla U_\Lambda(\vartheta, \cdot) = \frac{1}{\#\Lambda^\circ} \sum_{s \in \Lambda^\circ} (H_s - \gamma_s^\vartheta(H_s | \cdot)) = \nabla \tilde{U}_\Lambda(\vartheta, \cdot) + \frac{1}{\#\Lambda^\circ} \sum_{s \in \Lambda^\circ} (\tilde{\gamma}_s^\vartheta(H_s | \cdot) - \gamma_s^\vartheta(H_s | \cdot)) \quad \forall \Lambda \in \mathfrak{S};$$

$$(iii) \quad \nabla^2 \tilde{U}_\Lambda(\vartheta, \cdot) = -\frac{1}{\#\Lambda^\circ} \sum_{s \in \Lambda^\circ} \tilde{\gamma}_s^\vartheta \left(H_s (H_s - \tilde{\gamma}_s^\vartheta(H_s | \cdot))^T | \cdot \right) \leq 0 \quad \forall \Lambda \in \mathfrak{S}.$$

$$(iv) \quad \nabla^2 U_\Lambda(\vartheta, \cdot) = -\frac{1}{\#\Lambda^\circ} \sum_{s \in \Lambda^\circ} \gamma_s^\vartheta \left(H_s (H_s - \gamma_s^\vartheta(H_s | \cdot))^T | \cdot \right) \leq 0 \quad \forall \Lambda \in \mathfrak{S}.$$

Dem. Valen por el Lema 3.8.7 y el Lema 3.8.1(iii). □

Como último supuesto para probar la consistencia del estimador de PMV, se necesita pedir la siguiente “no correlación asintótica” respecto del modelo invariante asociado:

$$(H.17) \quad \tilde{\pi}^{\vartheta_0}(h_s H_s) = \tilde{\pi}^{\vartheta_0}(h_s) \tilde{\pi}^{\vartheta_0}(H_s) \quad \forall s \in \mathcal{S}.$$

En la Sección 6.4 del Apéndice se prueba que (H.17) vale cuando se trabaja con el modelo a posteriori del mapa de clases presentado en la Sección 3.5.

Lema 3.8.15 Para cada $\vartheta \in \Upsilon$, $\nabla U_n(\vartheta, \cdot) - \nabla \tilde{U}_n(\vartheta, \cdot) \rightarrow 0$ π^{ϑ} -p.p.

Dem. Por el Lema 3.8.14 se tiene que

$$\nabla U_n(\vartheta, \cdot) - \nabla \tilde{U}_n(\vartheta, \cdot) = \frac{1}{\#\Lambda_n^\circ} \sum_{s \in \Lambda_n^\circ} \left(\tilde{\gamma}_s^\vartheta(H_s | \cdot) - \frac{\tilde{\gamma}_s^\vartheta(h_s H_s | \cdot)}{\tilde{\gamma}_s^\vartheta(h_s | \cdot)} \right) = \frac{1}{\#\Lambda_n^\circ} \sum_{s \in \Lambda_n^\circ} \varepsilon_s,$$

donde para cada $s \in \mathcal{S}$, $\varepsilon_s := \tilde{\gamma}_s^\vartheta(H_s | \cdot) - \tilde{\gamma}_s^\vartheta(\tilde{h}_s H_s | \cdot)$, con $\tilde{h}_s := \frac{h_s}{\tilde{\gamma}_s^\vartheta(h_s | \cdot)}$. Por lo tanto sólo se debe probar que

$$\frac{1}{\#\Lambda_n^\circ} \sum_{s \in \Lambda_n^\circ} \varepsilon_s \rightarrow 0 \quad \tilde{\pi}^{\vartheta}\text{-p.p.} \quad (3.8.17)$$

Como $\tilde{\gamma}_s^\vartheta(\tilde{h}_s | \cdot) \equiv 1 \quad \forall s \in \mathcal{S}$, se tiene que $\tilde{\pi}^{\vartheta}(\tilde{h}_s) = 1 \quad \forall s \in \mathcal{S}$. Luego por ser $\tilde{\pi}^{\vartheta} \in \mathcal{G}(\tilde{\gamma}^{\vartheta})$ y luego por (H.17),

$$\tilde{\pi}^{\vartheta}(\varepsilon_s) = \tilde{\pi}^{\vartheta}(\tilde{\gamma}_s^\vartheta(\tilde{h}_s H_s | \cdot) - \tilde{\gamma}_s^\vartheta(H_s | \cdot)) = \tilde{\pi}^{\vartheta}(\tilde{h}_s H_s - H_s) = \tilde{\pi}^{\vartheta}(\tilde{h}_s) \tilde{\pi}^{\vartheta}(H_s) - \tilde{\pi}^{\vartheta}(H_s) = 0 \quad \forall s \in \mathcal{S}.$$

Además, para cada $s \in \mathcal{S}$, la función ε_s es \mathcal{F}_{∂_s} -medible y acotada. Luego por el Corolario 3.6.1, vale (3.8.17), y finaliza la prueba. □

Ahora procedamos a probar (T.1) y (T.2) para el caso no invariante.

Teorema 3.8.2 (Consistencia del Estimador de PMV) Sean $\mathcal{S} = \mathbb{Z}^d$; ∂ un sistema de vecindades que satisface (H.2)-(H.5); $\{\Lambda_n\}_{n \in \mathbb{N}}$ como se definió en (H.12); E un conjunto de estados que verifica (H.1); $\Upsilon \subseteq \mathbb{R}^m$, $m \in \mathbb{N}$, un espacio de parámetros que verifica (H.6) y (H.7); $\vartheta_0 \in \text{Int}(\Upsilon)$; $\{\Phi^\vartheta\}_{\vartheta \in \Upsilon}$ una familia de potenciales tales que se verifican (H.8)-(H.10), (H.13), (H.16) y (H.17); $\pi^{\vartheta_0} \in \mathcal{G}(\Phi^{\vartheta_0})$ cuya existencia la garantiza (H.11), y supongamos que se verifica (H.15). Entonces se verifican

- (i) $\pi^{\vartheta_0}(\{x \in E^{\mathcal{S}} : U_n(\cdot, x) \text{ es estrictamente cóncava } \forall n \gg 1\}) = 1$;
- (ii) $\pi^{\vartheta_0}(\{x \in E^{\mathcal{S}} : \#M_{\Lambda_n}(x) = 1 \quad \forall n \gg 1\}) = 1$;
- (iii) $\pi^{\vartheta_0}(\{x \in E^{\mathcal{S}} : \|\hat{\vartheta}_n(x) - \vartheta_0\| \rightarrow 0\}) = 1$.

Dem. Veamos (i): por el Lema 3.8.4 y el Lema 3.8.5, se tiene que

$$\liminf_{n \rightarrow \infty} \frac{N_n(\zeta)}{\#\Lambda_n^\circ} > 0 \quad \pi^{\vartheta_0}\text{-p.p.} \quad \forall \zeta \in E^{\partial_s},$$

Sea $x \in E^{\mathcal{S}}$ en este conjunto. Luego por el Lema 3.8.14(iv),

$$\nabla^2 U_n(\vartheta_0, x) = -\frac{1}{\#\Lambda_n^{\circ}} \sum_{s \in \Lambda_n^{\circ}} \gamma_s^{\vartheta_0} \left(H_s(H_s - \gamma_s^{\vartheta_0}(H_s|x))^T |x \right) \leq 0.$$

Luego, $U_n(\cdot, x)$ es cóncava y

$$\phi^T \nabla^2 U_n(\vartheta_0, x) \phi = -\frac{1}{\#\Lambda_n^{\circ}} \sum_{s \in \Lambda_n^{\circ}} \gamma_s^{\vartheta_0} \left(\left(\phi^T H_s - \gamma_s^{\vartheta_0}(\phi^T H_s|x) \right)^2 |x \right) \leq 0 \quad \forall \phi \in \mathbb{R}^m.$$

Sea $\vartheta \neq \vartheta_0$ y supongamos que $\phi^T \nabla^2 U_n(\vartheta, x) \phi = 0$ para algún ϕ no nulo. Entonces,

$$\phi^T H_s(\xi x^s) = \gamma_s^{\vartheta}(\phi^T H_s|x) \quad \forall \xi \in E \quad \forall s \in \mathcal{S}.$$

En particular, $H_s(\cdot x^s)$ es constante $\forall s \in \mathcal{S}$, y esto implica que

$$\tilde{\rho}_s^{\vartheta}(\xi x^s) = \frac{\exp\{\vartheta^T H_s(\xi x^s)\}}{\sum_{\tilde{\xi} \in E} \exp\{\vartheta^T H_s(\tilde{\xi} x^s)\}} = \frac{1}{\#E} \quad \forall \xi \in E \quad \forall s \in \mathcal{S}, \quad (3.8.18)$$

expresión que no depende de ϑ . Luego, (3.8.18) contradice (H.15) por un argumento análogo al utilizado en la prueba del Lema 3.8.8. Concluyendo que $U_n(\cdot, x)$ es estrictamente cóncava $\forall n \gg 1$, y en consecuencia valen (i) y (ii).

Veamos ahora (iii): consideremos $x \in E^{\mathcal{S}}$ en el conjunto de probabilidad 1 al cual pertenecen los elementos que verifican las condiciones dadas en el Teorema 3.8.1, los Lemas 3.8.8, 3.8.12 y 3.8.15. Para cada $n \in \mathbb{N}$, elijamos $\hat{\vartheta}_n(x) = \arg \max_{\vartheta} U_n(\vartheta, x)$ y $\tilde{\vartheta}_n(x) = \arg \max_{\vartheta} \tilde{U}_n(\vartheta, x)$. Luego por la consistencia en el caso invariante

$$\|\tilde{\vartheta}_n(x) - \vartheta_0\| \rightarrow 0,$$

y por el Lema 3.8.15

$$\nabla \tilde{U}_n(\hat{\vartheta}_n(x), x) = \nabla \tilde{U}_n(\tilde{\vartheta}_n(x), x) - \nabla U_n(\hat{\vartheta}_n(x), x) \rightarrow 0. \quad (3.8.19)$$

Como sólo debemos probar que $\|\hat{\vartheta}_n(x) - \vartheta_0\| \rightarrow 0$, supongamos que esto no vale. Luego existe $\varepsilon > 0$ tal que para cada $n \in \mathbb{N}$, existe $k_n \geq n$ tal que $\|\hat{\vartheta}_{k_n}(x) - \vartheta_0\| \geq 2\varepsilon$. Teniendo en cuenta (H.6), considerando

$$\bar{\vartheta}_n := \vartheta_0 + \varepsilon \frac{\hat{\vartheta}_n(x) - \vartheta_0}{\|\hat{\vartheta}_n(x) - \vartheta_0\|} \in \Upsilon,$$

se tiene que $\|\bar{\vartheta}_n - \vartheta_0\| = \varepsilon$. Luego por el Lema 3.8.8, se tiene que

$$\frac{|\tilde{U}_{k_n}(\vartheta_0, x) - \tilde{U}_{k_n}(\hat{\vartheta}_{k_n}(x), x)|}{\|\vartheta_0 - \hat{\vartheta}_{k_n}(x)\|} \geq \frac{|\tilde{U}_{k_n}(\vartheta_0, x) - \tilde{U}_{k_n}(\bar{\vartheta}_{k_n}, x)|}{\varepsilon} \quad (3.8.20)$$

En el Lema 3.8.12 se probó que

$$l := \liminf_{n \rightarrow \infty} \inf_{\|\vartheta - \vartheta_0\| = \varepsilon} (\tilde{U}_n(\vartheta_0, x) - \tilde{U}_n(\vartheta, x)) > 0. \quad (3.8.21)$$

Supongamos que

$$\liminf_{n \rightarrow \infty} |\tilde{U}_{k_n}(\vartheta_0, x) - \tilde{U}_{k_n}(\bar{\vartheta}_{k_n}, x)| < l/3.$$

Luego para cada $m \in \mathbb{N}$, existe $k_{n_m} \geq m$ tal que

$$\inf_{\|\vartheta - \vartheta_0\| = \varepsilon} (\tilde{U}_{k_{n_m}}(\vartheta_0, x) - \tilde{U}_{k_{n_m}}(\vartheta, x)) \leq \tilde{U}_{k_{n_m}}(\vartheta_0, x) - \tilde{U}_{k_{n_m}}(\tilde{\vartheta}_{k_{n_m}}, x) < \iota/2,$$

lo cual contradice (3.8.21). Implicando que

$$\liminf_{n \rightarrow \infty} |\tilde{U}_{k_n}(\vartheta_0, x) - \tilde{U}_{k_n}(\tilde{\vartheta}_{k_n}, x)| \geq \iota/3.$$

Luego por (3.8.20)

$$\liminf_{n \rightarrow \infty} \frac{|\tilde{U}_{k_n}(\vartheta_0, x) - \tilde{U}_{k_n}(\hat{\vartheta}_{k_n}(x), x)|}{\|\vartheta_0 - \hat{\vartheta}_{k_n}(x)\|} \geq \liminf_{n \rightarrow \infty} \frac{|\tilde{U}_{k_n}(\vartheta_0, x) - \tilde{U}_{k_n}(\tilde{\vartheta}_{k_n}, x)|}{\varepsilon} \geq \frac{\iota}{3\varepsilon}.$$

Por el Lema 3.8.8,

$$|\nabla \tilde{U}_{k_n}(\hat{\vartheta}_{k_n}(x), x)| \geq \frac{|\tilde{U}_{k_n}(\vartheta_0, x) - \tilde{U}_{k_n}(\hat{\vartheta}_{k_n}(x), x)|}{\|\vartheta_0 - \hat{\vartheta}_{k_n}(x)\|}$$

y por ende

$$\liminf_{n \rightarrow \infty} |\nabla \tilde{U}_{k_n}(\hat{\vartheta}_{k_n}(x), x)| \geq \liminf_{n \rightarrow \infty} \frac{|\tilde{U}_{k_n}(\vartheta_0, x) - \tilde{U}_{k_n}(\hat{\vartheta}_{k_n}(x), x)|}{\|\vartheta_0 - \hat{\vartheta}_{k_n}(x)\|} \geq \frac{\iota}{3\varepsilon},$$

lo cual contradice (3.8.19), y finaliza la prueba del Teorema. \square

De este modo se han probado dos resultados de consistencia fuerte para estimadores de PMV. En una primera etapa se extendió la prueba realizada por Geman y Graffigne en [4] a una familia de modelos exponenciales invariantes más grande. Además, se probó la consistencia fuerte para el caso en el que la especificación del modelo no es invariante por traslaciones. Para ello se reemplazó el supuesto de invarianza por los supuestos (H.16) de equivalencia microscópica y (H.17) de no correlación entre las funciones h_s y H_s .

3.9. Simulación

En la presente Sección abordamos el problema de obtener realizaciones de un determinado modelo. Esto permitirá analizar si el modelo de Potts es adecuado para representar al mapa de clases de una imagen. Además, permitirá analizar bajo simulación la performance de los estimadores a muestra finita.

Como se mencionó en la Sección 3.7, obtener una observación de π^{ϑ_0} , $\vartheta_0 \in \text{Int}(\Upsilon)$, es imposible debido a que \mathcal{S} es infinito. Sin embargo, fijando arbitrariamente $\Lambda \in \mathfrak{S}$ y $w \in E^{\mathcal{S}}$, las leyes $\gamma_{\Lambda}^{\vartheta_0}(\cdot|w)$ poseen el mismo parámetro, y puede ser muestreado por medio de algoritmos clásicos. Para ello se debe obtener una observación $x_{\Lambda} \in E^{\Lambda}$ con probabilidad proporcional a

$$\gamma_{\Lambda}^{\vartheta_0}(\{x_{\Lambda} w^{\Lambda}\}|w) = \rho_{\Lambda}^{\vartheta_0}(x_{\Lambda} w^{\Lambda}) = \pi_{\Lambda}^{\vartheta_0}(x_{\Lambda}|w^{\Lambda}) = \frac{\exp\{-H_{\Lambda}^{\vartheta_0}(x_{\Lambda} w^{\Lambda})\}}{Z_{\Lambda}^{\vartheta_0}(x_{\Lambda} w^{\Lambda})}, \quad (3.9.22)$$

y luego $x_{\Lambda} w^{\Lambda} \in E^{\mathcal{S}}$ es una observación de $\gamma_{\Lambda}^{\vartheta_0}(\cdot|w)$, o equivalentemente, $x_{\Lambda} \in E^{\Lambda}$ es una observación de $\pi_{\Lambda}^{\vartheta_0}(\cdot|w^{\Lambda})$. Entonces, el objetivo de la presente Sección es mostrar cómo podemos obtener una realización x_{Λ} .

En la presente Sección, trabajaremos en $\mathcal{S} = \mathbb{Z}^2$ bajo el supuesto (H.1) con $E = \{\ell_1, \ell_2, \dots, \ell_L\}$, $L = \#E \geq 2$, y supondremos que Λ es una grilla finita cuadrada, que en las simulaciones se consideró de tamaño 128×128 .

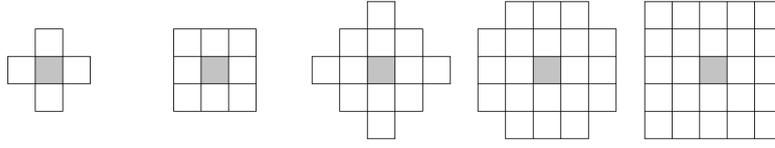


Figura 3.1: Vecindarios de orden 1, 2, 4, 5 y 8.

En la Figura 3.1 se muestran los vecindarios de orden 1, 2, 4, 5 y 8, los cuales satisfacen las condiciones (H.2)-(H.5). De ellos, obtendremos simulaciones considerando solamente los vecindarios de orden 2.

Bajo estas consideraciones, la presente Sección se organizará como sigue: comenzaremos explicando el clásico algoritmo Gibbs Sampler (GS) [6], que permite obtener una realización de las familias exponenciales definidas en (H.10). Posteriormente particularizaremos este algoritmo al modelo de Potts definido en la Sección 3.5, y se mostrarán realizaciones del mismo para diversos parámetros. Por último, se explicará y se obtendrán realizaciones del modelo de Potts isotrópico y sin campo externo por medio del algoritmo Swendsen-Wang [35, 36].

3.9.1. Gibbs Sampler (GS).

A pesar de que Λ y E son finitos, las posibles configuraciones de un campo X sobre Λ son $\#E^{\#\Lambda}$, cantidad que en la mayoría de los casos es inmanejable en tiempo finito. Luego, es necesario determinar un procedimiento iterativo que sea capaz de obtener una realización de estos campos, y el algoritmo Gibbs Sampler (GS) desarrollado por Geman y Geman en [6], realiza esta tarea.

La idea básica de este algoritmo es actualizar el estado de cada sitio, de a uno a la vez, teniendo en cuenta las etiquetas de su vecindad. Este proceso de actualización o relajación en la jerga del campo, se realiza siguiendo un esquema de visita, el cual no es más que una sucesión infinita de elementos de Λ que contiene infinitas veces a cada sitio. En particular, en la tesis se trabaja con un esquema de visita en el que no se revisita un píxel hasta que no se hayan visitado todos los píxeles de la grilla. Esto es, un esquema de visita que consiste en una sucesión de barridas, donde cada barrida consiste en visitar todos los píxeles de la grilla siguiendo un orden $s_1, s_2, \dots, s_{\#\Lambda}$ determinado.

Dado un mapa de inicio $x^{(0)} \in E^\Lambda$, luego de cada barrida, obtenemos una nueva configuración de etiquetas, que barrida tras barrida, genera una secuencia $x^{(1)}, x^{(2)}, \dots, x^{(n)}, \dots$ de elementos de E^Λ . Bajo determinadas leyes de transición esta sucesión converge a una realización de $\pi_\Lambda^{\vartheta_0}(\cdot | w^\Lambda)$. Como se dijo, cada transición de una realización a otra está conformada por un esquema de actualizaciones de etiquetas, las cuales deben ser realizadas de acuerdo a la verosimilitud de cada etiqueta condicional a las etiquetas vecinas. Esto es, supongamos que tenemos una configuración $x_\Lambda \in E^\Lambda$ y que deseamos actualizar la etiqueta que se encuentra sobre un determinado sitio s . Luego, se asignará la etiqueta $\ell \in E$ al sitio s con probabilidad dada por:

$$\pi_s^{\vartheta_0}(\ell | x_\Lambda^s w^\Lambda) \propto \exp \left\{ - \sum_{\Lambda \in \mathfrak{G}(s)} \Phi_\Lambda^{\vartheta_0}(\ell x_\Lambda^s w^\Lambda) \right\}.$$

En esta expresión, los píxeles del contorno de Λ tienen vecinos fuera Λ , y por ende actualizan su estado utilizando las entradas de w .

La convergencia de este algoritmo está probada en el Teorema A de [6]. En resumen, para obtener una realización del modelo sobre E^Λ se siguen los siguientes pasos:

- 1) obtener una realización de inicio $x^{(0)}$ de $X^{(0)}$.

- 2) elegir un esquema de visita adecuado.
- 3) realizar un barrido (visita de todos los pixeles), sorteando aleatoriamente un estado en cada visita, con probabilidad proporcional a

$$\exp \left\{ - \sum_{\Lambda \in \mathfrak{G}(s)} \Phi_{\Lambda}^{\partial_0}(\ell x_{\Lambda}^s w^{\Lambda}) \right\}$$

para cada $\ell \in E$, siendo s el pixel que se visita y $x \in E^{\Lambda}$ la configuración sobre E^{Λ} previa a la visita.

- 4) Se realizan tantas barridas como se necesiten para alcanzar algún criterio de convergencia.

3.9.2. Potts Sampler (PS).

Como se definió en la Sección 3.5, el modelo de Potts isotrópico y sin campo externo viene dado por el potencial conformado por las funciones

$$\Phi_{\Lambda}^{\beta}(x) := \begin{cases} -\beta & \text{si } \#\Lambda = 2, \Lambda = \{s, t\}, x_s = x_t, t \in \partial s \\ 0 & \text{caso contrario.} \end{cases} \quad \forall \Lambda \in \mathfrak{G} \quad (3.9.23)$$

Luego

$$\pi_s^{\beta}(\ell | x^s) \propto \exp\{\beta H_s(\ell)\},$$

donde $H_s(\ell) = \#\{t \in \partial s : x_s = \ell\}$.

En la presente sección se mostrará el clásico algoritmo Potts Sampler (PS), el cual tiene como objetivo obtener una realización de este modelo, y es un caso particular del algoritmo Gibbs Sampler (GS) visto en la Sección 3.9.1.

Particularizando el algoritmo visto en la Sección 3.9.1, con el potencial definido en (3.9.23) y con $X^{(0)}$ con distribución uniforme sobre E^{Λ} , se obtiene el siguiente algoritmo simulador de modelos de Potts.

- 1) Simular uniformemente una etiqueta para cada pixel de Λ .
- 2) Considerar el esquema de visita diagramado en la Figura 5.1.
- 3) Realizar una corrida actualizando aleatoriamente una etiqueta en cada visita, con probabilidad proporcional a

$$\pi_s(\ell | x_{\Lambda}^s w^{\Lambda}) \propto \exp\{\beta H_s(\ell)\}$$

para cada $\ell \in E$, siendo $s \in \Lambda$ el pixel que se visita y $x_{\Lambda} \in E^{\Lambda}$ la configuración sobre E^{Λ} previa a la visita.

- 4) Se realizan corridas hasta alcanzar algún criterio de convergencia.

En la Figura 3.2 se muestran realizaciones de 100 iteraciones de PS sobre Λ de tamaño 128×128 para distintos valores de β y L , y vecindades de orden 2.

3.9.3. Swendsen-Wang (SW).

La idea del algoritmo Swendsen-Wang [35, 36] es simular una observación $y = x_{\Lambda}$ de un campo aleatorio $Y = X_{\Lambda} \in E^{\Lambda}$ definido sobre un espacio de medida (Ω, \mathcal{A}, P) con distribución $\pi_{\Lambda}^{\beta}(\cdot | w^{\Lambda})$ dada en (3.9.22). Para ello se utiliza un arreglo aleatorio $V = \{V_{st} : s \in \Lambda, t \in \partial s\}$ auxiliar en el que, dado $Y = y$, sus componentes son independientes y tienen distribución condicional uniforme dada por

$$P(V_{st} = v_{st} | Y = y) \propto \exp\{-\beta 1(y_s = y_t)\} 1(0 \leq v_{st} \leq \exp\{\beta 1(y_s = y_t)\}), \quad (3.9.24)$$

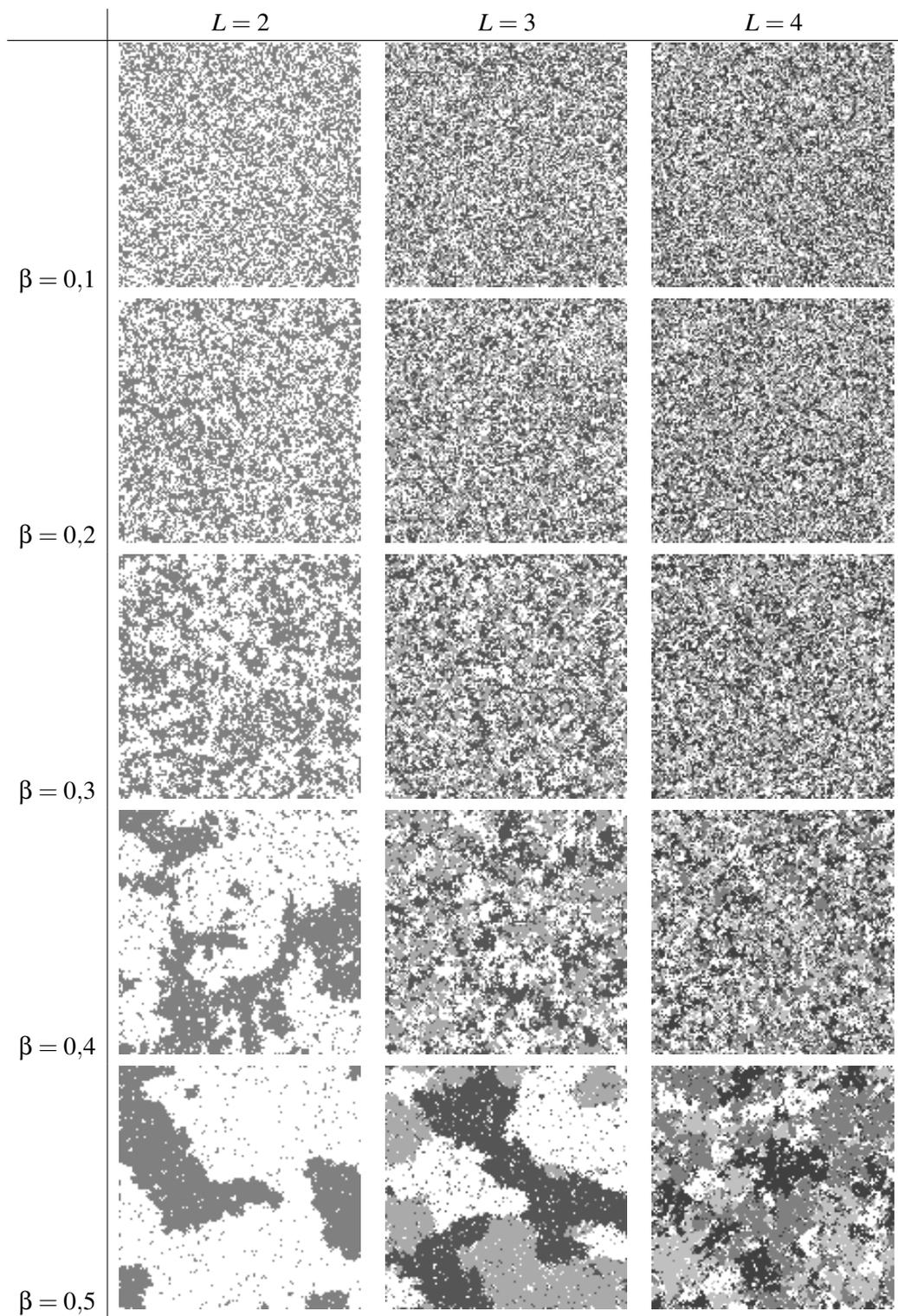


Figura 3.2: Realizaciones de 100 iteraciones de PS de tamaño 128×128 para distintos valores de β y L , y vecindades de orden 2.

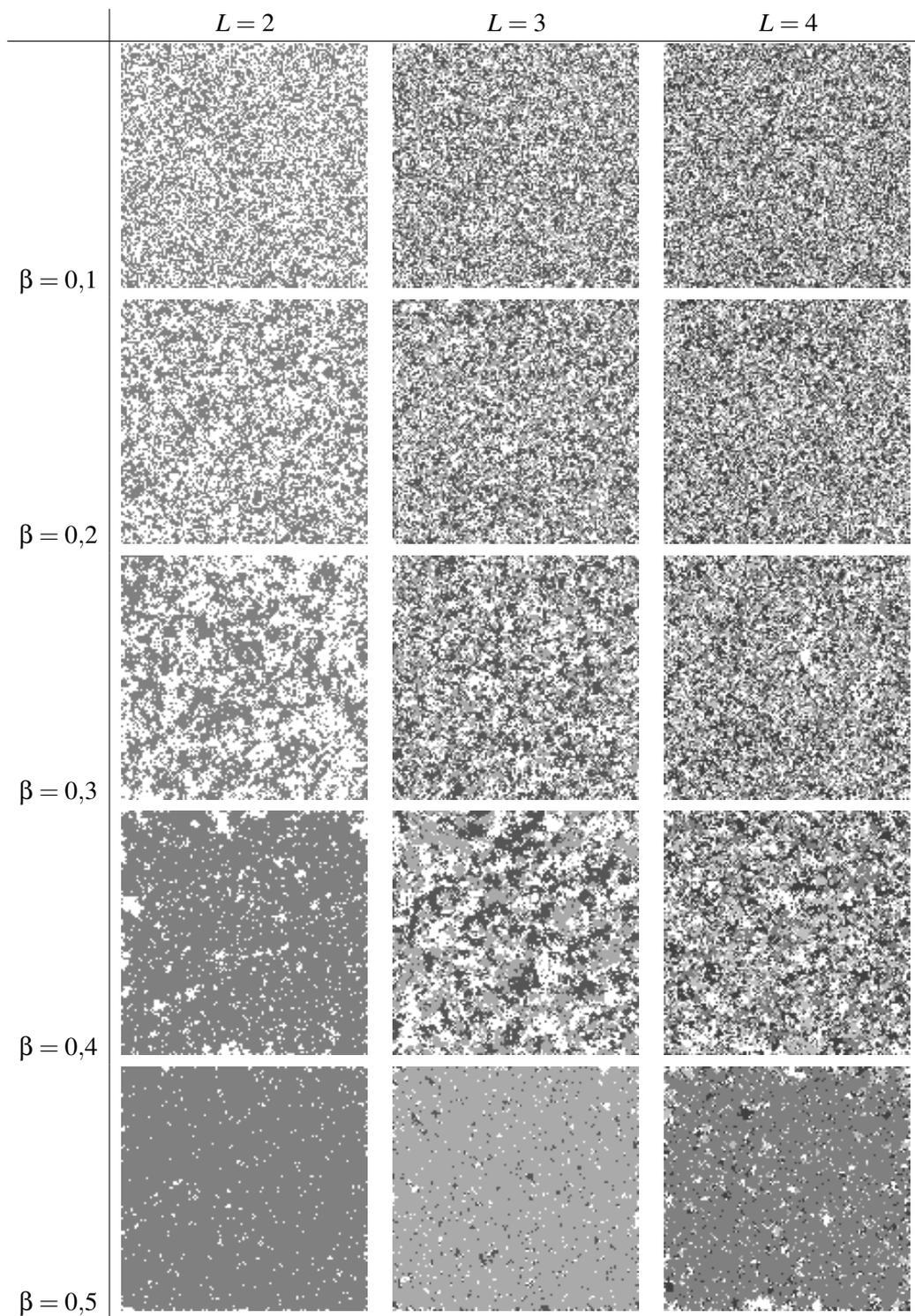


Figura 3.3: Realizaciones de 100 iteraciones de SW de tamaño 128×128 para distintos valores de β y L , y vecindades de orden 2.

donde V_{st} se denomina variable enlace entre Y_s y Y_t , y fijamos $y_s = w_s \forall s \in \partial\Lambda$. Luego

$$\begin{aligned}
P(Y = y|V = v) &\propto \pi_\Lambda^\beta(y|w^\Lambda) \prod_{s \in \Lambda, t \in \partial s} P(V_{st} = v_{st}|Y = y) \\
&\propto \prod_{s \in \Lambda, t \in \partial s} \exp\{\beta 1(y_s = y_t)\} \prod_{s \in \Lambda, t \in \partial s} \exp\{-\beta 1(y_s = y_t)\} 1(0 \leq v_{st} \leq \exp\{\beta 1(y_s = y_t)\}) \\
&= \prod_{s \in \Lambda, t \in \partial s} 1(0 \leq v_{st} \leq \exp\{\beta 1(y_s = y_t)\}) \\
&= \prod_{s \in \Lambda, t \in \partial s} 1\left(\frac{\ln v_{st}}{\beta} \leq 1(y_s = y_t)\right) \tag{3.9.25}
\end{aligned}$$

Ahora, por (3.9.24), dado $Y = y$ es sencillo simular U . Analicemos como simular Y dado $V = v$: sea $\Omega_v := \{y \in E^\Lambda : P(Y = y|V = v) > 0\}$ y sea $y \in \Omega_v$ una posible realización de $Y|V = v$. Luego por (3.9.25), se tiene que

$$\frac{\ln v_{st}}{\beta} \leq 1(y_s = y_t) \quad \forall s \in \Lambda, t \in \partial s, \tag{3.9.26}$$

pero esto implica que

$$P(Y = y|V = v) = \frac{\prod_{s \in \Lambda, t \in \partial s} 1\left(\frac{\ln v_{st}}{\beta} \leq 1(y_s = y_t)\right)}{\sum_{z \in E^\Lambda} \prod_{s \in \Lambda, t \in \partial s} 1\left(\frac{\ln v_{st}}{\beta} \leq 1(z_s = z_t)\right)} = \frac{1}{\sum_{z \in E^\Lambda} \prod_{s \in \Lambda, t \in \partial s} 1\left(\frac{\ln v_{st}}{\beta} \leq 1(z_s = z_t)\right)}$$

lo cual significa que dado $V = v$, Y distribuye uniformemente en Ω_v . Veamos entonces como está conformado este conjunto Ω_v de posibles realizaciones: siguiendo con $y \in \Omega_v$, por (3.9.26), si $v_{st} > 1$ entonces $y_s = y_t$; pero si $v_{st} \leq 1$, entonces y_s y y_t pueden tomar cualquier etiqueta. Luego

$$\Omega_v = \{y \in E^\Lambda : y_s = y_t \forall s, t \in \Lambda \text{ con } v_{st} > 1\}.$$

Note que lo único que interesa de las V_{st} es si son o no más grandes que 1. Luego, estas variables pueden ser vistas como variables binarias: vínculo o no vínculo; donde, por (3.9.24) valen

$$P(V_{st} > 1|Y_s = Y_t) = \frac{\exp\{\beta\} - 1}{\exp\{\beta\}} = 1 - \exp\{-\beta\}$$

y

$$P(V_{st} > 1|Y_s \neq Y_t) = 0. \tag{3.9.27}$$

Diremos que una configuración $x_\Lambda \in E^\Lambda$ tiene un agrupamiento en $\Delta \subset \mathcal{S}$ si toma el mismo estado sobre todos los sitios de Δ y si Δ es conexo maximal; esto es: dados $s, t \in \Delta$ distintos, se verifica que $x_s = x_t$; existen $s_0 = s, s_1, s_2, \dots, s_{n-1}, s_n = t \in \Delta$ tales que $s_{i-1} \in \partial s_i \forall i = 1, 2, \dots, n$; y no existe $\tilde{s} \in \Lambda \cap \partial\Delta$ tal que $x_s = x_{\tilde{s}}$.

En resumen, para simular una realización de (3.5.4), dado una configuración inicial $y^{(0)}$ se iteran hasta converger los siguientes dos pasos:

- 1) dada una configuración $Y = y$, se simula el campo V siguiendo (3.9.24).

- 2) dados los vínculos $V = v$, se particiona Λ en la familia $\mathcal{P}_u = \{S_1, S_2, \dots, S_K\}$ de subconjuntos de Λ , de modo tal que para cada $1 \leq k < k' \leq K$ se verifique

$$s, t \in S_k \Rightarrow v_{st} > 1 \quad \text{y} \quad s \in S_k, t \in S_{k'} \Rightarrow v_{st} \leq 1.$$

Luego se asigna a cada agrupamiento una etiqueta, eligiéndola uniformemente del conjunto E .

Note que en el paso 2 de este proceso iterativo contiene dos procesos: división y unión. En el proceso de división, cada agrupamiento del mapa puede llegar a partitionarse si en el paso 1 se desvinculan sus píxeles de modo tal que la partición \mathcal{P}_u envíe cada subdivisión a distintos agrupamientos. En el proceso de unión, cada par de agrupamientos adyacentes pueden llegar a unirse si les toca la misma etiqueta al hacer el sorteo uniforme de etiquetas. Luego si β no es grande, estos dos procesos se compensan, y no se producen mapas con sólo una etiqueta.

En la Figura 3.3 se muestran realizaciones de 100 iteraciones de SW de tamaño 128×128 para distintos valores de β y L , y vecindades de orden 2.

Para valores pequeños de β , ambos simuladores obtienen realizaciones parecidas a simple vista, pero a medida que crece β por encima de 0.3, SW presenta una transición más veloz entre mapas con pequeños agrupamientos a mapas en los que predomina una clase. En este sentido, la transición de PS es más lenta, presentando grandes agrupamientos para un rango de valores de β , antes de presentar mapas con prácticamente una clase.

Los programas de simulación realizados en esta tesis forman parte de un toolbox de segmentación markoviana que está siendo preparado por la Dra Flesia, y que sigue los postulados del programa internacional de Investigación Reproducible. Dicho toolbox se encuentra accesible en el sitio web <http://www.famaf.unc.edu.ar/~flesia/>

Capítulo 4

Estimación en el Modelo de Potts

4.1. Introducción

En la Sección 3.5 se observó que el modelo de Potts puede ser seleccionado para representar la distribución a priori del mapa de clases de una imagen, debido a que permite incorporar información contextual que vincule las etiquetas posicionadas sobre píxeles vecinos, de un modo sencillo y práctico. En el próximo Capítulo veremos que esto permite la obtención de segmentaciones homogéneas de imágenes, reduciendo las posibilidades de obtener un pixel clasificado aisladamente. Como los segmentadores contextuales requieren de una estimación del modelo para poder funcionar, es importante analizar una metodología eficiente para tal fin.

En la Sección 3.5 se mostró que si X es un campo aleatorio cuya distribución se rige de acuerdo a un modelo de Potts isotrópico sin campo externo, entonces sus densidades condicionales tienen la forma

$$\pi_{\Lambda}^{\beta}(x_{\Lambda}|w^{\Lambda}) \propto \exp\{\beta H_{\Lambda}(x_{\Lambda}w^{\Lambda})\} \quad \forall \Lambda \in \mathfrak{S}, \quad (4.1.1)$$

donde $w \in E^{\mathcal{S}}$ es una configuración prefijada arbitrariamente, y

$$H_{\Lambda}(y) = \#\{\Delta \in \mathfrak{S}(\Lambda) : \#\Delta = 2, \Delta = \{s, t\}, y_s = y_t, t \in \partial s\} \quad \forall y \in E^{\mathcal{S}},$$

Esto es, $H_{\Lambda}(y)$ es la cantidad de pares de vecinos con la misma etiqueta, con al menos uno de ellos en Λ . Además se mostró que este modelo posee especificaciones invariantes por traslaciones.

En particular, las distribuciones condicionales del modelo a priori sobre cada $s \in \mathcal{S}$, tienen la forma

$$\pi_s^{\beta}(x_s|x_{\Lambda \setminus \{s\}}w^{\Lambda}) \propto \exp\{\beta H_s(x_s)\},$$

donde $H_s(x_s) = \#\{t \in \partial s : x_t = x_s\}$ es una simplificación notacional, que permite independizarse simbólicamente de w , y hará mucho más sencilla la explicación de los estimadores del modelo.

Teniendo en cuenta la Proposición 3.2.7(ii), si notamos con $\tilde{\gamma}^{\beta}$ a la especificación correspondiente al modelo a priori, entonces la esperanza condicional de la función H_{Λ} , dada una observación w^{Λ} fuera de Λ , viene dada por

$$\tilde{\gamma}_{\Lambda}^{\beta}(H_{\Lambda}|w) = \sum_{x_{\Lambda} \in E^{\Lambda}} H_{\Lambda}(x_{\Lambda}w^{\Lambda}) \frac{\exp\{\beta H_{\Lambda}(x_{\Lambda}w^{\Lambda})\}}{\tilde{Z}_{\Lambda}^{\beta}(x_{\Lambda}w^{\Lambda})},$$

donde

$$\tilde{Z}_{\Lambda}^{\beta}(x_{\Lambda}w^{\Lambda}) = \sum_{y_{\Lambda} \in E^{\Lambda}} \exp\{\beta H_{\Lambda}(y_{\Lambda}w^{\Lambda})\}.$$

De este modo, la versión isotrópica y sin campo externo del modelo de Potts posee un único parámetro β a estimar. Este parámetro es conocido como parámetro de suavidad del modelo, y tiene la función de regular el tamaño de los agrupamientos de puntos con la misma etiqueta, ya que mientras más grande es β , los mapas más homogéneos son más predominantes probabilísticamente en el modelo. En la Literatura se han desarrollado diversos estimadores para modelos Markovianos en general, y para el modelo de Potts en particular.

Comencemos viendo el estimador más clásico, esto es, el Estimador de Máxima Verosimilitud (MV) de β respecto del modelo a priori. En este caso, como se mencionó en la Sección 3.1 y teniendo en cuenta la \mathcal{F}_Λ -medibilidad de $U_{\Lambda^\circ}^{MV}(\beta, \cdot)$, se tiene que

$$U_{\Lambda}^{MV}(\beta, x) = \ln \pi_{\Lambda^\circ}^\beta(x_{\Lambda^\circ} | x_{\Lambda \setminus \Lambda^\circ}) = \ln \pi_{\Lambda^\circ}^\beta(x_{\Lambda^\circ} | x_{\Lambda \setminus \Lambda^\circ} w^\Lambda) = \beta H_{\Lambda^\circ}(x_{\Lambda} w^\Lambda) - \ln \tilde{Z}_{\Lambda^\circ}^\beta(x_{\Lambda} w^\Lambda),$$

con $w^\Lambda \in E^{S \setminus \Lambda}$ prefijada arbitrariamente, que no influye en el proceso de estimación y que es agregada por cuestiones de coherencia notacional. De este modo, trabajar en Λ° , en lugar de trabajar sobre Λ , permite independizarnos de w , debido a que

$$\pi_{\Lambda^\circ}^\beta(x_{\Lambda^\circ} | x_{\Lambda \setminus \Lambda^\circ} w^\Lambda) = \pi_{\Lambda^\circ}^\beta(x_{\Lambda^\circ} | x^{\Lambda^\circ}) \quad \forall x, w \in E^S.$$

Por estos motivos eliminando a w de la notación, se trabajará simplemente con una configuración x en la que x_Λ es lo observado y x^Λ es elegido arbitrariamente. Luego, teniendo en cuenta el Lema 3.8.7(i), se tiene que

$$0 = \frac{\partial}{\partial \beta} U_{\Lambda}^{MV}(\beta, x) = H_{\Lambda^\circ}(x) - \frac{\partial}{\partial \beta} \ln \tilde{Z}_{\Lambda^\circ}^\beta(x) = H_{\Lambda^\circ}(x) - \tilde{\gamma}_{\Lambda^\circ}^\beta(H_{\Lambda^\circ} | x) \Leftrightarrow H_{\Lambda^\circ}(x) = \tilde{\gamma}_{\Lambda^\circ}^\beta(H_{\Lambda^\circ} | x).$$

Luego el estimador de Máxima Verosimilitud (MV) de β viene dado por

$$H_{\Lambda^\circ}(x) = \hat{\gamma}_{\Lambda^\circ}^{\beta^{MV}}(H_{\Lambda^\circ} | x). \quad (4.1.2)$$

Resolver esta ecuación es computacionalmente inviable, debido a la gran cantidad de operaciones necesarias para el cálculo de $\tilde{Z}_{\Lambda^\circ}^\beta$. Para tal fin, se necesita calcular la cantidad de pares de vecinos con la misma etiqueta en todos los mapas de E^Λ . Razón por la cual se debe buscar una alternativa más eficiente.

En el trabajo de Mcgrory et al. [37] se han propuesto expresiones recursivas exactas para computar $\tilde{Z}_{\Lambda^\circ}^\beta$. Sin embargo, estos métodos recursivos son posibles de llevar a cabo sólo en problemas pequeños. Younes [38] introdujo un algoritmo de gradiente estocástico basado en el algoritmo Simulated Annealing que iterativamente converge al estimador de MV de β . Los métodos Monte Carlo Markov Chain (MCMC) no pueden aplicarse en tiempo finito para estimar β , pero en los trabajos de Liu [39] y Risser [12], entre otros, se han diseñado algoritmos específicos para optimizarlos. Para aproximar el estimador de MV, se puede aplicar una particularización del método de estimación popular denominado EM (Expectation Maximization). Este procedimiento itera entre el cómputo de la esperanza de la log-verosimilitud de los datos y las etiquetas, dado β (paso E) y la actualización de β con el argumento que maximiza la esperanza (paso M). Para hacer calculable EM por medio de aproximaciones viables, existen diversas optimizaciones (ver el trabajo de Ibañes [40] y referencias allí).

Dejando de lado el estimador de MV, recientemente Okabayashi et al. [41] definieron el estimador de máxima verosimilitud de condicionales compuestas, el cual reemplaza la función objetivo propuesta por el estimador MV, por un producto de funciones de verosimilitud condicional sobre regiones elegidas convenientemente. Posteriormente, Friel [42] aplicó esta estrategia de estimación a procedimientos de segmentación de imágenes.

Por otro lado, en [43] Derin y Elliot propusieron el siguiente estimador de Mínimos Cuadrados (LSQR): si se cuenta con una observación $x_\Lambda \in E^\Lambda$ de $\pi_{\Lambda}^\beta(\cdot | w^\Lambda)$, entonces para cada $\ell \in E$ y $v \in E^{\#\partial s}$ se pueden definir las cantidades

$$N_{\Lambda}(\ell) := \#\{s \in \Lambda^{\circ} : x_s = \ell\} \quad \text{y} \quad N_{\Lambda}(\ell, \mathbf{v}) := \#\{s \in \Lambda^{\circ} : x_s = \ell, x_{\partial s} = \mathbf{v}\}.$$

Luego por el Lema 3.8.6

$$\frac{N_{\Lambda}(\ell, \mathbf{v})}{N_{\Lambda}(\mathbf{v})} \approx \pi_s^{\beta_0}(\ell|\mathbf{v}) = \frac{\exp\{\beta H_s(\ell)\}}{\sum_{\tilde{\ell} \in E} \exp\{\beta H_s(\tilde{\ell})\}},$$

implicando que

$$\frac{\exp\{\beta H_s(\ell)\}}{N_{\Lambda}(\ell, \mathbf{v})} \approx \frac{\sum_{\tilde{\ell} \in E} \exp\{\beta H_s(\tilde{\ell})\}}{N_{\Lambda}(\mathbf{v})}.$$

Como el lado derecho no depende de ℓ , entonces el lado izquierdo es aproximadamente constante en ℓ . Luego

$$\frac{\exp\{\beta H_s(\ell_1)\}}{\exp\{\beta H_s(\ell_2)\}} \approx \frac{N_{\Lambda}(\ell_1, \mathbf{v})}{N_{\Lambda}(\ell_2, \mathbf{v})} \quad \forall \ell_1, \ell_2 \in E \quad \forall \mathbf{v} \in E^{\#\partial s},$$

o equivalentemente,

$$\beta(H_s(\ell_1) - H_s(\ell_2)) \approx \ln \frac{N_{\Lambda}(\ell_1, \mathbf{v})}{N_{\Lambda}(\ell_2, \mathbf{v})} \quad \forall \ell_1, \ell_2 \in E \quad \forall \mathbf{v} \in E^{\#\partial s},$$

que es un sistema sobredeterminados de ecuaciones lineales, que puede ser resuelto por Mínimos Cuadrados. En caso de que algún $N_{\Lambda}(\ell, \mathbf{v}) = 0$, se excluyen las ecuaciones en las que interviene.

Por otro lado, en Ali et al. [44] se aproximó la función $U_{\Lambda^{\circ}}^{\text{prior}}(\cdot, x)$ con su polinomio de Taylor $p(\cdot, x)$ de orden 2 en un entorno de $\beta = 0$, para luego estimar β por medio del máximo global de dicho polinomio, esto es,

$$\hat{\beta} = L \frac{H_{\Lambda^{\circ}}(x)}{\#\Lambda^{\circ}}.$$

Estos métodos estiman basándose exclusivamente en una realización del modelo de Potts. Sin embargo, en el ámbito de la segmentación de imágenes uno no cuenta con tal realización, sino que se cuenta con los datos radiométricos I de la imagen sobre un conjunto finito de sitios Λ . Para poder estimar con estos procedimientos se debe segmentar la imagen, para con ella estimar. Ahora la cuestión es, ¿con qué procedimiento segmentamos?. Si usamos un segmentador no contextual como MV, obtendremos un mapa de clases muy granuloso que seguramente producirá subestimaciones de β (ver Capítulo 5), en cambio si usamos un procedimiento contextual que use la información que introduce el modelo de Potts, necesitaremos de una estimación del modelo, lo cual no tenemos aún. Sin embargo puede realizarse un procedimiento que itere entre segmentar y estimar. Pereyra et al. [13] afrontaron el problema de estimar β conjuntamente con los parámetros de un modelo Bayesiano, evitando el cálculo de $\tilde{Z}_{\Lambda}^{\beta}$. En esta propuesta la estimación de β se realiza utilizando un algoritmo likelihood-free Metropolis-Hastings, en el cual $\tilde{Z}_{\Lambda}^{\beta}$ se reemplazó por un procedimiento de aceptación y rechazo.

Todos estos métodos son imprecisos o costoso computacionalmente, a excepción de los estimadores de PMV [1], los cuales serán tratados en detalle en el contexto Potts en la próxima Sección.

4.2. Pseudo-Máxima Verosimilitud en el Modelo de Potts

En la presente Sección se particularizará el estimador de PMV [1] al modelo de Potts isotrópico. Esta contextualización se puede realizar debido a que los modelos de Potts pertenecen a la Familia Exponencial (H.10), tal y como se observó en la Sección 3.5, con

$$H_{\Lambda}(x) = \#\{\Delta \in \mathfrak{G}(\Lambda) : \#\Delta = 2, \Delta = \{s, t\}, x_s = x_t, t \in \partial s\}$$

y $h_\Lambda \equiv 1$ para el modelo a priori, y

$$h_\Lambda(x) = \prod_{s \in \Lambda} f_{x_s}(I_s),$$

para el modelo a posteriori. En particular $h_s(x) = f_{x_s}(I_s)$ y $H_s(y) = \#\{t \in \partial s : y_t = y_t\}$. El modelo a priori es el modelo invariante asociado al modelo a posteriori.

Teniendo en cuenta el estimador de PMV definido en la Sección 3.7, dada una observación $x_\Lambda \in E^\Lambda$, el estimador de PMV de β se puede definir siguiendo el modelo a priori o el modelo a posteriori. De este modo surgen dos estimadores de β , que se denotarán $\hat{\beta}_{\text{prior}}$ y $\hat{\beta}_{\text{post}}$, y se definirán a continuación.

4.2.1. A priori

En el caso de particular en que se esté trabajando con la distribución de Gibbs a priori correspondiente al mapa de clases (Modelo de Potts isotrópico sin campo externo), se tiene que

$$U_\Lambda^{\text{prior}}(\beta, x) = \frac{1}{\#\Lambda^\circ} \sum_{s \in \Lambda^\circ} (\beta H_s(x_s) - \ln Z_s^\beta(x)),$$

donde

$$\tilde{Z}_s^\beta(x) = \sum_{\ell \in E} \exp\{\beta H_s(\ell)\}.$$

Por el Lema 3.8.7(i), resulta

$$\frac{\partial}{\partial \beta} U_\Lambda^{\text{prior}}(\beta, x_\Lambda w^\Lambda) = \frac{1}{\#\Lambda^\circ} \sum_{s \in \Lambda^\circ} \left(H_s(x_s) - \frac{\partial}{\partial \beta} \ln \tilde{Z}_s^\beta(x) \right) = \frac{1}{\#\Lambda^\circ} \sum_{s \in \Lambda^\circ} \left(H_s(x_s) - \tilde{\gamma}_s^\beta(H_s|x) \right).$$

Luego el estimador de PMV correspondiente al modelo a priori viene dado por la solución de la siguiente ecuación

$$\sum_{s \in \Lambda^\circ} H_s(x_s) - \sum_{s \in \Lambda^\circ} \tilde{\gamma}_s^{\hat{\beta}_{\text{prior}}}(H_s|x) = 0. \quad (4.2.3)$$

Por la Proposición 3.2.7(ii)

$$\tilde{\gamma}_s^\beta(H_s|x) = \sum_{\ell \in E} H_s(\ell) \tilde{\rho}_s^\beta(\ell x^s) = \frac{\sum_{\ell \in E} H_s(\ell) \exp\{\beta H_s(\ell)\}}{\sum_{\ell \in E} \exp\{\beta H_s(\ell)\}},$$

y por ende, el estimador a priori $\hat{\beta}_{\text{prior}}$ debe ser solución de la ecuación

$$f_{\text{prior}}(\hat{\beta}_{\text{prior}}(x), x) = 0, \quad (4.2.4)$$

donde

$$f_{\text{prior}}(\beta, x) := \sum_{s \in \Lambda^\circ} H_s(x_s) - \sum_{s \in \Lambda^\circ} \frac{\sum_{\ell \in E} H_s(\ell) \exp\{\beta H_s(\ell)\}}{\sum_{\ell \in E} \exp\{\beta H_s(\ell)\}}. \quad (4.2.5)$$

Extrayendo factor común adecuadamente, Frery [15] propuso una reducción de los $\#\Lambda^\circ + 1$ términos de esta ecuación. Posteriormente, Levada [16] optimizó esta reducción. En particular, si se trabaja con vecindades de segundo orden, $\hat{\beta}_{\text{prior}}$ es la solución de la siguiente ecuación no lineal de 23 términos con única incógnita β

$$\sum_{s \in \Lambda^\circ} H_s(x_s) - \frac{8e^{8\beta}}{e^{8\beta} + L - 1} K_1 - \frac{7e^{7\beta} + e^\beta}{e^{7\beta} + e^\beta + L - 2} K_2 - \frac{6e^{6\beta} + 2e^{2\beta}}{e^{6\beta} + e^{2\beta} + L - 2} K_3 - \frac{6e^{6\beta} + 2e^\beta}{e^{6\beta} + 2e^\beta + L - 3} K_4$$

$$\begin{aligned}
& -\frac{5e^{5\beta} + 3e^{3\beta}}{e^{5\beta} + e^{3\beta} + L - 2}K_5 - \frac{5e^{5\beta} + 2e^{2\beta} + e^\beta}{e^{5\beta} + e^{2\beta} + e^\beta + L - 3}K_6 - \frac{5e^{5\beta} + 3e^\beta}{e^{5\beta} + 3e^\beta + L - 4}K_7 - \frac{8e^{4\beta}}{2e^{4\beta} + L - 2}K_8 \\
& -\frac{4e^{4\beta} + 3e^{3\beta} + e^\beta}{e^{4\beta} + e^{3\beta} + e^\beta + L - 3}K_9 - \frac{4e^{4\beta} + 4e^{2\beta}}{e^{4\beta} + 2e^{2\beta} + L - 3}K_{10} - \frac{4e^{4\beta} + 2e^{2\beta} + 2e^\beta}{e^{4\beta} + e^{2\beta} + 2e^\beta + L - 4}K_{11} - \frac{4e^{4\beta} + 4e^\beta}{e^{4\beta} + 4e^\beta + L - 5}K_{12} \\
& -\frac{6e^{3\beta} + 2e^{2\beta}}{2e^{3\beta} + e^{2\beta} + L - 3}K_{13} - \frac{6e^{3\beta} + 2e^\beta}{2e^{3\beta} + 2e^\beta + L - 4}K_{14} - \frac{3e^{3\beta} + 4e^{2\beta} + e^\beta}{e^{3\beta} + 2e^{2\beta} + e^\beta + L - 4}K_{15} - \frac{3e^{3\beta} + 2e^{2\beta} + 3e^\beta}{e^{3\beta} + e^{2\beta} + 3e^\beta + L - 5}K_{16} \\
& -\frac{3e^{3\beta} + 5e^\beta}{e^{3\beta} + 5e^\beta + L - 6}K_{17} - \frac{8e^{2\beta}}{4e^{2\beta} + L - 4}K_{18} - \frac{6e^{2\beta} + 2e^\beta}{3e^{2\beta} + 2e^\beta + L - 5}K_{19} - \frac{4e^{2\beta} + 4e^\beta}{2e^{2\beta} + 4e^\beta + L - 6}K_{20} \\
& -\frac{2e^{2\beta} + 6e^\beta}{e^{2\beta} + 6e^\beta + L - 7}K_{21} - \frac{8e^\beta}{8e^\beta + L - 8}K_{22} = 0, \tag{4.2.6}
\end{aligned}$$

donde los K_i son constantes que cuentan la cantidad de veces que se repite un determinado patrón en el mapa de clases. Por ejemplo, K_9 cuenta la cantidad de píxeles que tienen entre sus 8 vecinos de segundo orden, 4 con una determinada clase $\ell_1 \in E$, 3 con otra clase distinta $\ell_2 \in E$, y un único pixel con otra clase $\ell_3 \in E$, diferente de las anteriores. Nótese que no se tiene en cuenta la etiqueta del pixel visitado, y que en caso de que $L < 8$, la cantidad de términos en (4.2.6) es aún menor.

4.2.2. A posteriori

La estimación $\hat{\beta}_{\text{prior}}$ propuesta por Frery [15] y Levada [16], trabaja sobre la distribución de Gibbs a priori del mapa de clases, utilizando los datos I sólo a través de \hat{x} , lo que produce una reducción de información. Esta reducción puede ser suficiente o introducir una pérdida en la eficiencia, por lo cual, en la tesis se propone un estimador que incorpora esta información y se estudian los beneficios de incorporar directamente la información radiométrica a las estimaciones de β . Este nuevo estimador de β , que llamaremos estimador a posteriori y notaremos $\hat{\beta}_{\text{post}}$ se presentó en el trabajo Gimenez et al. [33]. Note que sea cual sea el modelo a priori que se considere, se puede definir análogamente un estimador a posteriori del parámetro del modelo.

Dada una observación I_Λ de los datos radiométricos de la imagen, las probabilidades sobre E^Λ se modifican debido a que algunos mapas son más o menos probables que antes según I_Λ . Esta nueva distribución sobre E^Λ recibe el nombre de distribución a posteriori del mapa de clases sobre Λ . En la Sección 3.5 se vio que la probabilidad condicional de observar x_Λ sobre E^Λ , dado que se observaron los datos radiométricos I_Λ y que la configuración sobre $\partial\Lambda$ es dada por el mapa $w \in E^S$ prefijado, viene dada por

$$\Pi_\Lambda^\beta(x_\Lambda | I_\Lambda, w^\Lambda) \propto \exp \left\{ \sum_{s \in \Lambda} \ln f_{x_s}(I_s) + \beta H_\Lambda(x_\Lambda w^\Lambda) \right\} \quad \forall \Lambda \in \mathfrak{S},$$

donde f_{x_s} es la densidad con la que emite datos radiométricas la clase asignada al pixel s . En particular, para cada $s \in \Lambda^\circ$,

$$\Pi_s^\beta(x_s | I_s, x_{\Lambda \setminus \{s\}} w^\Lambda) \propto \exp \{ \ln f_{x_s}(I_s) + \beta H_s(x_s) \}, \tag{4.2.7}$$

y por lo tanto, la función objetivo a maximizar viene dada por

$$U_\Lambda^{\text{post}}(\beta, x) = \frac{1}{\#\Lambda^\circ} \sum_{s \in \Lambda^\circ} \left(\ln f_{x_s}(I_s) + \beta H_s(x_s) - \ln \tilde{Z}_s^\beta(x) \right),$$

donde

$$Z_s^\beta(x) = \sum_{\ell \in E} f_\ell(I_s) \exp\{\beta H_s(\ell)\}.$$

Si notamos con γ^β a la especificación correspondiente al modelo a posteriori, por el Lema 3.8.7(i), resulta

$$\frac{\partial}{\partial \beta} U_\Lambda^{\text{post}}(\beta, x_\Lambda w^\Lambda) = \frac{1}{\#\Lambda^\circ} \sum_{s \in \Lambda^\circ} \left(H_s(x_s) - \frac{\partial}{\partial \beta} \ln Z_s^\beta(x) \right) = \frac{1}{\#\Lambda^\circ} \sum_{s \in \Lambda^\circ} \left(H_s(x_s) - \gamma_s^\beta(H_s|x) \right).$$

Luego el estimador de PMV correspondiente al modelo a priori viene dado por la solución de la siguiente ecuación

$$\sum_{s \in \Lambda^\circ} \{H_s(x_s) - \hat{\gamma}_s^{\text{post}}(H_s|x)\} = 0,$$

donde, por la Proposición 3.2.7(ii),

$$\gamma_s^\beta(H_s|x) = \frac{\sum_{\ell \in E} H_s(\ell) f_\ell(I_s) \exp\{\beta H_s(\ell)\}}{\sum_{\ell \in E} f_\ell(I_s) \exp\{\beta H_s(\ell)\}}.$$

Luego, el estimador a posteriori $\hat{\beta}_{\text{post}}$ debe ser solución de la siguiente ecuación no lineal en β de a lo sumo $\#\Lambda^\circ + 1$ términos

$$f_{\text{post}}(\hat{\beta}_{\text{post}}(x), x) = 0, \quad (4.2.8)$$

donde f_{post} es una función de $\mathbb{R} \times E^S$ en \mathbb{R} dada por

$$f_{\text{post}}(\beta, x) := \sum_{s \in \Lambda^\circ} H_s(x_s) - \frac{\sum_{\ell \in E} H_s(\ell) f_\ell(I_s) \exp\{\beta H_s(\ell)\}}{\sum_{\ell \in E} f_\ell(I_s) \exp\{\beta H_s(\ell)\}}. \quad (4.2.9)$$

En la tesis, los sistemas (4.2.6) y (4.2.8) se resolvieron por medio de una implementación en Matlab del algoritmo de Brent [45]. La cantidad de operaciones computacionales que requiere el cómputo de los estimadores $\hat{\beta}_{\text{prior}}$ y $\hat{\beta}_{\text{post}}$ son del mismo orden, debido a que contar configuraciones iguales para luego sumar pocos términos, es similar a sumar muchos términos, después de todo, igual se debe visitar toda la imagen. Sin embargo, ambas estimaciones se realizan a un muy bajo costo.

En las próximas Secciones se analizarán las propiedades teóricas de los estimadores de PMV en los modelos de Potts, y se analizarán sus cualidades a muestra finita por medio de simulaciones.

4.3. Consistencia

En la presente Sección probaremos que los estimadores definidos en la Sección 4.2 son consistentes. Comencemos con el estimador a priori $\hat{\beta}_{\text{prior}}$. Para ello se debe verificar que el modelo a priori satisfaga las condiciones (H.1)-(H.15), no considerandose las hipótesis (H.16) y (H.17) por tratarse de un modelo invariante.

Trabajando con $S = \mathbb{Z}^2$ y la familia $\{\Lambda_n\}_{n \in \mathbb{N}}$ definida en (H.12) con $d = 2$; las vecindades mostradas en la Figura 3.1 verifican trivialmente las condiciones (H.2)-(H.5). El conjunto E es el conjunto finito de todas las posibles etiquetas, y por lo tanto se cumple (H.1). El conjunto paramétrico con el que trabajaremos es $\Upsilon = \mathbb{R}$, el cual verifica (H.6) y (H.7).

El potencial del modelo a priori definido en la Sección 3.5 es

$$\Phi_{\Lambda}^{\beta}(x) := \begin{cases} -\beta & \text{si } \#\Lambda = 2, \Lambda = \{s, t\}, x_s = x_t, t \in \partial S; \\ 0 & \text{caso contrario,} \end{cases}$$

el cual verifica (H.8)-(H.10). Luego $\mathcal{G}(\tilde{\gamma}^{\beta})$ es no vacío y convexo, y se puede considerar la medida de Gibbs extremal supuesta en (H.11). Como se vio en la Sección 3.5, el modelo a priori satisface las condiciones de invarianza (H.13) y (H.14). La condición de identificabilidad (H.15) se verifica por la linealidad de las funciones de energía del modelo a priori. De este modo entonces el estimador a priori es consistente.

Analicemos la consistencia del estimador a posteriori. En la Sección 3.7 se mostró que cada modelo tiene su estimador de PMV, como lo es el estimador $\hat{\beta}_{\text{prior}}$ para el modelo a priori trabajado, y como lo es el estimador $\hat{\beta}_{\text{post}}$ para el modelo a posteriori. Como los modelos a priori y a posteriori contienen el mismo parámetro de suavidad β , ambos estimadores son también estimadores del parámetro del otro modelo, donde dejan de ser el estimador de PMV, para ser otro estimador diferente. Esto es, por ejemplo, el estimador a posteriori es también estimador del parámetro del modelo a priori, pero en el modelo a priori el estimador de PMV es $\hat{\beta}_{\text{prior}}$, mientras que $\hat{\beta}_{\text{post}}$ es otro estimador diferente que incluye la verosimilitud de los datos radiométricos. Luego la consistencia de $\hat{\beta}_{\text{post}}$ puede ser analizada desde dos perspectivas diferentes, dependiendo si las probabilidades se calculan con la medida de Gibbs correspondiente al modelo a priori o la correspondiente al modelo a posteriori. Debido a que con los supuestos hechos hasta el momento los modelos a priori y a posteriori son modelos asociados pero no son equivalentes microscópicamente por no satisfacer (H.16), probaremos la consistencia de $\hat{\beta}_{\text{post}}$ como estimador del parámetro del modelo a priori.

Teorema 4.3.1 (Consistencia de $\hat{\beta}_{\text{post}}$) Sean $\mathcal{S} = \mathbb{Z}^2$, E un conjunto de estados que verifica (H.1); ∂ un sistema de vecindades que satisface (H.2)-(H.5); $\{\Lambda_n\}_{n \in \mathbb{N}}$ como se definió en (H.12); $\{\tilde{\gamma}^{\beta}\}_{\beta \in \Upsilon}$ la familia de especificaciones correspondientes al modelo a priori, donde $\Upsilon \subset \mathbb{R}$ es tal que

$$(H.11') \quad \mathcal{G}(\tilde{\gamma}^{\beta}) = \{\pi^{\beta}\} \quad \forall \beta \in \Upsilon.$$

Entonces

- (i) $\pi^{\beta_0}(\{x \in E^{\mathcal{S}} : U_n^{\text{post}}(\cdot, x) \text{ es estrictamente cóncava } \forall n \gg 1\}) = 1$;
- (ii) $\pi^{\beta_0}(\{x \in E^{\mathcal{S}} : \#M_{\Lambda_n}^{\text{post}}(x) = 1 \forall n \gg 1\}) = 1$;
- (iii) $\pi^{\beta_0}(\{x \in E^{\mathcal{S}} : \|\hat{\beta}_{n, \text{post}}(x) - \beta_0\| \rightarrow 0\}) = 1$.

Dem. Ver Sección 6.3 del Apéndice. □

4.4. Propiedades

El objetivo de esta parte del trabajo es mostrar algunas propiedades que cumplen las funciones f_{prior} y f_{post} , y que son importantes informativamente para futuras justificaciones.

Debido a que las funciones $x \mapsto f_{\text{prior}}(\beta, x)$ y $x \mapsto f_{\text{post}}(\beta, x)$ dependen solamente de las entradas sobre Λ de x , se tiene que son funciones \mathcal{F}_{Λ} -medibles. Luego, dado un mapa de clases x_{Λ} observado, cada vez que escribamos x nos referimos a alguna configuración sobre $E^{\mathcal{S}}$ que contiene al mapa observado sobre Λ .

Luego, dado un mapa observado x_{Λ} , las raíces de las funciones $f_{\text{prior}}(\cdot, x)$ y $f_{\text{post}}(\cdot, x)$ son las estimaciones por PMV de β , basadas en tal mapa.

4.4.1. Gráficos

En la Figura 4.1 se muestran las gráficas de las funciones $f_{\text{prior}}(\cdot, x)$ y $f_{\text{post}}(\cdot, x)$, siendo x_Λ un mapa simulado con $L = 3$ y $\beta = 0,4$. Las raíces de estas funciones están muy próximas a 0,4, lo cual representa que ambos estimadores aproximaron bien el verdadero parámetro. Ambas funciones muestran una suavidad importante, a pesar de sus complejas expresiones. Además, son continuas, decrecientes y generalmente poseen una sola raíz como veremos a continuación.

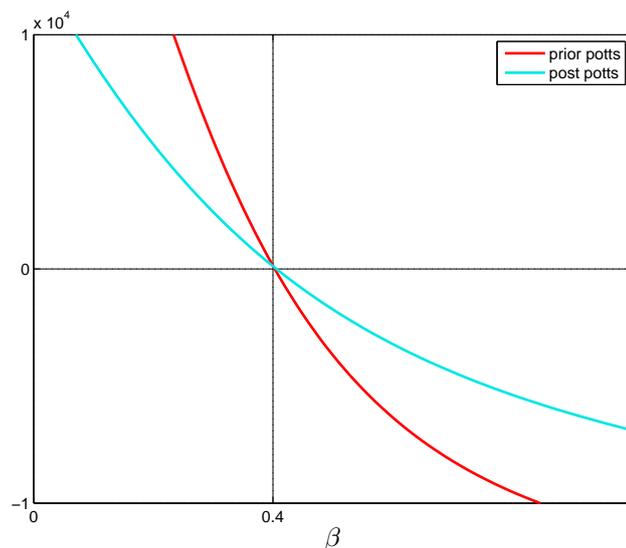


Figura 4.1: Gráfico de las funciones $f_{\text{prior}}(\cdot, x)$ y $f_{\text{post}}(\cdot, x)$.

Asíntotas

Por ser continuas sobre \mathbb{R} , las funciones no presentan asíntotas verticales. Probemos el siguiente resultado que nos permitirá analizar las asíntotas horizontales que poseen estas funciones.

Proposición 4.4.1 *Para cada $x \in E^S$, se verifican*

$$\lim_{\beta \rightarrow -\infty} f_{\text{prior}}(\beta, x) = \lim_{\beta \rightarrow -\infty} f_{\text{post}}(\beta, x) = \sum_{s \in \Lambda^\circ} (H_s(x_s) - \min_{\ell \in E} H_s(\ell)) \quad (4.4.10)$$

y

$$\lim_{\beta \rightarrow \infty} f_{\text{prior}}(\beta, x) = \lim_{\beta \rightarrow \infty} f_{\text{post}}(\beta, x) = \sum_{s \in \Lambda^\circ} (H_s(x_s) - \max_{\ell \in E} H_s(\ell)) \quad (4.4.11)$$

Dem.

$$\begin{aligned}
\lim_{\beta \rightarrow \infty} f_{\text{prior}}(\beta, x) &= \sum_{s \in \Lambda^\circ} H_s(x_s) - \sum_{s \in \Lambda^\circ} \lim_{\beta \rightarrow \infty} \frac{\sum_{\ell \in E} H_s(\ell) \exp\{\beta H_s(\ell)\}}{\sum_{\ell \in E} \exp\{\beta H_s(\ell)\}} \\
&= \sum_{s \in \Lambda^\circ} H_s(x_s) - \sum_{s \in \Lambda^\circ} \lim_{\beta \rightarrow \infty} \frac{\sum_{\ell \in E} H_s(\ell) \exp\left\{\beta \left(H_s(\ell) - \max_{\ell \in E} H_s(\ell)\right)\right\}}{\sum_{\ell \in E} \exp\left\{\beta \left(H_s(\ell) - \max_{\ell \in E} H_s(\ell)\right)\right\}} \\
&= \sum_{s \in \Lambda^\circ} H_s(x_s) - \sum_{s \in \Lambda^\circ} \frac{\sum_{\ell \in E} H_s(\ell) \lim_{\beta \rightarrow \infty} \exp\left\{\beta \left(H_s(\ell) - \max_{\ell \in E} H_s(\ell)\right)\right\}}{\sum_{\ell \in E} \lim_{\beta \rightarrow \infty} \exp\left\{\beta \left(H_s(\ell) - \max_{\ell \in E} H_s(\ell)\right)\right\}} \\
&= \sum_{s \in \Lambda^\circ} (H_s(x_s) - \max_{\ell \in E} H_s(\ell)),
\end{aligned}$$

donde la última igualdad vale por lo siguiente

$$\lim_{\beta \rightarrow \infty} \exp\left\{\beta \left(H_s(\ell) - \max_{\ell \in E} H_s(\ell)\right)\right\} = \begin{cases} 1 & \text{si } H_s(\ell) = \max_{\ell \in E} H_s(\ell) \\ 0 & \text{si } H_s(\ell) < \max_{\ell \in E} H_s(\ell). \end{cases}$$

De manera análoga se prueban las demás igualdades. □

Luego las funciones en cuestión son acotadas, ya que son continuas y presentan dos asíntotas horizontales que la acotan superior e inferiormente.

4.4.2. Derivadas

En el cálculo numérico se dice que un sistema de 2 ecuaciones lineales con dos incógnitas está mal condicionado si las pendientes de las rectas correspondientes a la cada ecuación son similares. Esto produce que cualquier error numérico ínfimo, traslade levemente cada recta, pudiendo desplazar en gran medida el punto de intersección entre ellas.

Este concepto se traslada a nuestro contexto, ya que si una curva tiene una derivada con poco módulo, entonces cualquier traslación ínfima de la curva produce que sus raíces cambien considerablemente. De este modo, toma particular importancia conocer el módulo de la derivada de las funciones (4.2.5) y (4.2.9). El siguiente resultado brinda esta información.

Proposición 4.4.2 *Para cada $x \in E^S$, se verifican*

$$\frac{d}{d\beta} f_{\text{prior}}(\beta, x) = - \sum_{s \in \Lambda^\circ} \tilde{\gamma}_s^\beta \left(\left(H_s - \tilde{\gamma}_s^\beta(H_s|x) \right)^2 |x \right) \leq 0, \quad (4.4.12)$$

y

$$\frac{d}{d\beta} f_{\text{post}}(\beta, x) = - \sum_{s \in \Lambda^\circ} \gamma_s^\beta \left(\left(H_s - \gamma_s^\beta(H_s|x) \right)^2 |x \right) \leq 0. \quad (4.4.13)$$

Dem. Derivando (4.2.5), se tiene que

$$\begin{aligned}
\frac{d}{d\beta} f_{\text{prior}}(\beta, x) &= - \sum_{s \in \Lambda^\circ} \sum_{\ell \in E} H_s(\ell) \frac{d}{d\beta} \frac{\exp\{\beta H_s(\ell)\}}{\sum_{\tilde{\ell} \in E} \exp\{\beta H_s(\tilde{\ell})\}} \\
&= - \sum_{s \in \Lambda^\circ} \sum_{\ell \in E} H_s(\ell) \frac{\exp\{\beta H_s(\ell)\} \sum_{\tilde{\ell} \in E} (H_s(\ell) - H_s(\tilde{\ell})) \exp\{\beta H_s(\tilde{\ell})\}}{\left(\sum_{\tilde{\ell} \in E} \exp\{\beta H_s(\tilde{\ell})\} \right)^2} \\
&= - \sum_{s \in \Lambda^\circ} \sum_{\ell \in E} H_s(\ell) \left(H_s(\ell) - \tilde{\gamma}_s^\beta(H_s|x) \right) \frac{\exp\{\beta H_s(\ell)\}}{\sum_{\tilde{\ell} \in E} \exp\{\beta H_s(\tilde{\ell})\}} \\
&= - \sum_{s \in \Lambda^\circ} \tilde{\gamma}_s^\beta \left(\left(H_s - \tilde{\gamma}_s^\beta(H_s|x) \right)^2 |x \right).
\end{aligned}$$

La no negatividad vale por la desigualdad

$$\left(H_s(x) - \tilde{\gamma}_s^\beta(H_s|x) \right)^2 \geq 0 \quad \forall x \in E^{\mathcal{S}},$$

y por lo tanto vale (4.4.12). De igual modo se prueba (4.4.13). □

Esta Proposición implica que para conocer la derivada (pendiente de la recta tangente) que tiene la curva correspondiente a la función (4.2.5) en un punto determinado β , se debe sumar sobre todos los píxeles la varianza entre las cantidades de píxeles vecinos con las distintas clases. Por ejemplo, en un mapa con 4 clases generado con $\beta = 1$, un determinado pixel s , puede tener 3 vecinos con una clase ℓ_1 , 2 vecinos con clase ℓ_2 , 2 vecinos con clase ℓ_3 y un vecino con la clase ℓ_4 , tal y como se diagrama a continuación:

$$x_{\partial s} = \begin{array}{|c|c|c|} \hline \ell_1 & \ell_1 & \ell_1 \\ \hline \ell_4 & \times & \ell_2 \\ \hline \ell_3 & \ell_3 & \ell_2 \\ \hline \end{array}$$

Luego, en este caso:

$$\pi_s(\ell_1|x_{\partial s}) = \frac{\exp\{\beta H_s(\ell_1)\}}{\sum_{i=1}^4 \exp\{\beta H_s(\ell_i)\}} = \frac{e^3}{e^3 + e^2 + e^2 + e^1} = 0,5344$$

$$\pi_s(\ell_2|x_{\partial s}) = \pi_s(\ell_3|x_{\partial s}) = 0,1966; \quad \pi_s(\ell_4|x_{\partial s}) = 0,0723.$$

Luego

$$\tilde{\gamma}_s^\beta(H_s|x) = 3 \times 0,5344 + 2 \times 0,1966 + 2 \times 0,1966 + 1 \times 0,0723 = 2,4621$$

y

$$\begin{aligned} \tilde{\gamma}_s^\beta \left(\left(H_s - \tilde{\gamma}_s^\beta (H_s|x) \right)^2 |x \right) &= (3 - 2,4621)^2 \times 0,5344 + (2 - 2,4621)^2 \times 0,1966 + (2 - 2,4621)^2 \times \\ &\quad \times 0,1966 + (1 - 2,4621)^2 \times 0,0723 = 0,3932 \end{aligned}$$

Repitiendo este procedimiento sobre todos los píxeles y sumando las varianzas se obtiene el módulo de la derivada de la curva correspondiente a la función (4.2.5) en $\beta = 1$.

Veamos ahora que sucede con la curva correspondiente al estimador a posteriori. Para comenzar suponemos que las clases emiten observaciones con los mismos parámetros. En tal caso las cantidades $f_\ell(I_s)$ serán constantes en ℓ , e iguales a p_s . Para ilustrar esta situación continuemos con el ejemplo anterior en el que fijamos la cantidad de clases en 4, $\beta = 1$, $H_s(\ell_1) = 3$, $H_s(\ell_2) = H_s(\ell_3) = 2$ y $H_s(\ell_4) = 1$. Luego

$$\Pi_s(\ell|x_{\partial s}, I_s) = \frac{\exp\{\ln p_s + \beta H_s(\ell)\}}{\sum_{i=1}^4 \exp\{\ln p_s + \beta H_s(\ell_i)\}} = \frac{\exp\{\beta H_s(\ell)\}}{\sum_{i=1}^4 \exp\{\beta H_s(\ell_i)\}} = \pi_s(\ell|x_{\partial s})$$

y por lo tanto, en tal caso, las derivadas son iguales.

Veamos ahora que sucede si la clase ℓ_1 emite observaciones con distintos parámetros, y continuemos suponiendo que las demás clases emiten observaciones con la misma distribución. Para abreviar notación, llamemos $\Delta_s = f_{\ell_1}(I_s) - p_s$. Luego

$$\begin{aligned} \Pi_s(\ell_1|x_{\partial s}, I_s) &= \frac{\exp\{\ln(p_s + \Delta_s) + \beta H_s(\ell_1)\}}{\sum_{i=1}^4 \exp\{\ln p(I_s|\ell_i) + \beta H_s(\ell_i)\}} = \frac{(p_s + \Delta_s) \exp\{\beta H_s(\ell_1)\}}{\Delta_s \exp\{\beta H_s(\ell_1)\} + \sum_{i=2}^4 \exp\{\ln p_s + \beta H_s(\ell_i)\}} = \\ &= \frac{(1 + \delta_s) \exp\{\beta H_s(\ell_1)\}}{\delta_s \exp\{\beta H_s(\ell_1)\} + \sum_{i=2}^4 \exp\{\beta H_s(\ell_i)\}}, \end{aligned}$$

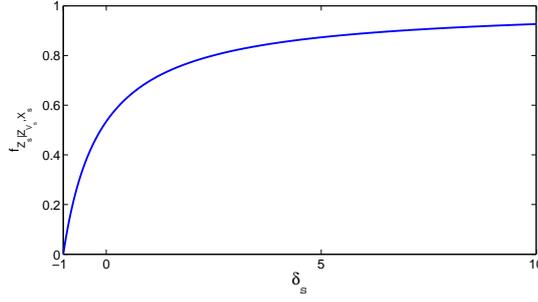
donde $\delta_s := \frac{\Delta_s}{p_s} \in (-1, \infty)$, pues $-p_s \leq \Delta_s \leq 1$.

Se verifica que $\delta_s = -1$ cuando $f_{\ell_1}(I_s) = 0$, esto es, cuando es improbable que la observación I_s provenga de la clase ℓ_1 ; y $\delta_s \rightarrow \infty$ a medida que crece $f_{\ell_1}(I_s)$ tendiendo a 1, esto es, cuando la verosimilitud de la clase ℓ_1 predomina sobre la de las demás clases. En la Figura 4.2 se aprecia la gráfica de $\Pi_s(\ell_1|x_{\partial s}, I_s)$ en función de δ_s . Esta función es creciente en δ_s y presenta una asíntota horizontal en 1. Esto implica que si $p_s \approx 0$, entonces δ_s es muy grande, y por ende, ℓ_1 logra un gran dominio sobre las restantes etiquetas.

En el caso extremo que $\Pi_s(\ell_1|x_{\partial s}, I_s) = 1 \forall s \in \Lambda^\circ$, se tiene que la varianza

$$\tilde{\gamma}_s^\beta \left(\left(H_s - \tilde{\gamma}_s^\beta (H_s|x) \right)^2 |x \right) = 0 \quad \forall s \in \Lambda^\circ,$$

y por lo tanto la derivada de la función (4.2.9) será menor en módulo que la derivada de la función (4.2.5). Luego intuitivamente uno puede pensar que a medida que las clases emitan observaciones más distantes, decrecerá el módulo de la derivada de la función (4.2.9). Esto implica que la curva corte al eje horizontal con un menor ángulo, y por lo tanto será más sensible a cualquier error de cálculo o del modelo. Demostremos esto teóricamente bajo el supuesto de unicidad de la medida de Gibbs correspondiente al modelo a priori, esto es:

Figura 4.2: Gráfico de $\Pi_s(\ell_1 | x_{\partial s}, I_s)$ en función de δ_s .

Proposición 4.4.3 Si $\mathcal{G}(\tilde{\gamma}^\beta) = \{\pi^\beta\}$, entonces

$$\pi^\beta \left(\left\{ x \in E^{\mathcal{S}} : \liminf_{n \rightarrow \infty} \left(\frac{d}{d\beta} f_{\text{prior}}(\beta, x_{\Lambda_n} w^{\Lambda_n}) - \frac{d}{d\beta} f_{\text{post}}(\beta, x_{\Lambda_n} w^{\Lambda_n}) \right) \geq 0 \forall \beta \in \mathbb{R} \right\} \right) = 1, \quad (4.4.14)$$

para cada $w \in E^{\mathcal{S}}$.

Dem. Por el Lema 6.4.1, para cada $s \in \mathcal{S}$, las funciones $h_s(X) = f_{X_s}(I_s)$ y $H_s(X) = \#\{t \in \partial s : X_s = X_t\}$ son independientes como variables aleatorias definidas sobre el espacio de Probabilidad (Ω, \mathcal{A}, P) definido en la Sección 3.5. Sean $w \in E^{\mathcal{S}}$, $\beta \in \mathbb{R}$ y $n \in \mathbb{N}$, por la \mathcal{F}_{Λ_n} -medibilidad de las funciones y por propiedades de la varianza,

$$\frac{d}{d\beta} f_{\text{prior}}(\beta, x_{\Lambda_n} w^{\Lambda_n}) = \frac{d}{d\beta} f_{\text{prior}}(\beta, x) = - \sum_{s \in \Lambda_n^{\circ}} \left[\tilde{\gamma}_s^\beta (H_s^2 | x) - \tilde{\gamma}_s^\beta (H_s | x)^2 \right],$$

y

$$\frac{d}{d\beta} f_{\text{post}}(\beta, x_{\Lambda_n} w^{\Lambda_n}) = \frac{d}{d\beta} f_{\text{post}}(\beta, x) = - \sum_{s \in \Lambda_n^{\circ}} \left[\gamma_s^\beta (H_s^2 | x) - \gamma_s^\beta (H_s | x)^2 \right] = - \sum_{s \in \Lambda_n^{\circ}} \left[\tilde{\gamma}_s^\beta (\tilde{h}_s H_s^2 | x) - \tilde{\gamma}_s^\beta (\tilde{h}_s H_s | x)^2 \right],$$

donde la última igualdad vale por el Lema 3.8.1 (iii), siendo $\tilde{h}_s := h_s / \tilde{\gamma}_s^\beta (h_s | \cdot)$. Luego $\tilde{\gamma}_s^\beta (\tilde{h}_s | \cdot) \equiv 1$ y

$$\begin{aligned} \frac{d}{d\beta} f_{\text{prior}}(\beta, x_{\Lambda_n} w^{\Lambda_n}) - \frac{d}{d\beta} f_{\text{post}}(\beta, x_{\Lambda_n} w^{\Lambda_n}) &= \sum_{s \in \Lambda_n^{\circ}} \left[\tilde{\gamma}_s^\beta (\tilde{h}_s H_s^2 | x) - \tilde{\gamma}_s^\beta (\tilde{h}_s H_s | x)^2 - \tilde{\gamma}_s^\beta (H_s^2 | x) + \tilde{\gamma}_s^\beta (H_s | x)^2 \right] \\ &= \sum_{s \in \Lambda_n^{\circ}} \left[\tilde{\gamma}_s^\beta ((\tilde{h}_s - 1) H_s^2 | x) + \tilde{\gamma}_s^\beta (H_s | x)^2 - \tilde{\gamma}_s^\beta (\tilde{h}_s H_s | x)^2 \right]. \end{aligned}$$

Por la desigualdad de Cauchy-Schwarz

$$\tilde{\gamma}_s^\beta (\tilde{h}_s H_s | x)^2 \leq \tilde{\gamma}_s^\beta (\tilde{h}_s | x)^2 \tilde{\gamma}_s^\beta (H_s | x)^2 = \tilde{\gamma}_s^\beta (H_s | x)^2 \quad \forall s \in \Lambda_n^{\circ}.$$

Luego

$$\frac{d}{d\beta} f_{\text{prior}}(\beta, x_{\Lambda_n} w^{\Lambda_n}) - \frac{d}{d\beta} f_{\text{post}}(\beta, x_{\Lambda_n} w^{\Lambda_n}) \geq \sum_{s \in \Lambda_n^{\circ}} \tilde{\gamma}_s^\beta ((\tilde{h}_s - 1) H_s^2 | x).$$

Como $\pi^\beta \in \mathcal{G}(\tilde{\gamma}^\beta)$, y posteriormente por la independencia entre h_s y H_s ,

$$\pi^\beta(\tilde{\gamma}_s^\beta((\tilde{h}_s - 1)H_s^2|\cdot)) = \pi^\beta((\tilde{h}_s - 1)H_s^2) = \pi^\beta(\tilde{h}_s - 1)\pi^\beta(H_s^2) = \pi^\beta(\tilde{\gamma}_s^\beta(\tilde{h}_s|\cdot) - 1)\pi^\beta(H_s^2) = 0.$$

Luego por el Corolario 3.6.1, resulta que

$$\frac{1}{\#\Lambda_n^\circ} \sum_{s \in \Lambda_n^\circ} \tilde{\gamma}_s^\beta((\tilde{h}_s - 1)H_s^2|x) \rightarrow 0 \quad \pi^\beta\text{-p.p.}$$

Por lo que

$$\liminf_{n \rightarrow \infty} \frac{1}{\#\Lambda_n^\circ} \left(\frac{d}{d\beta} f_{\text{prior}}(\beta, x_{\Lambda_n} w^{\Lambda_n}) - \frac{d}{d\beta} f_{\text{post}}(\beta, x_{\Lambda_n} w^{\Lambda_n}) \right) \geq \lim_{n \rightarrow \infty} \frac{1}{\#\Lambda_n^\circ} \sum_{s \in \Lambda_n^\circ} \tilde{\gamma}_s^\beta((\tilde{h}_s - 1)H_s^2|x) = 0 \quad \pi^\beta\text{-p.p.},$$

lo cual prueba el resultado. □

4.4.3. Raíces

Analizar las raíces de las funciones $f_{\text{prior}}(\cdot, x)$ y $f_{\text{post}}(\cdot, x)$ es equivalente a analizar la existencia y unicidad de los estimadores $\hat{\beta}_{\text{prior}}$ y $\hat{\beta}_{\text{post}}$ cuando la observación del modelo es x_Λ , y por lo tanto es importante investigar este tema.

Proposición 4.4.4 Si $x_\Lambda \in E^\Lambda$ es un mapa para el cual existen dos píxeles $s, t \in \Lambda^\circ$ no necesariamente distintos tales que

$$H_s(x_s) > \min_{\ell \in E} H_s(\ell) \quad \text{y} \quad U_t(x_t) < \max_{\ell \in E} U_t(\ell) \quad (4.4.15)$$

entonces las ecuaciones (4.2.4) y (4.2.8) tienen una única solución. En caso de verificarse sólo una de las condiciones de (4.4.15) ambas funciones no tienen raíces, y si no se verifica ninguna de estas condiciones entonces $f_{\text{prior}}(\cdot, x) = f_{\text{post}}(\cdot, x) \equiv 0$.

Dem. Por (4.4.15) y la Proposición 4.4.1, se tiene que

$$\lim_{\beta \rightarrow -\infty} f_{\text{prior}}(\beta, x) = \lim_{\beta \rightarrow -\infty} f_{\text{post}}(\beta, x) > 0 \quad \text{y} \quad \lim_{\beta \rightarrow \infty} f_{\text{prior}}(\beta, x) = \lim_{\beta \rightarrow \infty} f_{\text{post}}(\beta, x) < 0.$$

Luego, como ambas funciones son continuas, por el Teorema del valor intermedio, ambas funciones tienen raíces. Además, por la Proposición 4.4.2, las funciones (4.2.5) y (4.2.9) son estrictamente decrecientes, implicando la unicidad de raíces para ambas funciones. Por otro lado, si se verifica sólo la primera de las condiciones dadas en (4.4.15) (en caso de valer sólo la segunda, el análisis es simétrico), entonces

$$\lim_{\beta \rightarrow -\infty} f_{\text{prior}}(\beta, x) = \lim_{\beta \rightarrow -\infty} f_{\text{post}}(\beta, x) = 0 \quad \text{y} \quad \lim_{\beta \rightarrow \infty} f_{\text{prior}}(\beta, x) = \lim_{\beta \rightarrow \infty} f_{\text{post}}(\beta, x) < 0,$$

y como por la Proposición 4.4.2 las funciones son estrictamente decrecientes, queda que $y = 0$ es una asíntota horizontal y una cota inferior de las mismas, lo que implica que en este caso las funciones no tienen raíces. Por último, si no se cumple ninguna de las condiciones dadas en (4.4.15), entonces

$$\min_{\ell \in E} H_s(\ell) = \max_{\ell \in E} H_s(\ell) \quad \forall s \in \Lambda^\circ,$$

implicando que $H_s(\ell)$ sea constante en ℓ para cada s . Luego

$$f_{\text{prior}}(\beta, x) = \sum_{s \in \Lambda^\circ} H_s(x_s) - \sum_{s \in \Lambda^\circ} \frac{\#E H_s(x_s) \exp\{\beta H_s(x_s)\}}{\#E \exp\{\beta H_s(x_s)\}} = \sum_{s \in \Lambda^\circ} H_s(x_s) - \sum_{s \in \Lambda^\circ} H_s(x_s) = 0 \quad \forall \beta \in \mathbb{R}.$$

Estos resultados se prueban análogamente para f_{post} . □

Las condiciones dadas en (4.4.15) se verifican en general, salvo casos particulares rara vez observados en la práctica. Para ilustrar algunos ejemplos, supongamos que trabajamos con vecindades de segundo orden y que Λ es de 7×7 , al cual si le agregamos el conjunto $\partial\Lambda$ queda un mapa de 9×9 , como los que aparecen en la Figura 4.3.

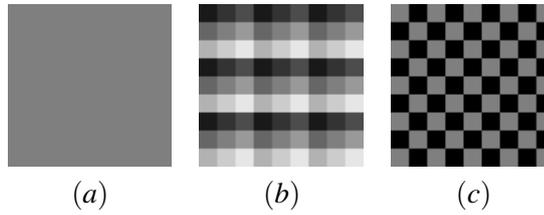


Figura 4.3: Mapas de clases atípicos para los cuales no se verifica (4.4.15) si se consideran vecindades de segundo orden.

Para el mapa (a): si $L \geq 2$, como

$$H_s(\ell) = 8; \quad \min_{\ell \in E} H_s(\ell) = 0 \quad \text{y} \quad \max_{\ell \in E} H_s(\ell) = 8 \quad \forall s \in \Lambda^\circ \forall \ell \in E$$

se tiene que ambos funciones no tienen raíces.

Para el mapa (b): como

$$H_s(\ell) = 0; \quad \min_{\ell \in E} H_s(\ell) = 0 \quad \text{y} \quad \max_{\ell \in E} H_s(\ell) = 1 \quad \forall s \in \Lambda^\circ \forall \ell \in E$$

se tiene que ambos funciones no tienen raíces.

Para el mapa (c): como

$$H_s(\ell) = 4; \quad \min_{\ell \in E} H_s(\ell) = \begin{cases} 4 & \text{si } L = 2 \\ 0 & \text{si } L > 2 \end{cases} \quad \text{y} \quad \max_{\ell \in E} H_s(\ell) = 4 \quad \forall s \in \Lambda^\circ \forall \ell \in E$$

se tiene que ambos funciones no tienen raíces si $L > 2$, y tiene infinitas raíces si $L = 2$.

4.5. ECM, Sesgo y Varianza bajo simulación

De este modo, se cuenta con dos estimadores de PMV consistentes de β , el que utiliza solamente el mapa observado, y el que además se vale directamente de la información radiométrica de la imagen observada. En el presente trabajo se comparan las estimaciones que proveen ambos estimadores por medio de simulaciones.

Como base para el trabajo, se simularon por medio del algoritmo Swendsen-Wang 100 modelos de Potts para cada combinación de las siguientes situaciones: β tomó los valores $\{0, 1; 0, 2; 0, 3; 0, 4; 0, 45; 0, 5; 0, 6; 0, 7; 0, 8; 0, 9; 1\}$, y la cantidad de clases L se varió entre 2, 3 y 4. Para cada una de estas situaciones, se simularon

datos radiométricos Gaussianos con el mismo desvío $\sigma = 15$ para cada clase, y medias distanciadas a $k = 1, 2, 3$ y 4 desvíos. La Figura 4.4 muestra en (a) la realización del modelo de Potts para $L = 2$ y $\beta = 0,3$; y en (b) los datos radiométricos Gaussianos con desvíos 15 y medias 70 y 100 respectivamente (distanciadas a $k = 2$ desvíos).

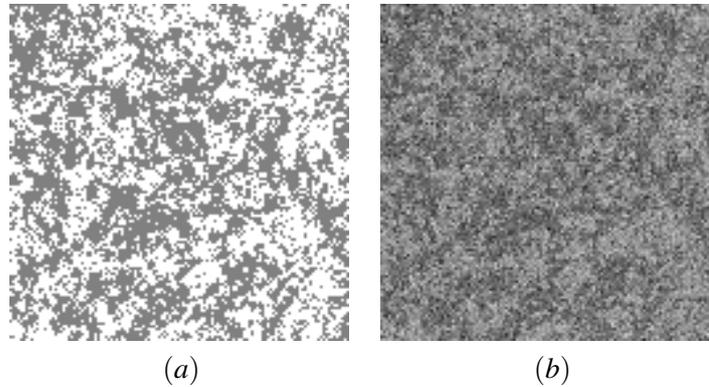


Figura 4.4: (a) Realización del Modelo de Potts con $L = 2$ y $\beta = 0,3$; (b) Datos Gaussianos con $k = 2$.

Para ilustrar cada una de las afirmaciones que se realicen en el trabajo se graficarán las funciones (4.2.5) y (4.2.9), como en la Figura 4.1, las cuales corresponden al caso $L = 3$; $\beta = 0,4$ y $k = 2$. En cada una de las gráficas se marcará con líneas llenas el eje horizontal correspondiente a β , y una línea vertical que señalará el valor real de β correspondiente al modelo. Como las respectivas raíces de (4.2.5) y (4.2.9) son las estimaciones que ofrecen cada uno de los estimadores, se estará en presencia de una buena estimación si la curva correspondiente pasa por la intersección de las líneas negras. De este modo, en el ejemplo dado en la Figura 4.1, se puede ver que ambos estimadores realizan estimaciones similares.

En la presente sección se analizará el Error Cuadrático Medio (ECM), el Sesgo y la Varianza de los estimadores en cuestión, en presencia de lo que llamaremos modelo puro. Esto es, cuando los estadísticos estiman basándose en la realización del modelo de Potts y los datos radiométricos simulados.

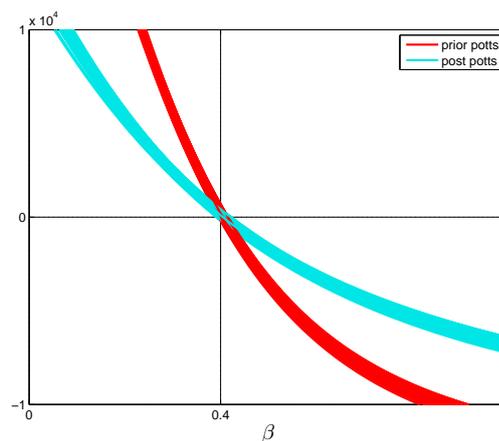


Figura 4.5: Haces de curvas de las funciones $f_{\text{prior}}(\cdot, x)$ en rojo y $f_{\text{post}}(\cdot, x)$ en azul, para el caso $L = 3$; $\beta = 0,4$ y $k = 2$.

Como tenemos 100 simulaciones del Modelo de Potts y de sus respectivos datos radiométricos para cada combinación de casos posibles, se puede graficar un haz de 100 curvas de las funciones (4.2.5) y (4.2.9) respectivamente para cada caso. En la Figura 4.5 se pueden observar dichos haces para el caso $L = 3$; $\beta = 0,4$; y $k = 2$.

Tal y como lo muestran esta gráfica y los Cuadros 4.1-4.4, ambos estimadores poseen sesgos y varianzas pequeñas y del mismo orden. Gráficamente esto se nota respectivamente, por medio de la traslación del haz respecto de la línea negra vertical y el ancho del haz de curvas al interceptar el eje horizontal. Esto produce que los estimadores posean un ECM insignificante y del mismo orden.

L=2						
β	\sqrt{ECM}		Media		Desvío	
	$\hat{\beta}_{prior}$	$\hat{\beta}_{post}$	$\hat{\beta}_{prior}$	$\hat{\beta}_{post}$	$\hat{\beta}_{prior}$	$\hat{\beta}_{post}$
0.1	0.006	0.007	0.101	0.102	0.006	0.006
0.2	0.006	0.007	0.203	0.203	0.005	0.006
0.3	0.008	0.007	0.306	0.306	0.005	0.005
0.4	0.005	0.005	0.403	0.403	0.004	0.004
0.45	0.005	0.006	0.452	0.452	0.005	0.005
0.5	0.006	0.006	0.499	0.499	0.006	0.006
0.6	0.040	0.039	0.609	0.612	0.038	0.037
0.7	0.019	0.021	0.685	0.683	0.012	0.013
0.8	0.034	0.036	0.771	0.769	0.018	0.019
0.9	0.053	0.051	0.852	0.854	0.024	0.023
1	0.082	0.082	0.923	0.924	0.032	0.033
L=3						
β	\sqrt{ECM}		Media		Desvío	
	$\hat{\beta}_{prior}$	$\hat{\beta}_{post}$	$\hat{\beta}_{prior}$	$\hat{\beta}_{post}$	$\hat{\beta}_{prior}$	$\hat{\beta}_{post}$
0.1	0.007	0.008	0.100	0.100	0.007	0.008
0.2	0.007	0.008	0.202	0.203	0.007	0.007
0.3	0.006	0.007	0.304	0.304	0.004	0.005
0.4	0.007	0.008	0.406	0.406	0.004	0.005
0.45	0.004	0.004	0.453	0.452	0.002	0.003
0.5	0.004	0.005	0.501	0.501	0.004	0.004
0.6	0.065	0.068	0.648	0.652	0.043	0.043
0.7	0.019	0.018	0.692	0.695	0.018	0.017
0.8	0.028	0.029	0.774	0.774	0.013	0.014
0.9	0.047	0.044	0.855	0.859	0.017	0.018
1	0.074	0.071	0.929	0.932	0.022	0.022
L=4						
β	\sqrt{ECM}		Media		Desvío	
	$\hat{\beta}_{prior}$	$\hat{\beta}_{post}$	$\hat{\beta}_{prior}$	$\hat{\beta}_{post}$	$\hat{\beta}_{prior}$	$\hat{\beta}_{post}$
0.1	0.007	0.008	0.101	0.101	0.007	0.008
0.2	0.006	0.007	0.199	0.199	0.006	0.007
0.3	0.006	0.006	0.302	0.301	0.005	0.006
0.4	0.006	0.007	0.404	0.403	0.005	0.006
0.45	0.006	0.006	0.454	0.454	0.004	0.004
0.5	0.004	0.004	0.502	0.502	0.003	0.004
0.6	0.093	0.103	0.682	0.694	0.043	0.043
0.7	0.031	0.034	0.719	0.723	0.024	0.025
0.8	0.023	0.022	0.779	0.780	0.010	0.010
0.9	0.043	0.039	0.858	0.863	0.013	0.015
1	0.070	0.064	0.932	0.938	0.018	0.020

Cuadro 4.1: ECM, media y desvío de los estimadores computados sobre 100 realizaciones del modelo de Potts con distintos valores de β , datos Gaussianos con $k = 1$.

L=2						
β	\sqrt{ECM}		Media		Desvío	
	$\hat{\beta}_{prior}$	$\hat{\beta}_{post}$	$\hat{\beta}_{prior}$	$\hat{\beta}_{post}$	$\hat{\beta}_{prior}$	$\hat{\beta}_{post}$
0.1	0.006	0.008	0.101	0.101	0.006	0.008
0.2	0.006	0.008	0.203	0.203	0.005	0.007
0.3	0.008	0.008	0.306	0.306	0.005	0.006
0.4	0.005	0.006	0.403	0.404	0.004	0.005
0.45	0.005	0.007	0.452	0.451	0.005	0.007
0.5	0.006	0.007	0.499	0.500	0.006	0.007
0.6	0.040	0.037	0.609	0.607	0.038	0.037
0.7	0.019	0.016	0.685	0.690	0.012	0.013
0.8	0.034	0.032	0.771	0.774	0.018	0.020
0.9	0.053	0.047	0.852	0.859	0.024	0.024
1	0.082	0.077	0.923	0.929	0.032	0.031
L=3						
β	\sqrt{ECM}		Media		Desvío	
	$\hat{\beta}_{prior}$	$\hat{\beta}_{post}$	$\hat{\beta}_{prior}$	$\hat{\beta}_{post}$	$\hat{\beta}_{prior}$	$\hat{\beta}_{post}$
0.1	0.007	0.010	0.100	0.100	0.007	0.010
0.2	0.007	0.009	0.202	0.202	0.007	0.009
0.3	0.006	0.008	0.304	0.304	0.004	0.007
0.4	0.007	0.010	0.406	0.407	0.004	0.006
0.45	0.004	0.006	0.453	0.453	0.002	0.005
0.5	0.004	0.007	0.501	0.501	0.004	0.007
0.6	0.065	0.074	0.648	0.657	0.043	0.047
0.7	0.019	0.020	0.692	0.695	0.018	0.019
0.8	0.028	0.026	0.774	0.777	0.013	0.014
0.9	0.047	0.044	0.855	0.860	0.017	0.021
1	0.074	0.066	0.929	0.941	0.022	0.030
L=4						
β	\sqrt{ECM}		Media		Desvío	
	$\hat{\beta}_{prior}$	$\hat{\beta}_{post}$	$\hat{\beta}_{prior}$	$\hat{\beta}_{post}$	$\hat{\beta}_{prior}$	$\hat{\beta}_{post}$
0.1	0.007	0.010	0.101	0.101	0.007	0.010
0.2	0.006	0.008	0.199	0.199	0.006	0.008
0.3	0.006	0.008	0.302	0.302	0.005	0.008
0.4	0.006	0.008	0.404	0.403	0.005	0.008
0.45	0.006	0.008	0.454	0.455	0.004	0.006
0.5	0.004	0.006	0.502	0.501	0.003	0.006
0.6	0.093	0.101	0.682	0.686	0.043	0.052
0.7	0.031	0.040	0.719	0.723	0.024	0.032
0.8	0.023	0.021	0.779	0.784	0.010	0.014
0.9	0.043	0.036	0.858	0.870	0.013	0.022
1	0.070	0.063	0.932	0.942	0.018	0.026

Cuadro 4.2: ECM, media y desvío de los estimadores computados sobre 100 realizaciones del modelo de Potts con distintos valores de β , datos Gaussianos con $k = 2$.

L=2						
β	\sqrt{ECM}		Media		Desvío	
	$\hat{\beta}_{prior}$	$\hat{\beta}_{post}$	$\hat{\beta}_{prior}$	$\hat{\beta}_{post}$	$\hat{\beta}_{prior}$	$\hat{\beta}_{post}$
0.1	0.006	0.011	0.101	0.101	0.006	0.011
0.2	0.006	0.011	0.203	0.202	0.005	0.011
0.3	0.008	0.011	0.306	0.305	0.005	0.009
0.4	0.005	0.010	0.403	0.405	0.004	0.008
0.45	0.005	0.009	0.452	0.451	0.005	0.009
0.5	0.006	0.009	0.499	0.500	0.006	0.009
0.6	0.040	0.045	0.609	0.615	0.038	0.043
0.7	0.019	0.020	0.685	0.694	0.012	0.019
0.8	0.034	0.031	0.771	0.785	0.018	0.027
0.9	0.053	0.052	0.852	0.864	0.024	0.038
1	0.082	0.073	0.923	0.938	0.032	0.041

L=3						
β	\sqrt{ECM}		Media		Desvío	
	$\hat{\beta}_{prior}$	$\hat{\beta}_{post}$	$\hat{\beta}_{prior}$	$\hat{\beta}_{post}$	$\hat{\beta}_{prior}$	$\hat{\beta}_{post}$
0.1	0.007	0.012	0.100	0.100	0.007	0.012
0.2	0.007	0.012	0.202	0.202	0.007	0.012
0.3	0.006	0.011	0.304	0.303	0.004	0.010
0.4	0.007	0.011	0.406	0.407	0.004	0.009
0.45	0.004	0.008	0.453	0.453	0.002	0.008
0.5	0.004	0.009	0.501	0.501	0.004	0.009
0.6	0.065	0.075	0.648	0.650	0.043	0.055
0.7	0.019	0.031	0.692	0.696	0.018	0.031
0.8	0.028	0.028	0.774	0.786	0.013	0.025
0.9	0.047	0.040	0.855	0.871	0.017	0.028
1	0.074	0.063	0.929	0.949	0.022	0.037

L=4						
β	\sqrt{ECM}		Media		Desvío	
	$\hat{\beta}_{prior}$	$\hat{\beta}_{post}$	$\hat{\beta}_{prior}$	$\hat{\beta}_{post}$	$\hat{\beta}_{prior}$	$\hat{\beta}_{post}$
0.1	0.007	0.014	0.101	0.101	0.007	0.014
0.2	0.006	0.012	0.199	0.197	0.006	0.012
0.3	0.006	0.012	0.302	0.302	0.005	0.011
0.4	0.006	0.012	0.404	0.401	0.005	0.012
0.45	0.006	0.010	0.454	0.455	0.004	0.009
0.5	0.004	0.009	0.502	0.501	0.003	0.009
0.6	0.093	0.107	0.682	0.686	0.043	0.063
0.7	0.031	0.049	0.719	0.728	0.024	0.040
0.8	0.023	0.024	0.779	0.789	0.010	0.021
0.9	0.043	0.042	0.858	0.869	0.013	0.030
1	0.070	0.061	0.932	0.948	0.018	0.032

Cuadro 4.3: ECM, media y desvío de los estimadores computados sobre 100 realizaciones del modelo de Potts con distintos valores de β , datos Gaussianos con $k = 3$.

L=2						
β	\sqrt{ECM}		Media		Desvío	
	$\hat{\beta}_{prior}$	$\hat{\beta}_{post}$	$\hat{\beta}_{prior}$	$\hat{\beta}_{post}$	$\hat{\beta}_{prior}$	$\hat{\beta}_{post}$
0.1	0.006	0.019	0.101	0.100	0.006	0.019
0.2	0.006	0.018	0.203	0.202	0.005	0.018
0.3	0.008	0.015	0.306	0.304	0.005	0.015
0.4	0.005	0.016	0.403	0.405	0.004	0.015
0.45	0.005	0.015	0.452	0.451	0.005	0.015
0.5	0.006	0.016	0.499	0.499	0.006	0.016
0.6	0.040	0.047	0.609	0.612	0.038	0.046
0.7	0.019	0.029	0.685	0.697	0.012	0.029
0.8	0.034	0.037	0.771	0.789	0.018	0.036
0.9	0.053	0.052	0.852	0.874	0.024	0.045
1	0.082	0.075	0.923	0.962	0.032	0.065

L=3						
β	\sqrt{ECM}		Media		Desvío	
	$\hat{\beta}_{prior}$	$\hat{\beta}_{post}$	$\hat{\beta}_{prior}$	$\hat{\beta}_{post}$	$\hat{\beta}_{prior}$	$\hat{\beta}_{post}$
0.1	0.007	0.020	0.100	0.099	0.007	0.020
0.2	0.007	0.020	0.202	0.202	0.007	0.020
0.3	0.006	0.018	0.304	0.306	0.004	0.017
0.4	0.007	0.016	0.406	0.406	0.004	0.015
0.45	0.004	0.014	0.453	0.453	0.002	0.013
0.5	0.004	0.014	0.501	0.500	0.004	0.014
0.6	0.065	0.078	0.648	0.647	0.043	0.062
0.7	0.019	0.031	0.692	0.697	0.018	0.031
0.8	0.028	0.036	0.774	0.787	0.013	0.033
0.9	0.047	0.049	0.855	0.877	0.017	0.044
1	0.074	0.063	0.929	0.967	0.022	0.054

L=4						
β	\sqrt{ECM}		Media		Desvío	
	$\hat{\beta}_{prior}$	$\hat{\beta}_{post}$	$\hat{\beta}_{prior}$	$\hat{\beta}_{post}$	$\hat{\beta}_{prior}$	$\hat{\beta}_{post}$
0.1	0.007	0.024	0.101	0.101	0.007	0.024
0.2	0.006	0.021	0.199	0.195	0.006	0.020
0.3	0.006	0.021	0.302	0.303	0.005	0.021
0.4	0.006	0.021	0.404	0.401	0.005	0.021
0.45	0.006	0.016	0.454	0.455	0.004	0.015
0.5	0.004	0.015	0.502	0.500	0.003	0.015
0.6	0.093	0.120	0.682	0.698	0.043	0.069
0.7	0.031	0.060	0.719	0.737	0.024	0.047
0.8	0.023	0.037	0.779	0.789	0.010	0.036
0.9	0.043	0.055	0.858	0.870	0.013	0.047
1	0.070	0.064	0.932	0.958	0.018	0.048

Cuadro 4.4: ECM, media y desvío de los estimadores computados sobre 100 realizaciones del modelo de Potts con distintos valores de β , datos Gaussianos con $k = 4$.

Capítulo 5

Segmentaciones

Como se mencionó anteriormente, una segmentación no es más que una asignación de una etiqueta a cada píxel de una imagen. Estas pueden ser supervisadas o no supervisadas, dependiendo de si necesitan o no de una muestra de entrenamiento. El clásico Segmentador MV es un ejemplo de segmentador supervisado, y el algoritmo EM brinda un ejemplo de segmentador no supervisado. Los resultados del segmentador MV dependen de la calidad del entrenamiento (ver Frery et al. [46]), mientras que los resultados de EM dependen de la correcta selección del modelo. Si se trabaja con imágenes ruidosas en las que los modelos de emisión se entrecruzan, ambos segmentadores presentan segmentaciones heterogéneas, las cuales en muchos contextos de trabajo son indeseables. En la práctica se busca que la clase asignada a un determinado píxel esté vinculada a la de sus vecinos, evitando la existencia de píxeles clasificados aisladamente a una determinada clase. Esta problemática se soluciona mediante la inclusión de información contextual al proceso de asignación de etiquetas, y los modelos Markovianos son ideales para tal fin.

En el ámbito de la segmentación de imágenes digitales intervienen dos componentes aleatorias: el mapa de clases X y la información radiométrica I . La distribución de X se denomina modelo a priori del mapa de clases, mientras que la distribución condicional de X dado I es el modelo a posteriori del mismo. En la Sección 3.5 se mostró que si la ley a priori del mapa de clases es un modelo de Potts, entonces bajo determinadas hipótesis sobre el modelo de emisión, la distribución a posteriori también lo será. En la Sección 5.1 se mostrará que esto vale en general para distribuciones de Gibbs generadas por potenciales. La estimación de la segmentación óptima para una determinada imagen es equivalente a encontrar máximos globales o modas de la distribución del modelo a posteriori. En la Sección 5.2 se mostrará como el algoritmo Simulated Annealing (SA) (Geman y Geman [6]), que es una leve modificación de GS, converge iterativamente a un máximo global de una distribución de Gibbs, y que por lo tanto, se puede utilizar para hallar segmentaciones óptimas. La convergencia de SA depende del cumplimiento de una cota sobre un esquema de enfriamiento que lo lleva a ser un algoritmo con una convergencia muy lenta. El no cumplimiento de esta cota produce que la convergencia de SA sea a un máximo local, que generalmente no es un máximo global. En la Sección 5.3 se definirá el algoritmo Iterated Conditional Modes (ICM) (Besag [14]), el cual suele definirse como el algoritmo SA con esquema de enfriado extremo, que no respeta la cota de enfriamiento, y que converge rápidamente al máximo local más cercano a la segmentación de inicio introducida al mismo. Luego, esta optimización en la velocidad de convergencia que presenta SA, hace a ICM más sensible a la elección de la segmentación de inicio. Sin embargo, en el contexto de segmentaciones bajo modelos de Potts, en la Sección 5.4 se verá que el algoritmo ICM brinda segmentaciones homogéneas de calidad, y menos sensibles a la incorrecta selección de las muestras de entrenamiento, como se mostró en el trabajo de Frery et al. [46].

El modelo de Potts isotrópico contiene un parámetro desconocido β denominado temperatura inversa, el cual regula el grado de homogeneidad del mapa, y que puede ser estimado eficientemente por PMV. Cada mo-

delo tiene su propio estimador de PMV, y como tanto el modelo a priori como el modelo a posteriori comparten el parámetro β , el mismo puede estimarse con el estimador de PMV correspondiente al modelo a priori, o al correspondiente al modelo a posteriori. La diferencia entre estos estimadores es que este último agrega directamente al proceso de estimación los datos radiométricos de la imagen, mientras que el estimador a priori sólo los incorpora indirectamente por medio de una segmentación de los mismos. La situación ideal a la hora de estimar, es que las estimaciones se basen en correctas realizaciones del modelo, esto es, que los mapas de clases de las imágenes sean realizaciones del modelo de Potts por ejemplo. Sin embargo, en la práctica se cuenta solamente con los datos radiométricos de la imagen, los cuales deben ser segmentados, para recién proceder con la estimación. Esto implica que la calidad de los datos radiométricos influya en el proceso de estimación, y por ende, es intuitivamente correcto incluirlos para de este modo contar con el total de la información disponible. En la Sección 4.5 se mostró evidencia de que bajo modelo ambos estimadores presentan resultados con errores cuadráticos medios realmente despreciables a muestra finita, no mostrando diferencias significativas entre ellos. En la Sección 5.5 se analiza la sensibilidad de ambos estimadores ante la presencia de un modelo contaminado, esto es, cuando las estimaciones se basan en segmentaciones de los datos, y no en realizaciones reales del modelo. Se mostrará evidencia de que incorporar los datos al proceso de estimación produce estimaciones muy inestables, debidas a que si bien los datos más el modelo contextual conforman la masa de información con la que contamos, al desviarnos del modelo incorporando segmentaciones de los datos, toda la información que incorporan los datos al proceso de estimación se convierte en una fuente de error importante que debe ser dejada de lado. Estos hechos serán denominados informalmente como *suficiencia del mapa de clases*, debido a que la realización del modelo es suficiente para realizar estimaciones precisas del parámetro del modelo, y debido a que agregar los datos no mejora la perfecta precisión de las estimaciones por PMV.

El Capítulo finaliza con una discusión sobre el rol de las estimaciones en el proceso de segmentación por medio de ICM. La elección del parámetro óptimo a la hora de aplicar ICM, u otro segmentador, es un tema que se ha discutido bastante en la Literatura. En el contexto Potts, Jackson y Landgrebe [47] utilizaron el algoritmo ICM con β fijo, y mostraron que una clasificación contextual con muestras pequeñas alcanza una precisión comparable a la obtenida por MV con una muestra de entrenamiento mucho mayor. Arbia et al. [48] también utilizó un valor fijo para este parámetro, elegido por prueba y error, en una configuración de clasificación en dos clases con datos simulados. Melgani y Sapico [49] propusieron un enfoque que denominaron de perturbación mínima, en el cual estiman una vez el modelo y lo mantienen fijo en todas las iteraciones. Besag [50], Arnold y Strauss [51] y Frery et al. [15] propusieron una estimación iterativa de β con la información disponible a través del estimador basado en el modelo a priori. En la tesis se mostrará evidencia de que ICM funciona mejor si se reestima β , debido a que la estimación original basada en la segmentación por medio de MV es baja e insuficiente para que ICM logre homogeneizar correctamente el mapa de clases. Además, se incorpora evidencia de que si se trabaja con imágenes con clases homogéneas que no se mezclan localmente, entonces sus mapas de clases ajustan mejor a modelos de Potts con vecindades de mayor orden, pero que si se quiere trabajar con vecindades de segundo orden, un proceso de filtrado previo es necesario. Por ejemplo, en [52] se muestra evidencia de que segmentar con vecindades de primer orden en el contexto Potts no es una buena opción. Existen en la Literatura trabajos tales como el de Ji y Seymour [53], y más recientemente el de Grelaud et al. [54], en los que se proponen metodologías de selección del modelo.

5.1. Criterio MAP: Máximo a Posteriori.

Supongamos que el espacio de estados es el conjunto de todas las posibles etiquetas

$$E = \{\ell_1, \ell_2, \dots, \ell_L\}.$$

Generalizando lo mostrado en la Sección 3.5, bajo determinados supuestos, si la distribución del modelo a priori del mapa de clases X es π , inducida por el potencial $\Phi = \{\Phi_\Lambda\}_{\Lambda \in \mathfrak{S}}$, y el modelo de emisión en cada pixel para $\ell \in E$ viene dado por las leyes f_ℓ , entonces dada una observación $\mathbf{I} = [I_s]_{s \in \mathcal{S}}$ del modelo de emisión conjunto, la distribución a posteriori del mapa de clases es una distribución de Gibbs Π inducida por el potencial $\Phi^I = \{\Phi_\Lambda^I\}_{\Lambda \in \mathfrak{S}}$ dado por

$$\Phi_\Lambda^I(x) = \begin{cases} \Phi_s(x) - \ln f_{x_s}(I_s) & \text{si } \Lambda = \{s\}, s \in \mathcal{S} \\ \Phi_\Lambda(x) & \text{en otro caso,} \end{cases} \quad (5.1.1)$$

ya que en tal caso la función de energía sobre $\Lambda \in \mathfrak{S}$ viene dada por

$$H_\Lambda^{\Phi^I}(x) = \sum_{\Delta \in \mathfrak{S}(\Lambda)} \Phi_\Delta^I(x) = - \sum_{s \in \Lambda} \ln f_{x_s}(I_s) + \sum_{\Delta \in \mathfrak{S}(\Lambda)} \Phi_\Delta(x) = - \sum_{s \in \Lambda} \ln f_{x_s}(I_s) + H_\Lambda^\Phi(x),$$

y la densidad condicional a posteriori de observar la configuración de clases $x_\Lambda \in E^\Lambda$, dado que se observaron los datos radiométricos $I_\Lambda \in (\mathbb{R}^B)^\Lambda$ y que se fijó por cuestiones teóricas la configuración de clases $x^\Lambda \in E^{\mathcal{S} \setminus \Lambda}$, viene dada por

$$\Pi_\Lambda(x_\Lambda | x^\Lambda, I_\Lambda) \propto \exp \left\{ -H_\Lambda^{\Phi^I}(x) \right\} = \exp \left\{ \sum_{s \in \Lambda} \ln f_{x_s}(I_s) - H_\Lambda^\Phi(x) \right\}.$$

Como la segmentación óptima es la más probable dadas las observaciones, el problema de segmentación se reduce a encontrar el Máximo a Posteriori (MAP), esto es, encontrar un mapa de clases $x_\Lambda^* \in E^\Lambda$ que verifique

$$x_\Lambda^* = \arg \max_{x_\Lambda \in E^\Lambda} \Pi_\Lambda(x_\Lambda | x^\Lambda, I_\Lambda), \quad (5.1.2)$$

lo cual requiere hallar máximos globales de una condicional de una distribución de Gibbs.

5.2. Simulated Annealing (SA)

El objetivo es ahora mostrar el algoritmo Simulated Annealing (SA) [6], el cual tiene como meta resolver (5.1.2). Se verá como una modificación en Gibbs Sampler puede llevar a encontrar las modas de las distribuciones en estudio.

Dada una distribución de Gibbs π con funciones de energía H_Λ^Φ inducidas por el potencial Φ , se denota con π^T , siendo $T > 0$, a la distribución de Gibbs con funciones de energía H_Λ^Φ/T y densidades condicionales

$$\pi_\Lambda^T(x_\Lambda | x^\Lambda, I_\Lambda) \propto \exp \left\{ -H_\Lambda^\Phi(x)/T \right\},$$

donde T es conocida en el ámbito de la física como la constante de temperatura. Todas las distribuciones π_Λ^T tienen los mismos máximos globales, y mientras más chico sea T , se tendrán distribuciones con modas más marcadas, y simulaciones por medio de GS con mayor probabilidad de encontrarse cerca de las modas de la distribución π_Λ . Luego, se puede pensar en un esquema de enfriado $T(t)$, que verifique

$$\lim_{t \rightarrow \infty} T(t) = 0, \quad (5.2.3)$$

con el cual, en cada corrida de GS, se utilice la distribución $\pi_\Lambda^{T(t)}$ en lugar de π_Λ .

Esto es, supongamos que en el t -ésimo barrido tenemos una configuración $x_\Lambda \in E^\Lambda$ y que deseamos actualizar la etiqueta que se encuentra sobre un determinado sitio $s \in \Lambda$. Luego, se asignará la etiqueta $\ell \in E$ al sitio s con probabilidad dada por:

$$\pi_s^{T(t)}(\ell|x^s) \propto \exp \left\{ - \sum_{\Lambda \in \mathfrak{S}(s)} \Phi_\Lambda(\ell x^s) \right\}.$$

El algoritmo SA propone que en cada visita a cada sitio $s \in \Lambda$ se asigne la etiqueta $\ell \in E$ con probabilidad dada por:

$$\pi_s^{T(t)}(\ell|x^s) \propto \exp \left\{ - \sum_{\Lambda \in \mathfrak{S}(s)} \Phi_\Lambda(\ell x^s) / T(t) \right\},$$

donde $\mathfrak{S}(s) = \{\Lambda \in \mathfrak{S} : s \in \Lambda\}$. Sea

$$\Omega_0 = \left\{ \xi \in E^\Lambda : H_\Lambda^\Phi(\xi x^\Lambda) = \min_{\eta} H_\Lambda^\Phi(\eta x^\Lambda) \right\}$$

el conjunto de modas o máximos globales de $\pi_\Lambda(\cdot|x^\Lambda)$ y sea

$$\Delta = \max_{\xi} H_\Lambda^\Phi(\xi x^\Lambda) - \min_{\xi} H_\Lambda^\Phi(\xi x^\Lambda).$$

El Teorema B de [6] muestra que SA converge uniformemente a un elemento de Ω_0 , siempre que se respete la siguiente cota sobre el esquema de enfriado:

$$T(t) \geq \frac{\Delta * \#\Lambda}{\ln(t)}. \quad (5.2.4)$$

En resumen, para hallar las modas de la distribución de Gibbs condicional $\pi_\Lambda(\cdot|x^\Lambda)$ sobre E^Λ , SA propone seguir los siguientes pasos:

- 1) obtener una realización de inicio $x^{(0)} \in E^\Lambda$ con algún procedimiento.
- 2) elegir un esquema de visita adecuado y un esquema de enfriado que satisfaga (5.2.3) y (5.2.4).
- 3) realizar un barrido (visita de todos los pixeles), sorteando aleatoriamente un estado en cada visita, con probabilidad proporcional a

$$\exp \left\{ - \sum_{\Lambda \in \mathfrak{S}(s)} \Phi_\Lambda(\ell x^s) / T(t) \right\}$$

para cada $\ell \in E$, siendo $s \in \Lambda$ el pixel que se visita y $x \in E^S$ la configuración previa a la visita.

- 4) realizar tantas barridas como se necesiten para alcanzar algún criterio de convergencia.

5.3. Iterated Conditional Modes (ICM)

Si bien SA converge a un máximo global, en general es muy lento (por (5.2.4)) y pesado computacionalmente. Por este motivo es que Besag en [14, 55] lo modificó, proponiendo un esquema de enfriado extremo. Esto es, en el paso 3) de SA, se elige la etiqueta que maximiza la probabilidad de ser observada dado el contexto. Este cambio produce que el algoritmo converja rápidamente a un máximo local, que eventualmente puede ser un máximo global. Este algoritmo recibe el nombre de ICM y será expuesto a continuación.

Notemos con

$$\Omega_1 = \left\{ \xi \in E^\Lambda : H_\Lambda^\Phi(\xi|x^\Lambda) = \min_{s \in \Lambda, \ell \in E} H_\Lambda^\Phi(\ell x^s) \right\}$$

al conjunto de los mínimos locales de $\pi_\Lambda(\cdot|x^\Lambda)$.

Como ya se dijo, SA con un esquema de enfriado extremo no tiene asegurada la convergencia a un máximo global, ya que no cumple con la cota inferior de enfriado requerido, sin embargo, ICM converge a un máximo local, que circunstancialmente puede ser un máximo global [14, 55]. En resumen, para hallar un máximo local de una condicional $\pi_\Lambda(\cdot|x^\Lambda)$ de una distribución de Gibbs, ICM propone:

- 1) obtener una configuración de inicio $x^{(0)} \in E^\Lambda$ con algún procedimiento;
- 2) elegir un esquema de visita sobre Λ ;
- 3) realizar un barrido (visita de todos los pixeles de Λ), asignando a cada sitio $s \in \Lambda$ la etiqueta $\ell \in E$ que maximice

$$\exp \left\{ - \sum_{\Lambda \in \mathfrak{S}(s)} \Phi_\Lambda(\ell x^s) \right\},$$

o equivalentemente, que maximice

$$- \sum_{\Lambda \in \mathfrak{S}(s)} \Phi_\Lambda(\ell x^s),$$

siendo $x \in E^\Lambda$ la configuración previa a la visita.

- 4) se realizan tantas barridas como se necesiten para alcanzar algún criterio de convergencia.

Luego ICM resuelve parcialmente (5.1.2), ya que encuentra un máximo local, que puede no ser un máximo global. Sin embargo su costo computacional es despreciable a comparación del requerido por SA, y en el contexto de las segmentaciones, las soluciones que brinda son en general satisfactorias y mejores que las obtenidas por procedimientos no contextuales como MV y EM, como veremos a continuación.

5.4. ICM y el modelo de Potts.

En la presente sección, particularizaremos el algoritmo ICM visto en la Sección 5.3 al contexto de segmentaciones basadas en modelos de Potts. De este modo mostraremos un segmentador de imágenes sencillo matemática y computacionalmente.

Supongamos que la distribución a priori del mapa de clases es un modelo de Potts isotrópico sin campo externo con distribuciones condicionales dadas por

$$\pi_\Lambda^\beta(x_\Lambda|x^\Lambda) \propto \exp\{\beta H_\Lambda(x)\}.$$

Luego, teniendo en cuenta lo visto en la Sección 3.5, dada una imagen I_Λ , la distribución a posteriori del mapa de clases x_Λ viene dada por la distribución de Gibbs cuyas condicionales tienen la forma

$$\Pi_\Lambda^\beta(x_\Lambda|x^\Lambda, I_\Lambda) \propto \exp \left\{ \sum_{s \in \Lambda} \ln f_{x_s}(I_s) + \beta H_\Lambda(x) \right\}, \quad (5.4.5)$$

las cuales son inducidas por el potencial $\Phi^{\beta, I} = \{\Phi_{\Lambda}^{\beta, I}\}_{\Lambda \in \mathcal{G}}$ dado por

$$\Phi_{\Lambda}^{\beta, I}(x) = \begin{cases} -\ln f_{x_s}(I_s) & \text{si } \Lambda = \{s\}, s \in S; \\ -\beta & \text{si } \Lambda = \{s, t\} \text{ y } x_s = x_t, \text{ con } s \in \partial_t; \\ 0 & \text{en otro caso.} \end{cases}$$

Esto significa que si la distribución a priori del mapa de clases se rige por un modelo de Potts isotrópico sin campo externo, entonces la distribución a posteriori se regirá por un modelo de Potts isotrópico con campo externo $(-\ln f_{x_s}(I_s))_{s \in S}$.

En particular, las probabilidades locales de la distribución a posteriori vienen dadas por

$$\Pi_s^{\beta}(\ell | x^s, I_s) \propto \exp\{\ln f_{\ell}(I_s) + \beta H_s(\ell)\}. \quad (5.4.6)$$

Luego dada una imagen I_{Λ} , se debe hallar la segmentación x_{Λ}^* que satisface (5.1.2), o hallar la solución subóptima propuesta por ICM. En este contexto, ICM propone:

- 1) pensar que la clasificación original obtenida por MV o EM es una realización de un campo inicial $X^{(0)}$ con distribución arbitraria;
- 2) elegir un esquema de visita;
- 3) realizar una corrida (visita de todos los pixeles), cambiando la etiqueta de cada pixel por la etiqueta $\ell \in E$ que maximice

$$g(\ell) = \ln f_{\ell}(I_s) + \beta H_s(\ell) \quad (5.4.7)$$

donde s es el pixel que se visita, x es la segmentación previa a la visita, I_s es la observación original de la imagen en el pixel s ;

- 4) realizar corridas hasta alcanzar algún criterio de convergencia.

El primer término de (5.4.7) es equivalente al que usa el clasificador MV. El segundo término es la componente contextual, y si $\beta > 0$, pone más peso en aquellas clases que rodean al pixel s . La influencia contextual es cuantificada por el valor del parámetro β . Cuando $\beta = 0$ la regla se reduce a la dada por MV, pero cuando $\beta \rightarrow \infty$ el efecto es al revés, es decir, los datos observados no tienen influencia en la regla. Este parámetro es desconocido y, por lo tanto, debe ser especificado en forma determinística, o estimado basándose en la segmentación inicial, o re-estimado siguiendo algún criterio.

A	B	C	A	B	C	A	B	C
D	E	F	D	E	F	D	E	F
G	H	I	G	H	I	G	H	I
A	B	C	A	B	C	A	B	C
D	E	F	D	E	F	D	E	F
G	H	I	G	H	I	G	H	I
A	B	C	A	B	C	A	B	C
D	E	F	D	E	F	D	E	F
G	H	I	G	H	I	G	H	I

Figura 5.1: Diagrama del esquema de visita utilizado en las implementaciones.

El esquema de visita elegido es un punto importante de analizar, debido a que una mala elección del mismo, puede provocar que en la práctica el algoritmo muestre un comportamiento oscilatorio a partir de una determinada iteración. Esto sucede por ejemplo si el esquema de visita sigue el orden lexicográfico de los píxeles de Λ . En la tesis se utilizó el orden diagramado en la Figura 5.1, y consiste en dividir la imagen en ventanas de 3×3 , y actualizar las etiquetas con la misma letra al mismo tiempo, siguiendo el orden dado por el abecedario. Si se trabaja con vecindades de orden 2, este diagrama garantiza la convergencia a un único mapa, evitando oscilaciones. Si se desea trabajar con vecindades de orden mayor, se debe elegir un esquema de visita en el que dos píxeles visitados sucesivamente no contengan vecinos en común. De este modo, el criterio de convergencia utilizado a lo largo de la tesis es que no se produzca ningún cambio de etiqueta de una iteración a la otra.

En la Figura 5.2 se muestra la imagen de la mandíbula de una rata, con segmentaciones por medio de MV e ICM con distintos valores de β . Supongamos que la imagen tiene 3 clases principales homogéneas: hueso-diente, cartílago y fondo; cada una con un modelo de emisión Gaussiano con medias 91,82; 66,38; 51,62 y desvíos 16,30; 12,01; 6,25 respectivamente. Claramente, a medida que aumenta β las clasificaciones crecen en homogeneidad, y por lo tanto determinar el parámetro correcto es importante en el proceso de segmentación de imágenes.

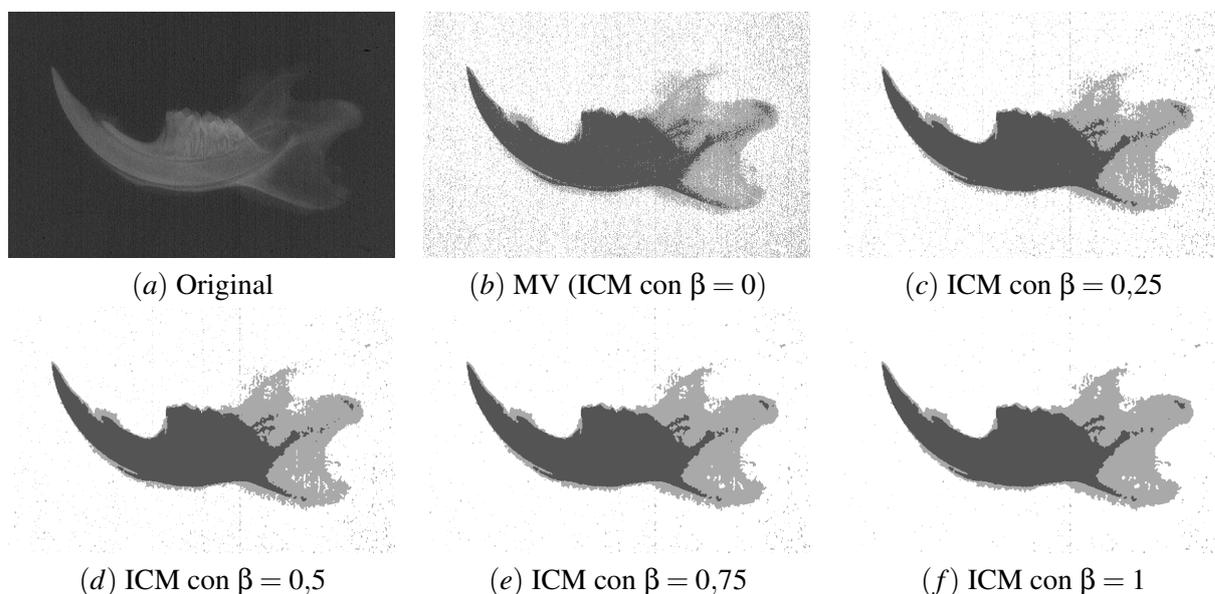


Figura 5.2: (a) Imagen Original; (b) MV de (a); (c)-(f) ICM de (a) con distintos β .

Ante la presencia de modelos de emisión multimodales, es más adecuado ajustar mezclas de gaussianas (EM) por clase, para luego clasificar por MV. En la Figura 5.3 se muestra en (a) la imagen Landsat sobre la ciudad de San Juan-Argentina. Para obtener (b), se consideró una muestra de entrenamiento por clase y se ajustó una mezcla de dos Gaussianas por clase, para luego clasificar por MV. Al igual que en el ejemplo anterior, a pesar de que la imagen es de mucho mejor calidad, se pueden observar zonas heterogéneas dentro del mapa de clases. Una vez más se puede ver en (c)-(f) que ICM presenta mapas más homogéneos a medida que crece β . Sin embargo se debe tener cuidado con el sobresuavizado, el cual no es deseable en diversas aplicaciones. Las segmentaciones (g)-(j) presentan zooms de las imágenes (a),(c),(e),(f) respectivamente, con el fin de mostrar el efecto que produce la elección de β arbitrariamente grande. La resolución espacial que posee la imagen produce que en lugar de ver el fino río que sale del dique en la zona derecha del lago, se observe solo la vegetación de

sus costados. Por tal motivo, los mapas de clases clasifican como vegetación a tal río (ver (h)-(j)). Esta región del mapa que se segmenta como vegetación es fina, y es clasificada a la clase vegetación por la evidencia radiométrica que presenta, a pesar de que la evidencia contextual pretende clasificar a tal región como montaña. Luego, valores grandes de β producen que ICM entrecorte el río, tal y como se muestra en la zona inferior de (j). Por ello es que es importante la determinación de β a la hora de segmentar. En las próximas secciones abordaremos el problema de la estimación del parámetro, y de la relación entre estimación y clasificación de imágenes.

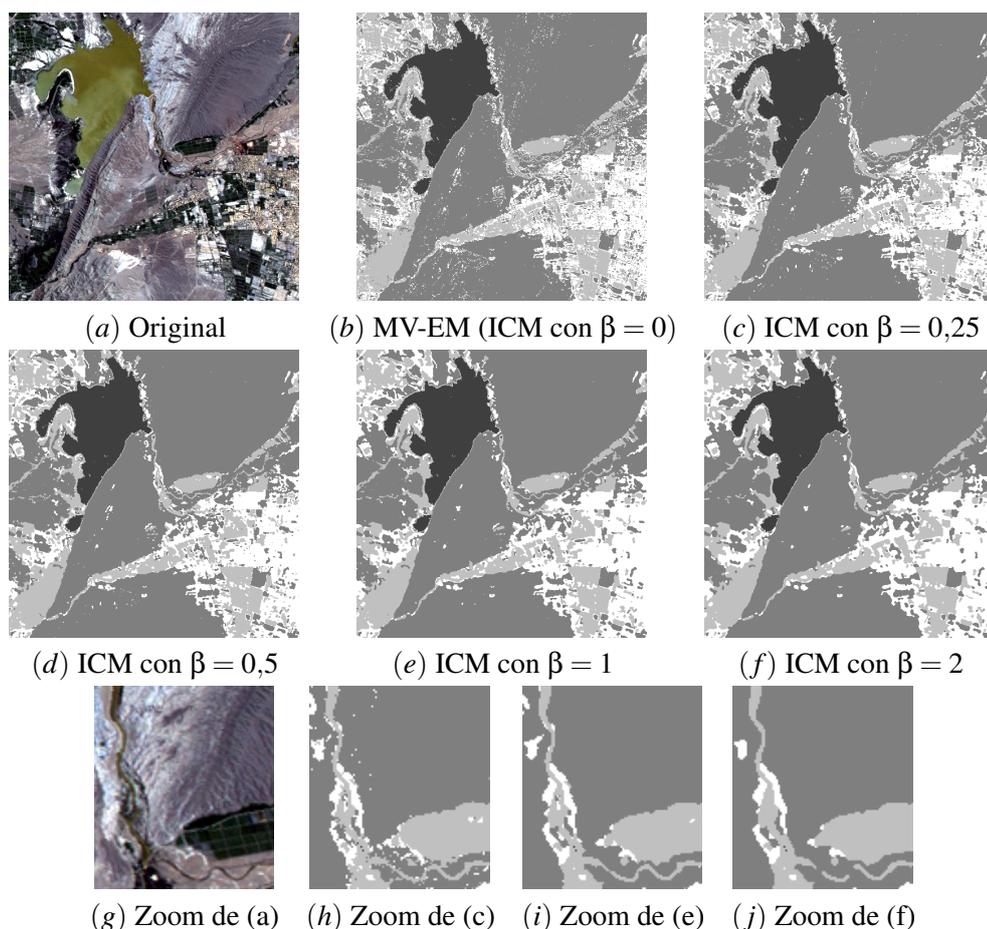


Figura 5.3: (a) Imagen Original; (b) EM-MV de (a); (c)-(f) ICM de (a) con distintos β ; (g)-(j) Zoom de las imágenes (a),(c),(e),(f) respectivamente.

5.5. Sensibilidad

En el presente Capítulo se vincularán los dos temas principales que se trataron en la tesis: Segmentación y Estimación. Todos los análisis se harán en el contexto Potts, pero pueden adaptarse a modelos más generales. Por un lado, cuando se habló de segmentaciones por intermedio de ICM en la Sección 5.4, se observó la necesidad de proponer un valor para β , debido a que tal elección influye en la calidad de la segmentación. Por otro lado, cuando se habló de estimaciones en la Sección 4.1, se observó la necesidad de segmentar los

datos radiométricos, para de este modo tener un mapa de clases con el cual estimar. Esto muestra la fuerte interrelación que existe entre Segmentar y Estimar.

En la Sección 4.5 se mostró la gran precisión a muestra finita de los estimadores de PMV, pero nada se ha dicho aún en la tesis sobre la influencia del procedimiento elegido para segmentar en la precisión de las estimaciones. El hecho de utilizar una segmentación de I en lugar de una realización del modelo, produce que las estimaciones se realicen fuera de las hipótesis bajo las cuales su performance es óptima. El rendimiento de los estimadores en este modelo contaminado se denominará sensibilidad de los estimadores. Finalizaremos el Capítulo analizando la sensibilidad de las segmentaciones, esto es, cuan influyente es la correcta selección y estimación del modelo en el proceso de segmentación.

5.5.1. Sensibilidad de los Estimadores.

La sensibilidad es una propiedad que valora la performance del estimador ante la presencia de datos que no son exactamente realizaciones del modelo. En el contexto de la segmentación de imágenes, es interesante ver, en primera instancia, cómo se comportan los estimadores si en lugar de utilizar el modelo de Potts simulado, se utiliza el mapa de clases que resulta de aplicar el clásico segmentador bayesiano de Máxima Verosimilitud (MV) con los parámetros reales del modelo. En segunda instancia, se hará un estudio análogo, pero utilizando la segmentación dada por ICM.

En la Sección 4.4.2 se habló sobre la importancia que tiene la derivada de las funciones $f_{\text{prior}}(\cdot, x)$ y $f_{\text{post}}(\cdot, x)$ en la sensibilidad de los estimadores $\hat{\beta}_{\text{prior}}$ y $\hat{\beta}_{\text{post}}$, y se calcularon las mismas. Allí se dedujo que el estimador $\hat{\beta}_{\text{post}}$ es más sensible por poseer una menor pendiente. En la presente Sección incorporaremos evidencias de tal inestabilidad.

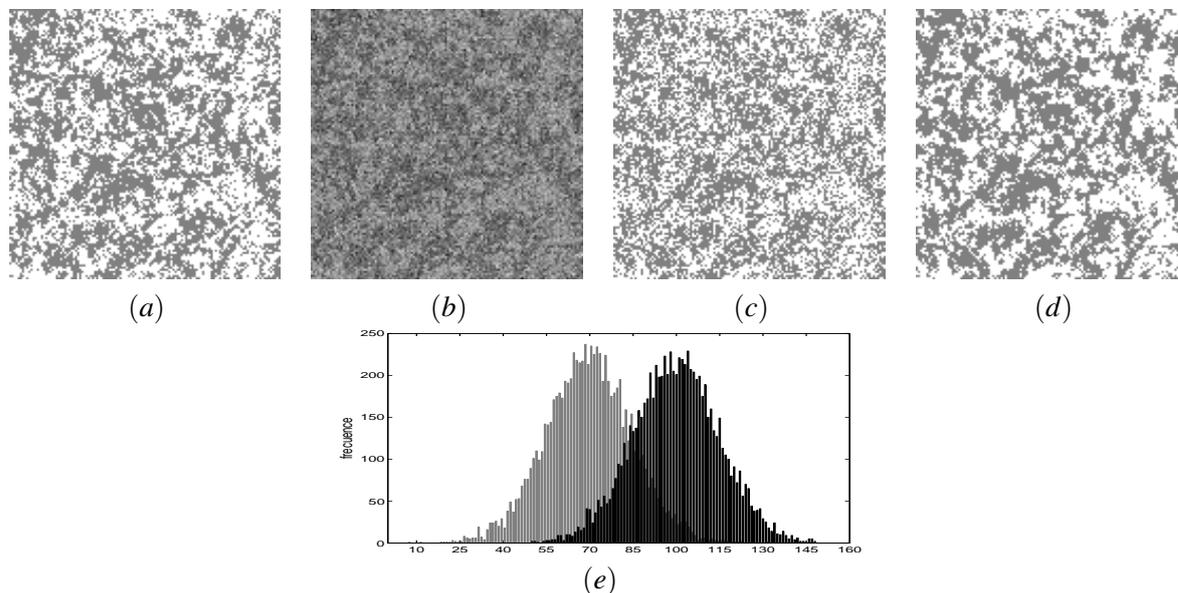


Figura 5.4: (a) Realización del Modelo de Potts con $L = 2$ y $\beta = 0,3$; (b) Datos Gaussianos con $k = 2$; (c) MV de (b); (d) ICM con $\beta = 0,3$ fijo y punto de inicio (c); (e) histograma de (b) por clase según (a).

En la Figura 5.4 se muestra en (a) una realización x_Λ del modelo de Potts con $L = 2$ y $\beta = 0,3$ por medio

de 100 iteraciones de SW; en (b) los datos radiométricos Gaussianos I_Λ con desvíos iguales a 15 y medias 70 y 100 respectivamente (distanciadas a $k = 2$ desvíos); en (e) los histogramas de los modelos de emisión de cada clase; en (c) la clasificación $\hat{x}_{MV,\Lambda}$ por MV de I_Λ ; y en (d) la clasificación $\hat{x}_{ICM,\Lambda}$ por medio de ICM de I_Λ con mapa de inicio $\hat{x}_{MV,\Lambda}$ y $\beta = 0,3$ fijo. La base de datos con la que se realizará el análisis está conformada por estas cuatro imágenes para cada combinación

$$\{(L, \beta, k) : L \in \{2, 3, 4\}, k \in \{1, 2, 3, 4\}, \beta \in \{0.1, 0.2, 0.3, 0.4, 0.45, 0.5, 0.6, 0.7, 0.8, 0.9, 1\}\}.$$

Caso MV

En la presente sección veremos la performance de los estimadores cuando en lugar de utilizar la realización del modelo de Potts, se utiliza la segmentación $\hat{x}_{MV,\Lambda}$ de los datos radiométricos I_Λ sobre $\Lambda \in \mathfrak{S}$. Por cuestiones notacionales, consideraremos una configuración $\hat{x}_{MV} \in E^{\mathfrak{S}}$, que sobre Λ contiene al mapa de clases $\hat{x}_{MV,\Lambda} \in E^\Lambda$. Este análisis muestra evidencia de que al segmentar por medio de ICM, la estimación inicial de β , que se basa en la segmentación no contextual que brinda MV, es una subestimación del parámetro real, y que por lo tanto, es necesario re-estimarlo.

En cada uno de los gráficos de la Figura 5.5 se ilustra la sensibilidad de los estimadores para mapas x_Λ simulados con: $L = 2$; $\beta = 0,3$; y k variando de 1 a 4. En cada gráfico aparecen en rojo (cyan) las gráficas correspondientes respectivamente a las funciones $f_{\text{prior}}(\cdot, x)$ y $f_{\text{prior}}(\cdot, \hat{x}_{MV})$ ($f_{\text{post}}(\cdot, x)$ y $f_{\text{post}}(\cdot, \hat{x}_{MV})$). En línea llena aparecen las gráficas correspondientes al modelo puro, y en línea entre cortada las gráficas de las funciones $f_{\text{prior}}(\cdot, \hat{x}_{MV})$ y $f_{\text{post}}(\cdot, \hat{x}_{MV})$ del modelo contaminado.

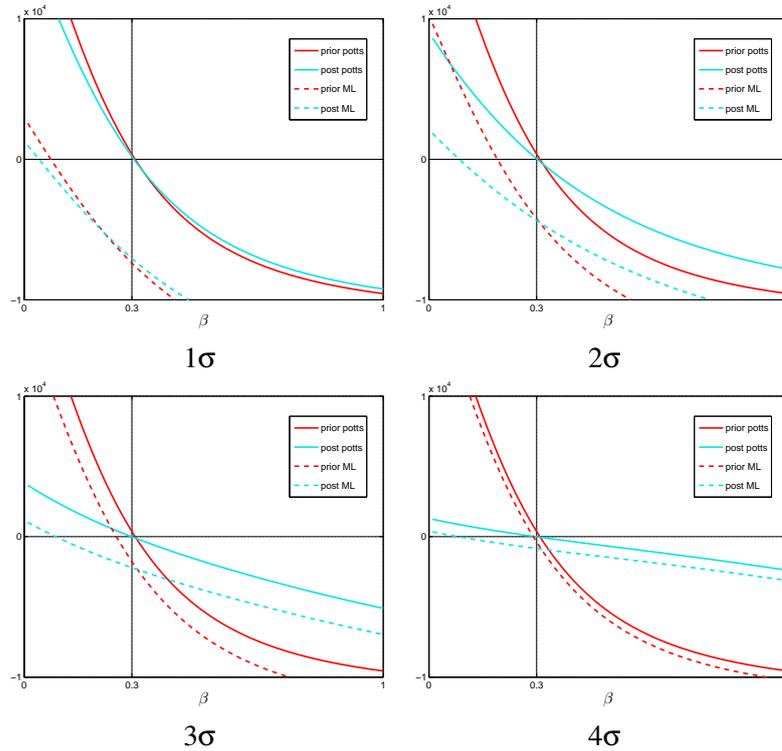


Figura 5.5: Sensibilidad de los estimadores para $L = 2$ y $\beta = 0,3$ al variar k . Muestra como se modifica la pendiente de las curvas correspondientes al modelo a posteriori y como se acercan las curvas correspondientes a los modelos contaminados y a la de los modelos puros, a medida que se distancian los modelos de emisión.

Las curvas rojas en línea llena son muy similares en todos los gráficos ya que k no afecta a $f_{\text{prior}}(\cdot, x)$. Por otro lado, las curvas en cian y en línea llena muestran el efecto analizado en la Sección 4.4.2, esto es, a medida que la diferencia entre las medias es mayor, la curva tiende a convertirse en una recta horizontal. A pesar de esto, se observa que ambos estimadores son precisos al trabajar sobre el modelo puro, tal y como se estudió en la Sección 4.5. Las curvas entre cortadas son casi paralelas a las curvas correspondientes al modelo puro. Mientras más distanciadas estén las medias, más parecida será la segmentación producida por MV al mapa original, y por ende las curvas del modelo contaminado estarán más cerca de las producidas sobre el modelo puro. A pesar de que el desplazamiento que produce trabajar con el mapa estimado por MV es del mismo orden para ambos estimadores, el estimador a posteriori se ve más influenciado por presentar curvas con menor pendiente. De este modo, hay dos conceptos complementarios al distanciarse las medias: curvas paralelas del mismo color más cerca y mayor sensibilidad producida por la disminución de la pendiente de las curvas. Estos dos conceptos se complementan en el estimador a posteriori haciendo casi indistinto para su mala performance, la distancia entre las medias. En el caso del estimador a priori el segundo concepto no se presenta, y por ende las estimaciones son mejores cuando las medias están lejos.

Se puede apreciar que $f_{\text{prior}}(\cdot, \hat{x}_{MV}) \leq f_{\text{prior}}(\cdot, x)$ y que $f_{\text{post}}(\cdot, \hat{x}_{MV}) \leq f_{\text{post}}(\cdot, x)$. Esto se debe a que la altura de las gráficas la regula el primer término de sus respectivas expresiones. Este término es mayor en el modelo puro porque los mapas son más homogéneos que los datos por MV. Esto implica que las estimaciones con el modelo contaminado sean menores y, por ende, sesgadas a izquierda. La Figura 5.7 muestra el sesgo de ambos estimadores en función de β , variando L entre 2 y 4, y para k entre 1 y 4, cuando el modelo se contamina con \hat{x}_{MV} . Claramente ambos estimadores presentan un marcado sesgo a izquierda, que sólo el estimador a priori lo reduce cuando k es grande y β toma valores menores que 0,5.

Por otro lado, la varianza de los estimadores sigue siendo pequeña al trabajar con el modelo contaminado tal y como se muestra en la Figura 5.6, en la cual, al igual que en la Figura 4.5, se muestran los haces de curvas correspondientes a ambos estimadores en presencia del modelo contaminado, para el caso de $L = 2$; $\beta = 0,2$; y k igual a 1 y 4.

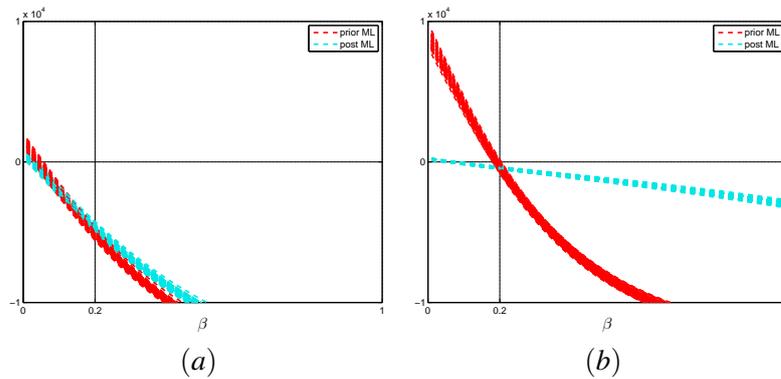


Figura 5.6: Haces de curvas de las funciones $f_{\text{prior}}(\cdot, \hat{x}_{MV})$ y $f_{\text{post}}(\cdot, \hat{x}_{MV})$, para el caso $L = 2$ y $\beta = 0,2$. En (a) aparece el caso $k = 1$, y en (b) el caso $k = 4$. Muestra el sesgo a izquierda y la reducida varianza que presentan los estimadores al trabajar con el modelo contaminado por la segmentación por MV.

En este sentido entonces, se observa que, si se utiliza la segmentación por MV como entrada para los estimadores, entonces es conveniente utilizar el estimador a priori. Sin embargo, no es conveniente estimar basándose en tal mapa.

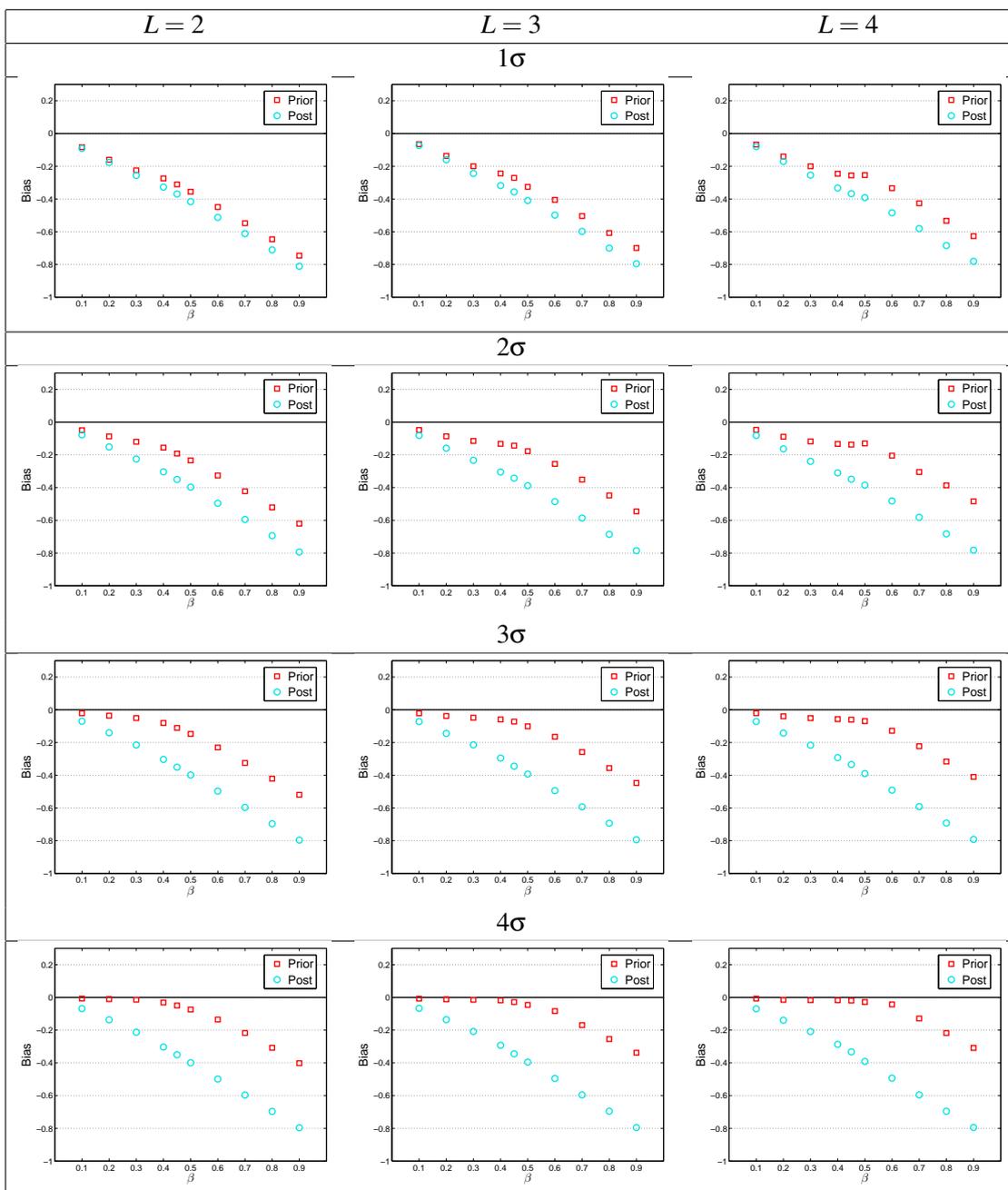


Figura 5.7: Sesgo de los estimadores al utilizar \hat{x}_{MV} , al variar k entre 1 y 4, y L entre 2 y 4. Muestra evidencia del sesgo negativo que poseen ambos estimadores producido por estimar basándose en mapas de clases ruidosos obtenidos por el clasificador no contextual MV.

Caso ICM

Comencemos haciendo un análisis bajo simulación de la velocidad de convergencia de ICM. Utilizando la base de datos con datos radiométricos para distintos valores de β , L y k , se segmentó por ICM cada una de las imágenes y se consideró la cantidad de iteraciones que necesitó ICM para converger. En la Figura 5.8 se

representan tales cantidades para los distintos casos planteados en la tesis. En una primera instancia se observa que ICM converge rápidamente en a lo sumo 10 iteraciones en la mayoría de situaciones. Se puede apreciar además que para todo posible valor de L , a medida que los modelos de emisión se distancian, ICM converge en menos iteraciones, debido a que cuando los modelos de emisión son muy parecidos, ICM requiere de un mayor esfuerzo para homogeneizar el mapa de clases dado por MV. Por otro lado, al variar L , no se observan diferencias en la cantidad de iteraciones hasta converger, excepto en el caso $k = 1$, donde al aumentar la cantidad de clases con modelos de emisión similares, ICM aumenta la cantidad de iteraciones necesarias. Por último, se observa que β no influye en la cantidad de iteraciones, a excepción del caso $k = 1$.

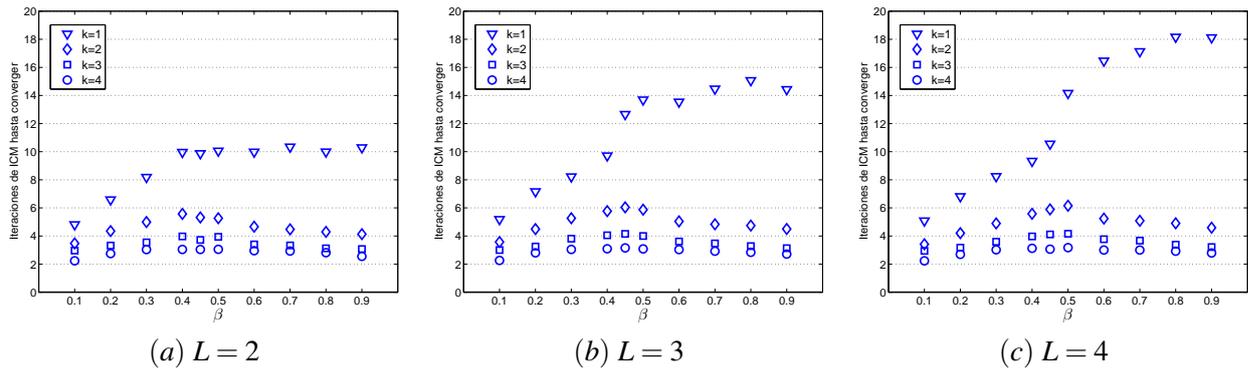


Figura 5.8: Cantidad de iteraciones que requiere ICM hasta converger.

Continuando con el análisis de sensibilidad, veamos como influye la inclusión de la segmentación por ICM como entrada para ambos estimadores. Análogamente que en el caso MV, notaremos con \hat{x}_{ICM} a alguna configuración sobre E^S que contenga sobre Λ la segmentación por medio de ICM de los datos radiométricos observados I_Λ . Esta segmentación por medio de ICM se realizará utilizando β fijo e igual al valor real de simulación. En la Figura 5.9 se pueden observar las gráficas de las funciones $f_{\text{prior}}(\cdot, \hat{x}_{MV})$, $f_{\text{prior}}(\cdot, x)$ y $f_{\text{prior}}(\cdot, \hat{x}_{ICM})$ en rojo, y de las funciones $f_{\text{post}}(\cdot, \hat{x}_{MV})$, $f_{\text{post}}(\cdot, x)$ y $f_{\text{post}}(\cdot, \hat{x}_{ICM})$ en cyan, para diferentes diferencias de medias. En línea llena aparecen las curvas correspondientes al modelo puro, y respectivamente, en línea entrecortada y en línea de puntos las correspondientes al modelo contaminado con los mapas provenientes de MV e ICM (esto es, cuando en las estimaciones en lugar de usar el mapa simulado, usamos \hat{x}_{MV} y \hat{x}_{ICM}). Se puede ver que $f_{\text{prior}}(\cdot, \hat{x}_{MV}) \leq f_{\text{prior}}(\cdot, \hat{x}_{ICM})$ y $f_{\text{post}}(\cdot, \hat{x}_{MV}) \leq f_{\text{post}}(\cdot, \hat{x}_{ICM})$, y esto se debe a que ICM homogeneiza el mapa de inicio \hat{x}_{MV} , provocando estimaciones más grandes que las provenientes de contaminar el modelo con \hat{x}_{MV} .

Además, se verifica que $f_{\text{prior}}(\cdot, x) \leq f_{\text{prior}}(\cdot, \hat{x}_{ICM})$ y $f_{\text{post}}(\cdot, x) \leq f_{\text{post}}(\cdot, \hat{x}_{ICM})$. Esto se debe a que \hat{x}_{ICM} es un máximo local de (5.4.5), y como tal, presenta una mayor suavidad que la que contiene una simple realización del modelo (5.4.5), que no es necesariamente una moda del modelo.

Al igual que antes, se puede observar que las curvas de $f_{\text{post}}(\cdot, \hat{x}_{ICM})$ pierden pendiente al aumentar k . Por otro lado, se puede notar que, a medida que crece k , las curvas de los modelos contaminados se acercan más a las curvas del modelo puro. Esto se debe a que mientras más distantes estén las medias, a ICM le costará más generar regiones conexas en el mapa, porque la evidencia contextual será débil en comparación de la evidencia radiométrica.

En las gráficas de la Figura 5.10 se muestra el sesgo de ambos estimadores en función de β , al variar L entre 2 y 4, y k entre 1 y 4. Se puede observar que ambos estimadores presentan un sesgo positivo, proveniente de sus respectivas sobreestimaciones. Solamente se aprecian sesgos negativos mínimos para valores muy grandes de β . Se puede notar que k influye en mayor medida que L , en la precisión de las estimaciones. Para $k = 4$ la

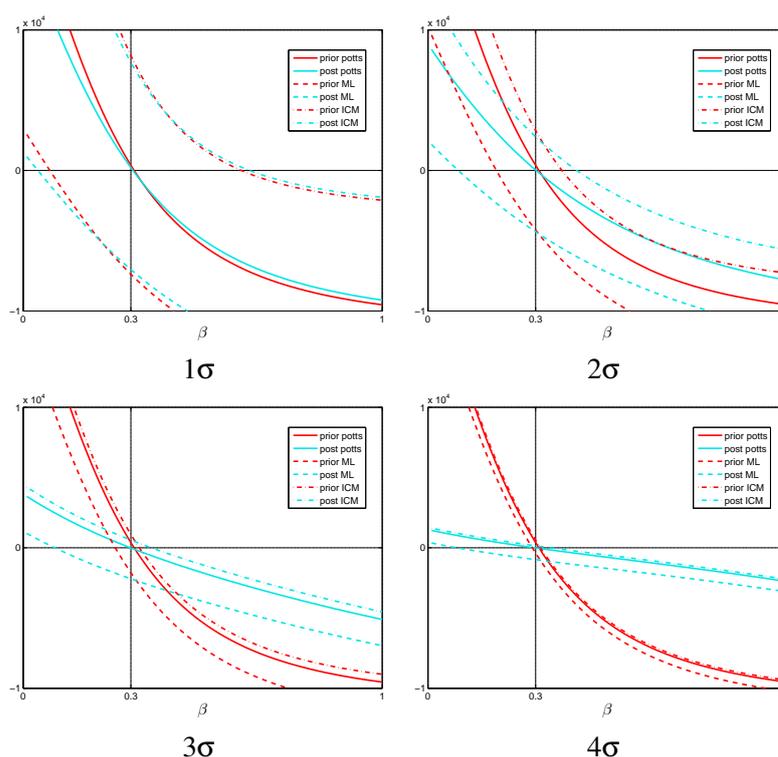


Figura 5.9: Sensibilidad de los estimadores para distintas distancias entre las medias para 2 clases y $\beta = 0,3$. Muestra cómo se modifica la pendiente de las curvas correspondientes al modelo a posteriori y como se acercan las curvas correspondientes a los modelos contaminados por ambos segmentadores y a la de los modelos puros, a medida que se distancian los modelos de emisión.

cantidad de clases casi no influye en el sesgo de los estimadores. En todos los casos analizados el estimador a priori muestra una mayor precisión.

Tal y como se hizo en las Figuras 4.5 y 5.6, al tener 100 repeticiones del experimento para cada configuración (L, β, k) , podemos realizar haces de 100 curvas que nos muestre la varianza y el sesgo de los estimadores. En la Figura 5.11 se muestran los haces obtenido para $L = 2$ y $\beta = 0,2$, al modificar la diferencia entre los modelos de emisión. Se aprecia nuevamente que los estimadores mejoran a medida que crece k . Además, se puede notar que los estimadores presentan una reducida varianza.

En base a lo mostrado en la sección, se concluye que ambos estimadores presentan una buena performance a muestra finita cuando los modelos no están contaminados. En tal caso, la varianza y el sesgo de ambos estimadores es insignificante.

De este modo entonces, para analizar la sensibilidad de los estimadores, se estudió sus performance ante la ausencia del mapa original simulado, reemplazándolo por la segmentación de los datos dada por MV y por ICM. En el caso MV ambos estimadores mostraron estimaciones sesgadas a izquierda. Al aumentar k se observaron dos situaciones complementarias: por un lado, segmentaciones por MV más parecidas a las simuladas originalmente, y por el otro lado, mayor inestabilidad en las estimaciones. Segmentaciones similares juegan un rol positivo en ambos estimadores, ya que introduce poca contaminación al modelo puro. La inestabilidad cumple un rol negativo sólo en el estimador a posteriori y no se contempla en el estimador a priori. Por ende, cuanto más distantes están los modelos de emisión de las clases, mejor es la performance del estimador a priori. Por el otro lado, el estimador a posteriori compensa los conceptos complementarios mencionados, produciendo

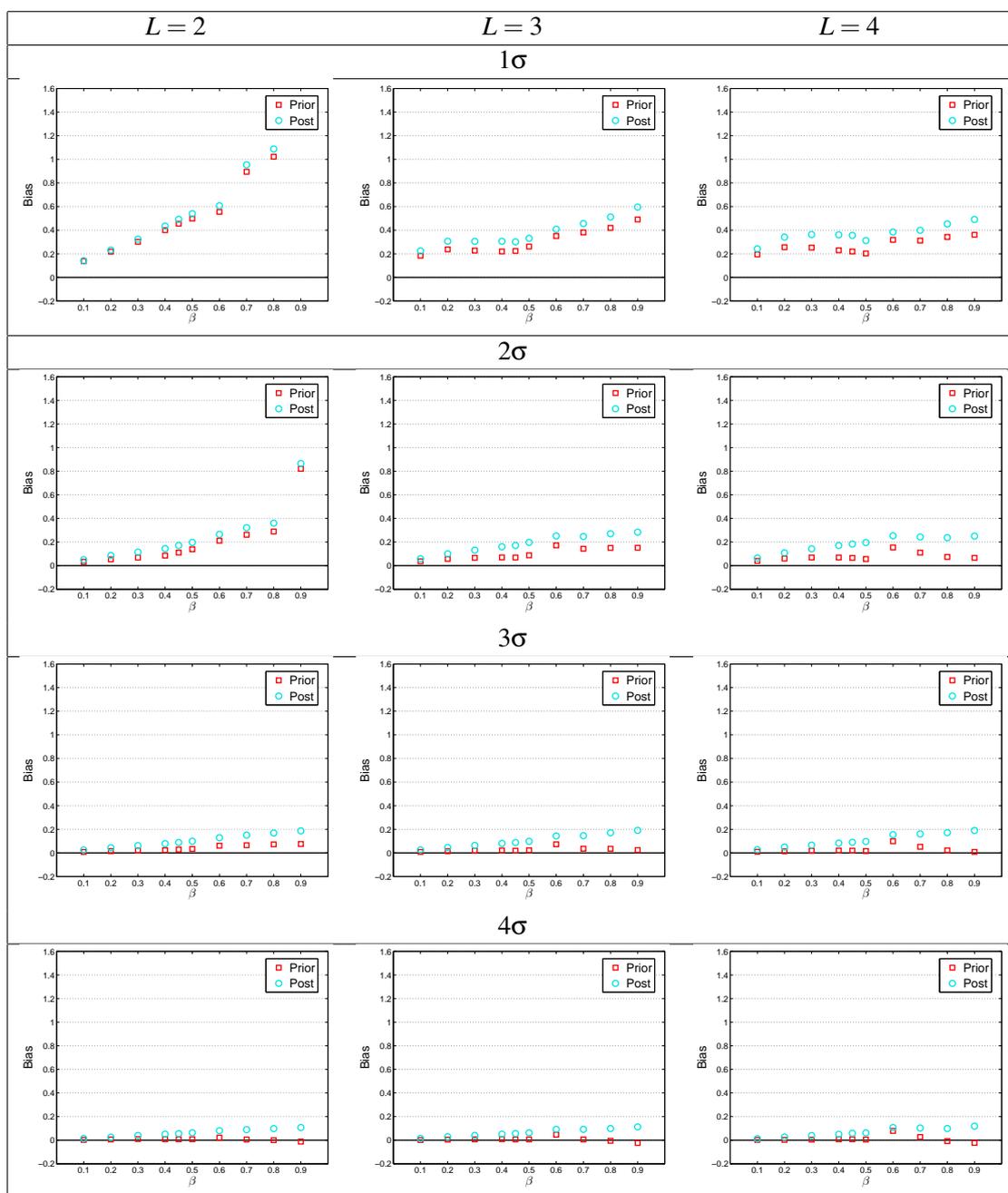


Figura 5.10: Sesgo de los estimadores al utilizar \hat{x}_{ICM} , al variar L entre 2 y 4, y k entre 1 y 4. Muestra evidencia del sesgo positivo que poseen ambos estimadores producido por estimar basándose en mapas que son modas locales del modelo obtenidos por ICM, y no en realizaciones del mismo.

estimaciones malas y del mismo orden ante distintos k . De este modo se concluye que el estimador a priori es menos sensible en este contexto. En el caso ICM, ambos estimadores funcionan mejor cuando k es grande, independientemente de la cantidad de clases. En este plano, y en todos los casos analizados, el estimador a priori presenta una mejor performance que el estimador a posteriori.

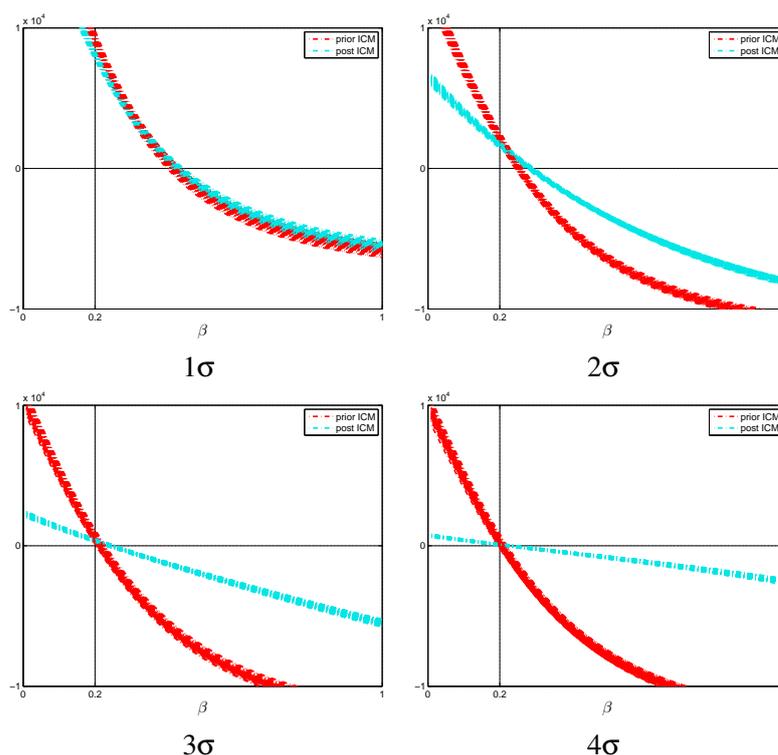


Figura 5.11: Haces de curvas de las funciones $f_{\text{prior}}(\cdot, \hat{x}_{ICM})$ y $f_{\text{post}}(\cdot, \hat{x}_{ICM})$, para el caso $L = 2$; $\beta = 0,2$; y k variando entre 1 y 4. Muestra cómo varía el sesgo positivo y cómo se mantiene la reducida varianza de ambos estimadores, a medida que se distancian los modelos de emisión.

5.5.2. Sensibilidad de las Segmentaciones

En la presente Sección se incorporan algunas reflexiones sobre la influencia del procedimiento utilizado para estimar el modelo dentro del algoritmo ICM. La elección del parámetro óptimo al aplicar ICM, es un tema discutido en la Literatura en los últimos 20 años, y para el cual se han propuesto distintas alternativas: fijar β por prueba y error [47, 48], estimar β una vez y mantenerlo fijo hasta la convergencia [9, 49, 56], o actualizarlo en cada una de sus iteraciones [15, 46, 50, 51].

Para ser gráficos en las observaciones, en la Figura 5.12 se muestra en (a) la imagen de la mandíbula de una rata que tiene 3 clases principales homogéneas: hueso-diente, cartílago y fondo. Supongamos que cada clase sigue una distribución Gaussiana con medias 91,82; 66,38; 51,62 y desvíos 16,30; 12,01; 6,25. En (b) se muestra la segmentación MV de (a), para el cual $\hat{\beta}_{\text{prior}} = 0,5034$. En (c) se muestra la segmentación obtenida por medio de ICM con β fijo e igual a 0,5034. Sin duda alguna, cualquier valor de $\beta > 0$ que se utilice en ICM homogeneizará la segmentación (a), pero a la vista de los resultados, usar el β chico y fijo estimado en base a MV, no es suficiente. Luego, si se desea estimar β en el proceso de segmentación, no podemos quedarnos con este β inicial, sino que debemos re-estimarlo iteración tras iteración, para que de este modo se retroalimenten los procesos de homogeneización de la segmentación y de crecimiento pertinente de β . En (d) se puede observar el resultado obtenido al aplicar ICM sobre (a) con este procedimiento, y a continuación se muestran los β resultantes en este proceso de re-estimación iteración tras iteración:

MV	1	2	3	4	convergencia
0,5034	0,5610	0,5673	0,5674	0,5675	0,5675.

Se observa que ante la presencia de una imagen ruidosa, estos procedimientos no poseen la fuerza homogeneizadora suficiente que requiere el grado de heterogeneidad de la segmentación inicial dada por MV que se le imputó a ICM. Como ICM converge al máximo local más cercano a esta segmentación inicial, una solución a este inconveniente es aplicarle algún pre-procesamiento de filtrado de ruido. Esto producirá una segmentación por medio de MV más homogénea, que le otorgará un punto de inicio más adecuado a ICM, y una estimación de β mayor. Estos factores generalmente producen un resultado más adecuado.

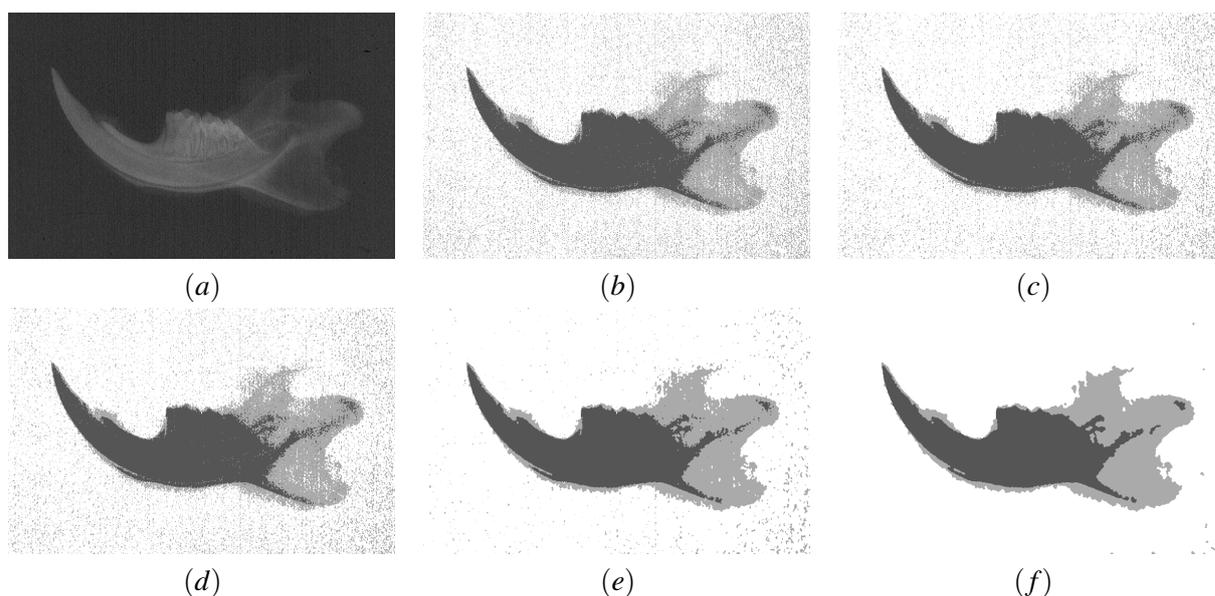


Figura 5.12: (a) Imagen Original; (b) MV de (a); (c)-(f) ICM de (a) con distintas estrategias para elegir β : (c) β estimado en base a (a) y fijado iteración tras iteración; (d)-(f) segmentación lograda cuando β es reestimado iteración tras iteración, pero los valores de inicio son diferentes.

En (f) se muestra el resultado de segmentar (a) con ICM y $\beta = 50$ fijo. Para este nuevo mapa, se verifica que $\hat{\beta}_{\text{prior}} = 4,4459$. Esto da una especie de cota superior a las estimaciones que se obtendrían en un proceso de re-estimación. Luego, para analizar el mejor resultado hipotético que se lograría con un proceso de re-estimación, en (e) mostramos la segmentación de (a) con ICM y $\beta = 4,4459$ fijo.

De este modo, no se pretende decir que siempre debemos fijar β grande al aplicar ICM, debido a que esto provoca un sobresuavizado que en determinadas circunstancias no es deseable, en el caso de que la medida de Gibbs exista.

Un ejemplo de ello puede ser si se trabaja con un conjunto de imágenes médicas, en las que el objeto a identificar sea pequeño, y un sobresuavizado tenderá a eliminarlo. Otro ejemplo es el observado en la Figura 5.3, en el que un β grande entrecorta un río. Sin embargo, cuando las imágenes son ruidosas y cuentan con un mapa de clases real desconocido pero que puede suponerse homogéneo, como en el caso de la Figura 5.12, es conveniente hacer un filtrado previo de la imagen para que el procedimiento de estimación-segmentación no colapse.

Por otro lado, si uno observa las realizaciones del modelo de Potts obtenidas en la Sección 3.9 para vecindades de segundo orden, difícilmente encuentre mapas con grandes agrupamientos de píxeles con la misma clase como los que contiene (f). Luego, el colapso de la interacción estimación-segmentación puede deberse a que los mapas de clases muy homogéneos no ajustan a un modelo de Potts con vecindades de segundo orden, y se deba

suponer vecindades de orden superior. En un reporte técnico, Morris [52] recomienda no segmentar utilizando el modelo de Potts como ley a priori, debido a que una gran cantidad de imágenes reales poseen mapas de clases que no ajustan a tales modelos. Esta conclusión es exagerada debido a que, además de que los modelos de Potts contienen una simpleza que no puede ser desaprovechada, este análisis se hizo por medio de simulaciones del modelo con vecindades de primer orden, lo cual irremediablemente produce conclusiones sesgadas sobre las bondades de segmentar en base a modelos de Potts. Sin embargo, un inconveniente que presenta el modelo de Potts isotrópico es que su estadístico suficiente $H_s(X)$ presenta una gran variabilidad en su valor medio cuando β se encuentra en el rango de valores correspondientes a imágenes que generalmente surgen en la práctica, tal y como lo afirma Morris [52]. Una vez identificada la clase de imágenes con la que se trabajará, es necesario hacer un análisis que especifique el grado de vecindades más adecuado para trabajar. En los trabajos Ji y Seymour [53], y más recientemente en un trabajo de Grelaud et al. [54], entre otros, se proponen metodologías de selección de modelo para modelos Markovianos. Otro análisis conveniente que es recomendable hacer, son los test de hipótesis que analicen la verosimilitud de la hipótesis de isotropía. Estos conceptos muestran que el juicio de la bondad de los modelos de Potts, no puede realizarse en base a simulaciones del modelo en su versión isotrópica y con vecindades de órdenes inferiores.

Capítulo 6

Apéndice

6.1. Demostraciones de resultados del Capítulo 3

Dem. Proposición 3.2.1:

Veamos (i): por la definición de las σ -álgebras \mathcal{F}_Λ , sólo se debe probar que

$$\Lambda \subseteq \Delta \Leftrightarrow \mathcal{F}_\Lambda \subseteq \mathcal{F}_\Delta.$$

Definamos para cada $\Lambda, \Delta \subseteq \mathcal{S}$ con $\Lambda \subseteq \Delta$ la función $\sigma_{\Delta \rightarrow \Lambda} : \mathcal{E}^\Delta \rightarrow \mathcal{E}^\Lambda$ dada por $\sigma_{\Delta \rightarrow \Lambda}(y) = y_\Lambda \forall y \in E^\Delta$. Luego $\sigma_\Lambda = \sigma_{\Delta \rightarrow \Lambda} \circ \sigma_\Delta$, y por lo tanto

$$A \in \mathcal{F}_\Lambda \Leftrightarrow \exists B \in \mathcal{E}^\Delta : A = \sigma_\Lambda^{-1}(B) = \sigma_\Delta^{-1}(\sigma_{\Delta \rightarrow \Lambda}^{-1}(B)).$$

Veamos la ida: si $\Lambda \subseteq \Delta$ y consideramos $A \in \mathcal{F}_\Lambda$, entonces $A = \sigma_\Lambda^{-1}(\sigma_{\Delta \rightarrow \Lambda}^{-1}(B))$, con $B \in \mathcal{E}^\Delta$. Luego como $\sigma_{\Delta \rightarrow \Lambda}^{-1}(B) \in \mathcal{E}^\Delta$, resulta que $A \in \mathcal{F}_\Delta$. Veamos ahora la vuelta: sea $s \in \Lambda$ y $B \in \mathcal{E}^{\{s\}}$ tal que $B \neq E^{\{s\}}$. Luego $\sigma_\Lambda^{-1}(\sigma_{\Lambda \rightarrow \{s\}}^{-1}(B)) \in \mathcal{F}_\Lambda \subseteq \mathcal{F}_\Delta$. Por lo que existe $C \in \mathcal{E}^\Delta$ tal que $\sigma_\Delta^{-1}(C) = \sigma_\Lambda^{-1}(\sigma_{\Lambda \rightarrow \{s\}}^{-1}(B)) = \sigma_{\{s\}}^{-1}(B)$. Luego $\sigma_{\{s\}}(\sigma_\Delta^{-1}(C)) = B$. Si se verificase que $s \notin \Delta$, entonces valdría que $\sigma_{\{s\}}(\sigma_\Delta^{-1}(C)) = E^{\{s\}}$, lo cual no es cierto, pues $B \neq E^{\{s\}}$.

Veamos (ii): para la ida consideremos $x_1, x_2 \in E^\mathcal{S}$ tales que $\sigma_\Lambda(x_1) = \sigma_\Lambda(x_2)$. Si $f(x_1) \neq f(x_2)$, entonces existen dos abiertos disjuntos G_1 y G_2 en \mathbb{R}^m , tales que $f(x_1) \in G_1$ y $f(x_2) \in G_2$. Luego por la \mathcal{F}_Λ -medibilidad de f , se tiene que $f^{-1}(G_1), f^{-1}(G_2) \in \mathcal{F}_\Lambda$ y además $f^{-1}(G_1) \cap f^{-1}(G_2) = \emptyset$. Luego existen $B_1, B_2 \in \mathcal{E}^\Lambda$ disjuntos tales que $f^{-1}(G_1) = \sigma_\Lambda^{-1}(B_1)$ y $f^{-1}(G_2) = \sigma_\Lambda^{-1}(B_2)$. Sin embargo, como $x_1 \in f^{-1}(G_1)$ y $x_2 \in f^{-1}(G_2)$, se tiene que $\sigma_\Lambda(x_1) = \sigma_\Lambda(x_2) \in B_1 \cap B_2 = \emptyset$, lo cual es un absurdo. Veamos la vuelta: fijado $w \in E^\mathcal{S}$ y dado un abierto G en \mathbb{R}^m , usando la \mathcal{F}_Λ -medibilidad de f se tiene que

$$f(x) \in G \Leftrightarrow f(x_\Lambda w^\Lambda) \in G \Leftrightarrow x_\Lambda \in \{y \in E^\Lambda : f(yw^\Lambda) \in G\} \Leftrightarrow x \in \sigma_\Lambda^{-1}(\{y \in E^\Lambda : f(yw^\Lambda) \in G\}).$$

Luego

$$f^{-1}(G) = \{x \in E^\mathcal{S} : f(x) \in G\} = \sigma_\Lambda^{-1}(\{y \in E^\Lambda : f(yw^\Lambda) \in G\}) \in \mathcal{F}_\Lambda.$$

Veamos (iii): dado un abierto G en \mathbb{R}^m , se tiene que $f^{-1}(G) \in \mathcal{F}_\Delta \subseteq \mathcal{F}_\Lambda$, lo cual finaliza la prueba.

Veamos (iv): dados $\Lambda \subseteq \Delta$, sean $x, y \in E^\mathcal{S}$ tales que $\sigma_{\Delta \setminus \Lambda}(x) = \sigma_{\Delta \setminus \Lambda}(y)$. Luego, para cada $\xi \in E^\Lambda$, se tiene que

$$f(\xi x^\Lambda) = f(\xi y_{\Delta \setminus \Lambda} x^{\Lambda \cup \Delta}) = f(\xi y^\Lambda)$$

donde la última igualdad vale por (ii) y la \mathcal{F}_Δ -medibilidad de f . Luego

$$\int_{E^\Lambda} f(\xi x^\Lambda) \lambda^\Lambda(d\xi) = \int_{E^\Lambda} f(\xi y^\Lambda) \lambda^\Lambda(d\xi),$$

lo cual, por (ii), culmina la prueba de la Proposición. \square

Dem. Proposición 3.2.2:

Sean $x, y \in E^S$ tales que $\sigma_{\Lambda^*}(x) = \sigma_{\Lambda^*}(y)$ y sea $\Delta \in \mathfrak{S}(\Lambda)$. Veamos que

$$\Phi_\Delta^\vartheta(x) = \Phi_\Delta^\vartheta(y) \quad \forall \Delta \in \mathfrak{S}(\Lambda). \quad (6.1.1)$$

Por (H.9), si $\Delta \notin \mathcal{C}$, entonces $\Phi_\Delta^\vartheta(x) = 0 = \Phi_\Delta^\vartheta(y)$; pero si $\Delta \in \mathcal{C}$, veamos que $\Delta \subseteq \Lambda^*$, ya que si esto vale, entonces por la Proposición 3.2.1(ii), se tiene también que $\Phi_\Delta^\vartheta(x) = \Phi_\Delta^\vartheta(y)$. Veamos la inclusión entonces: sea $t \in \Delta$. Si $t \in \Lambda$ entonces $t \in \Lambda^*$, pero si no, como $\Delta \in \mathfrak{S}(\Lambda)$, existe $s \in \Lambda \cap \Delta$. Luego $s \neq t$ y como $\Delta \in \mathcal{C}$, se tiene que $t \in \partial s \subseteq \Lambda^*$. Luego vale (6.1.1), y por la definición de H_Λ^ϑ y la Proposición 3.2.1(ii), finaliza la prueba de la Proposición. \square

Dem. Proposición 3.2.3:

Veamos (i): por la Proposición 3.2.1(iv) y la Proposición 3.2.2, se tiene que la función Z_Λ^ϑ es $\mathcal{F}_{\Lambda^* \setminus \Lambda}$ -medible. Como $\partial\Lambda = \Lambda^* \setminus \Lambda$, culmina la prueba de (i).

Veamos (ii): por definición de ρ_Λ^ϑ , la prueba es inmediata, debido a que la función $\exp\{-H_\Lambda^\vartheta\}$ es \mathcal{F}_{Λ^*} -medible y la función Z_Λ^ϑ es $\mathcal{F}_{\partial\Lambda}$ -medible, con $\partial\Lambda \subseteq \Lambda^*$. \square

Dem. Proposición 3.2.4:

Veamos que se cumplen (i)-(iv) de la Definición 3.2.2. Consideremos $\vartheta \in \Upsilon$ y $\Lambda \in \mathfrak{S}$.

Comencemos con (i): dado $x \in E^S$, las funciones $\gamma_\Lambda^\vartheta(\cdot|x)$ son no negativas por ser la integral de una función no negativa. Además, $\gamma_\Lambda^\vartheta(\emptyset|x) = 0$. Para probar la σ -aditividad, consideremos una familia $\{A_n\}_{n=1}^\infty \subseteq \mathcal{F}$ de conjuntos medibles disjuntos dos a dos. Luego

$$\gamma_\Lambda^\vartheta\left(\bigcup_{n=1}^\infty A_n|x\right) := \int_{E^\Lambda} 1_{\bigcup_{n=1}^\infty A_n}(\xi x^\Lambda) \rho_\Lambda^\vartheta(\xi x^\Lambda) \lambda^\Lambda(d\xi) = \sum_{n=1}^\infty \int_{E^\Lambda} 1_{A_n}(\xi x^\Lambda) \rho_\Lambda^\vartheta(\xi x^\Lambda) \lambda^\Lambda(d\xi) = \sum_{n=1}^\infty \gamma_\Lambda^\vartheta(A_n|x).$$

Veamos (ii): si $A \in \mathcal{F}$, entonces la función $x \mapsto 1_A(x) \rho_\Lambda^\vartheta(x)$ es \mathcal{F} -medible. Luego por la Proposición 3.2.1(iv), culmina la prueba de (ii).

Veamos (iii): sean $B \in \mathcal{J}_\Lambda$ y $x \in E^S$. Luego existe $C \in \mathcal{E}^{S \setminus \Lambda}$ tal que $\sigma_{S \setminus \Lambda}^{-1}(C) = B$. Por lo que

$$\gamma_\Lambda^\vartheta(B|x) := \int_{E^\Lambda} 1_B(\xi x^\Lambda) \rho_\Lambda^\vartheta(\xi x^\Lambda) \lambda^\Lambda(d\xi) = 1_C(x^\Lambda) \int_{E^\Lambda} \rho_\Lambda^\vartheta(\xi x^\Lambda) \lambda^\Lambda(d\xi) = 1_C(x^\Lambda) = 1_B(x).$$

Veamos (iv): dados $\Lambda, \Delta \in \mathfrak{S}$, con $\Lambda \subset \Delta$, segmentaremos esta prueba en la siguiente sucesión de resultados:

$$(a) \quad \xi, \eta \in E^\Lambda \text{ y } \zeta \in E^{S \setminus \Lambda} \Rightarrow \rho_\Lambda^\vartheta(\xi\zeta) \rho_\Delta^\vartheta(\eta\zeta) = \rho_\Delta^\vartheta(\xi\zeta) \rho_\Lambda^\vartheta(\eta\zeta);$$

$$(b) \xi \in E^\Lambda \text{ y } \zeta \in E^{S \setminus \Lambda} \Rightarrow \rho_\Delta^\vartheta(\xi\zeta) = \rho_\Lambda^\vartheta(\xi\zeta) \int_{E^\Lambda} \rho_\Delta^\vartheta(\eta\zeta) \lambda^\Lambda(d\eta);$$

$$(c) A \in \mathcal{F} \text{ y } x \in E^S \Rightarrow \gamma_\Delta^\vartheta(A|x) = \gamma_\Lambda^\vartheta \gamma_\Lambda^\vartheta(A|x).$$

Veamos (a):

$$H_\Delta^\vartheta(\xi\zeta) - H_\Lambda^\vartheta(\xi\zeta) = \sum_{A \in \mathfrak{G}(\Delta) \setminus \mathfrak{G}(\Lambda)} \Phi_A^\vartheta(\xi\zeta) = \sum_{A \in \mathfrak{G}(\Delta): A \subseteq \Lambda^c} \Phi_A^\vartheta(\xi\zeta) = (\diamond)$$

Como Φ_A^ϑ es \mathcal{F}_A -medible y sumamos sobre $A \subseteq \Lambda^c$, se tiene que

$$(\diamond) = \sum_{A \in \mathfrak{G}(\Delta): A \subseteq \Lambda^c} \Phi_A^\vartheta(\eta\zeta) = H_\Delta^\vartheta(\eta\zeta) - H_\Lambda^\vartheta(\eta\zeta) \quad \forall \xi, \eta \in E^\Lambda \quad \forall \zeta \in E^{S \setminus \Lambda}. \quad (6.1.2)$$

Como la función Z_Λ^ϑ es \mathcal{J}_Λ -medible, por la Proposición 3.2.1(ii), se tiene que

$$Z_\Lambda^\vartheta(\xi\zeta) = Z_\Lambda^\vartheta(\eta\zeta) \quad \forall \xi, \eta \in E^\Lambda \quad \forall \zeta \in E^{S \setminus \Lambda}. \quad (6.1.3)$$

Como la función Z_Δ^ϑ es \mathcal{J}_Δ -medible, por la Proposición 3.2.1(ii), se tiene que

$$Z_\Delta^\vartheta(\xi\zeta) = Z_\Delta^\vartheta(\eta\zeta) \quad \forall \xi, \eta \in E^\Lambda \quad \forall \zeta \in E^{S \setminus \Lambda}. \quad (6.1.4)$$

De (6.1.2), (6.1.3) y (6.1.4) se tiene que

$$\begin{aligned} \rho_\Lambda^\vartheta(\xi\zeta) \rho_\Delta^\vartheta(\eta\zeta) &= \frac{\exp\{-H_\Lambda^\vartheta(\xi\zeta) - H_\Delta^\vartheta(\eta\zeta)\}}{Z_\Lambda^\vartheta(\xi\zeta) Z_\Delta^\vartheta(\eta\zeta)} = \frac{\exp\{-H_\Delta^\vartheta(\xi\zeta) - H_\Lambda^\vartheta(\eta\zeta)\}}{Z_\Lambda^\vartheta(\eta\zeta) Z_\Delta^\vartheta(\xi\zeta)} = \\ &= \rho_\Lambda^\vartheta(\eta\zeta) \rho_\Delta^\vartheta(\xi\zeta) \quad \forall \xi, \eta \in E^\Lambda \quad \forall \zeta \in E^{S \setminus \Lambda}. \end{aligned}$$

Veamos ahora (b): por (a), se tiene que

$$\begin{aligned} \rho_\Delta^\vartheta(\xi\zeta) &= \rho_\Delta^\vartheta(\xi\zeta) \int_{E^\Lambda} \rho_\Lambda^\vartheta(\eta\zeta) \lambda^\Lambda(d\eta) = \int_{E^\Lambda} \rho_\Delta^\vartheta(\xi\zeta) \rho_\Lambda^\vartheta(\eta\zeta) \lambda^\Lambda(d\eta) = \int_{E^\Lambda} \rho_\Lambda^\vartheta(\xi\zeta) \rho_\Delta^\vartheta(\eta\zeta) \lambda^\Lambda(d\eta) = \\ &= \rho_\Lambda^\vartheta(\xi\zeta) \int_{E^\Lambda} \rho_\Delta^\vartheta(\eta\zeta) \lambda^\Lambda(d\eta) \quad \forall \xi \in E^\Lambda \quad \forall \zeta \in E^{S \setminus \Lambda}. \end{aligned}$$

Veamos ahora (c): sean $A \in \mathcal{F}$ y $x \in E^S$.

$$\gamma_\Delta^\vartheta(A|x) = \int_{E^\Delta} 1_A(\zeta x^\Delta) \rho_\Delta^\vartheta(\zeta x^\Delta) \lambda^\Delta(d\zeta) = \int_{E^{\Delta \setminus \Lambda}} \int_{E^\Lambda} 1_A(\xi \alpha x^\Delta) \rho_\Delta^\vartheta(\xi \alpha x^\Delta) \lambda^\Delta(d\xi) \lambda^{\Delta \setminus \Lambda}(d\alpha) = (\heartsuit)$$

Luego por (b),

$$\begin{aligned} (\heartsuit) &= \int_{E^{\Delta \setminus \Lambda}} \int_{E^\Lambda} 1_A(\xi \alpha x^\Delta) \rho_\Lambda^\vartheta(\xi \alpha x^\Delta) \int_{E^\Lambda} \rho_\Delta^\vartheta(\eta \alpha x^\Delta) \lambda^\Delta(d\eta) \lambda^\Delta(d\xi) \lambda^{\Delta \setminus \Lambda}(d\alpha) = \\ &= \int_{E^{\Delta \setminus \Lambda}} \int_{E^\Lambda} \int_{E^\Lambda} 1_A(\xi \alpha x^\Delta) \rho_\Lambda^\vartheta(\xi \alpha x^\Delta) \rho_\Delta^\vartheta(\eta \alpha x^\Delta) \lambda^\Delta(d\xi) \lambda^\Delta(d\eta) \lambda^{\Delta \setminus \Lambda}(d\alpha) = \\ &= \int_{E^{\Delta \setminus \Lambda}} \int_{E^\Lambda} \rho_\Delta^\vartheta(\eta \alpha x^\Delta) \int_{E^\Lambda} 1_A(\xi \alpha x^\Delta) \rho_\Lambda^\vartheta(\xi \alpha x^\Delta) \lambda^\Delta(d\xi) \lambda^\Delta(d\eta) \lambda^{\Delta \setminus \Lambda}(d\alpha) = \\ &= \int_{E^{\Delta \setminus \Lambda}} \int_{E^\Lambda} \rho_\Delta^\vartheta(\eta \alpha x^\Delta) \gamma_\Lambda^\vartheta(A|\eta \alpha x^\Delta) \lambda^\Delta(d\eta) \lambda^{\Delta \setminus \Lambda}(d\alpha) = \\ &= \int_{E^\Delta} \rho_\Delta^\vartheta(\zeta x^\Delta) \gamma_\Lambda^\vartheta(A|\zeta x^\Delta) \lambda^\Delta(d\zeta) = \int_{E^S} \gamma_\Lambda^\vartheta(A|w) \gamma_\Delta^\vartheta(dw|x) = \gamma_\Delta^\vartheta \gamma_\Lambda^\vartheta(A|x). \end{aligned}$$

□

Dem. Proposición 3.2.6:

Sea $\pi \in \mathcal{P}(E^{\mathcal{S}}, \mathcal{F})$.

Veamos (i): sean $\Lambda, \Delta \in \mathfrak{S}$ disjuntos y sean f y g funciones \mathcal{F}_Λ -medible y \mathcal{F}_Δ -medible respectivamente. Por la Proposición 3.2.1(ii), existen dos funciones $\phi_f : E^\Lambda \rightarrow \mathbb{R}$ y $\phi_g : E^\Delta \rightarrow \mathbb{R}$ tales que $f(x) = \phi_f(x_\Lambda)$ y $g(x) = \phi_g(x_\Delta) \forall x \in E^{\mathcal{S}}$. Luego

$$\begin{aligned} \pi(fg) &= \int f(x)g(x)\pi(dx) = \int_{E^{\mathcal{S} \setminus (\Lambda \cup \Delta)}} \int_{E^\Lambda} \int_{E^\Delta} f(\xi\eta\zeta)g(\xi\eta\zeta)(\pi \circ \sigma_\Delta^{-1})(d\eta)(\pi \circ \sigma_\Lambda^{-1})(d\xi)(\pi \circ \sigma_{\mathcal{S} \setminus (\Lambda \cup \Delta)}^{-1})(d\zeta) = \\ &= \int_{E^{\mathcal{S} \setminus (\Lambda \cup \Delta)}} \int_{E^\Lambda} \int_{E^\Delta} \phi_f(\xi)\phi_g(\eta)(\pi \circ \sigma_\Delta^{-1})(d\eta)(\pi \circ \sigma_\Lambda^{-1})(d\xi)(\pi \circ \sigma_{\mathcal{S} \setminus (\Lambda \cup \Delta)}^{-1})(d\zeta) = \\ &= \int_{E^\Lambda} \phi_f(\xi)(\pi \circ \sigma_\Lambda^{-1})(d\xi) \int_{E^\Delta} \phi_g(\eta)(\pi \circ \sigma_\Delta^{-1})(d\eta) = \\ &= \int_{E^\Lambda} \phi_f(\xi)(\pi \circ \sigma_\Lambda^{-1})(d\eta)(\pi \circ \sigma_\Lambda^{-1})(d\xi)(\pi \circ \sigma_{\mathcal{S} \setminus (\Lambda \cup \Delta)}^{-1})(d\zeta) \times \\ &\quad \times \int_{E^\Delta} \phi_g(\eta)(\pi \circ \sigma_\Delta^{-1})(d\eta)(\pi \circ \sigma_\Delta^{-1})(d\xi)(\pi \circ \sigma_{\mathcal{S} \setminus (\Lambda \cup \Delta)}^{-1})(d\zeta) = \\ &= \int f(x)\pi(dx) \int g(x)\pi(dx) = \pi(f)\pi(g). \end{aligned}$$

Veamos (ii): sean f y g \mathcal{F}_Λ -medibles con $\pi(\{x \in E^{\mathcal{S}} : f(x) = g(x)\}) = 1$. Supongamos que existe $x \in E^{\mathcal{S}}$ tal que $f(x) \neq g(x)$. Si $A = \sigma_\Lambda^{-1}(\{x_\Lambda\})$, entonces $A \subseteq \{x \in E^{\mathcal{S}} : f(x) \neq g(x)\}$, y por lo tanto $\pi(A) = 0$. Por la Proposición 3.2.5 y la \mathcal{F}_Λ -medibilidad de las especificaciones,

$$\gamma_\Lambda^\vartheta(A|w) = \gamma_\Lambda^\vartheta(A|x_\Lambda w^\Lambda) = \gamma_\Lambda^\vartheta(\sigma_\Lambda^{-1}(\{x_\Lambda\})|x_\Lambda w^\Lambda) = \gamma_\Lambda^\vartheta(\sigma_\Lambda^{-1}(\{(x_\Lambda w^\Lambda)_\Lambda\})|x_\Lambda w^\Lambda) = \rho_\Lambda^\vartheta(x_\Lambda w^\Lambda) > 0 \quad \forall w \in E^{\mathcal{S}}.$$

Luego $\pi(A) = \pi\gamma_\Lambda^\vartheta(A) = \int \gamma_\Lambda^\vartheta(A|w)\pi(dw) > 0$, lo cual contradice lo antes afirmado, y finaliza la prueba. \square

Dem. Proposición 3.2.7:

Sean $\vartheta \in \Upsilon$, $\Lambda \in \mathfrak{S}$ y $\Delta \subseteq \mathcal{S}$.

Veamos (i): si $A \in \mathcal{F}_\Delta$, entonces la función 1_A es \mathcal{F}_Δ -medible. Luego si probamos (iii), estaremos probando (i) también.

Veamos (ii): si $f = 1_A$, con $A \in \mathcal{F}$, entonces la igualdad vale por (3.2.2). Si f es una función simple, entonces la igualdad vale por la linealidad de la integral. Si f es una función integrable no negativa, entonces la igualdad vale por el Teorema de Convergencia Dominada. Si f es una función integrable, la igualdad vale considerando sus partes positiva y negativa por separado.

Veamos (iii): si f es una función \mathcal{F}_Δ -medible. Por la Proposición 3.2.3(ii), la función ρ_Λ^ϑ es \mathcal{F}_{Λ^*} -medible. Luego la función $x \mapsto f(x)\rho_\Lambda^\vartheta(x)$ es $\mathcal{F}_{\Delta \cup \Lambda^*}$ -medible. Luego por (ii) y la Proposición 3.2.1(iv), la función $\gamma_\Lambda^\vartheta(A|\cdot)$ es $\mathcal{F}_{(\Delta \cup \Lambda^*) \setminus \Lambda}$ -medible, lo cual concluye la prueba, pues $(\Delta \cup \Lambda^*) \setminus \Lambda = \partial\Lambda \cup (\Delta \setminus \Lambda)$.

Veamos (iv): la σ -aditividad y la no negatividad de $\pi\gamma_\Lambda^\vartheta$ valen por la σ -aditividad y la no negatividad de γ_Λ^ϑ . Por otro lado, $\pi\gamma_\Lambda^\vartheta(\emptyset) = \int \gamma_\Lambda^\vartheta(\emptyset|x)\pi(dx) = 0$. \square

Dem. Proposición 3.2.8:

Por (H.8) y el Teorema 4.23(a) de [27] se tiene que $\mathcal{G}(\gamma^\vartheta) \neq \emptyset$. Veamos ahora la convexidad: dados $\pi_1^\vartheta, \pi_2^\vartheta \in \mathcal{G}(\gamma^\vartheta)$ y $0 \leq c \leq 1$, se tiene que

$$(c\pi_1^\vartheta + (1-c)\pi_2^\vartheta)\gamma_\Lambda^\vartheta = c\pi_1^\vartheta\gamma_\Lambda^\vartheta + (1-c)\pi_2^\vartheta\gamma_\Lambda^\vartheta = c\pi_1^\vartheta + (1-c)\pi_2^\vartheta \quad \forall \Lambda \in \mathfrak{S},$$

y por ende $c\pi_1^\vartheta + (1-c)\pi_2^\vartheta \in \mathcal{G}(\gamma^\vartheta)$. □

Dem. Teorema 3.2.1:

Teniendo en cuenta (H.11) y (H.12), la parte (i) corresponde al Teorema (7.12)(a) de [27]. Veamos ahora (ii): dado $A \in \mathcal{F}$, se tiene que $1_A \in \mathcal{L}^\infty(E^S, \mathcal{F}, \pi^\vartheta, \mathbb{R})$. Luego por la parte (i), por ser $\gamma_{\Lambda_n}^\vartheta(1_A|x) = \gamma_{\Lambda_n}^\vartheta(A|x)$ y por ser $\pi^\vartheta(1_A) = \pi^\vartheta(A)$, resulta el Teorema. □

Dem. Proposición 3.3.1:

Por la Proposición 3.2.7(iii), se tiene que $\gamma_\Lambda^\vartheta(f|\cdot)$ es $\mathcal{F}_{\partial\Lambda}$ -medible. Tomemos $A \in \mathcal{F}_{\partial\Lambda}$ y veamos que

$$\int_A \gamma_\Lambda^\vartheta(f|\cdot) d\pi^\vartheta = \int_A f d\pi^\vartheta.$$

Comencemos suponiendo que $f = 1_B$, con $B \in \mathcal{F}_\Lambda$. Luego, por la Proposición 3.2.7(ii),

$$\int_A \gamma_\Lambda^\vartheta(1_B|x) \pi^\vartheta(dx) = \int_A \int_{E^\Lambda} 1_B(\xi x^\Lambda) \rho_\Lambda^\vartheta(\xi x^\Lambda) \lambda^\Lambda(d\xi) \pi^\vartheta(dx) = \int_{E^S} \int_{E^\Lambda} 1_A(x) 1_B(\xi x^\Lambda) \rho_\Lambda^\vartheta(\xi x^\Lambda) \lambda^\Lambda(d\xi) \pi^\vartheta(dx) = (*)$$

Por ser $A \in \mathcal{F}_{\partial\Lambda}$, se tiene que $1_A(x) = 1_A(\xi x^\Lambda)$. Luego $1_A(x) 1_B(\xi x^\Lambda) = 1_{A \cap B}(\xi x^\Lambda)$, y por ende

$$\begin{aligned} (*) &= \int_{E^S} \int_{E^\Lambda} 1_{A \cap B}(\xi x^\Lambda) \rho_\Lambda^\vartheta(\xi x^\Lambda) \lambda^\Lambda(d\xi) \pi^\vartheta(dx) = \int_{E^S} \gamma_\Lambda^\vartheta(A \cap B|x) \pi^\vartheta(dx) = \\ &= \pi^\vartheta \gamma_\Lambda^\vartheta(A \cap B) = \pi^\vartheta(A \cap B) = \int_A 1_B d\pi^\vartheta. \end{aligned}$$

Si f es una función simple, la igualdad vale la aditividad de la integral y de $\gamma_\Lambda^\vartheta(\cdot|x)$. Si f es integrable y no negativa, el resultado vale por el Teorema de Convergencia Dominada, y si f es integrable, entonces el resultado se prueba considerando sus partes positivas y negativas por separado. □

Dem. Proposición 3.3.2:

Por la Proposición 3.2.5 se tiene que $\pi_\Lambda^\vartheta(x_\Lambda|x^\Lambda) = \gamma_\Lambda^\vartheta(\sigma_\Lambda^{-1}(\{x_\Lambda\})|x) = \rho_\Lambda^\vartheta(x) \quad \forall x \in E^S$. □

Dem. Proposición 3.4.2:

Por la Proposición 3.4.1, (i) es equivalente a decir que la familia $\{\rho_\Lambda^\vartheta\}_{\Lambda \in \mathfrak{G}}$ es invariante por traslaciones. Teniendo en cuenta esto, tomemos $\vartheta \in \Upsilon$ y probemos la ida: dados $\Lambda \in \mathfrak{G}$ y $s \in \mathcal{S}$,

$$\rho_{\Lambda+s}^\vartheta \circ \theta_s = \rho_{\Lambda+s}^\vartheta \circ \theta_{-s}^{-1} = \rho_\Lambda^\vartheta,$$

y por (H.10), se tiene que

$$H_{\Lambda+s,0} \circ \theta_s + \vartheta^T H_{\Lambda+s} \circ \theta_s = H_{\Lambda,0} + \vartheta^T H_\Lambda.$$

Luego

$$(H_{\Lambda,0} - H_{\Lambda+s,0} \circ \theta_s) + \vartheta^T (H_\Lambda - H_{\Lambda+s} \circ \theta_s) = 0 \quad \forall \vartheta \in \Upsilon,$$

lo cual prueba la ida. Ahora veamos la vuelta: sean $x \in E^{\mathcal{S}}$, $s \in \mathcal{S}$ y $\Lambda \in \mathfrak{G}$. Por (H.13) y (H.14) se tiene que

$$h_\Lambda(x) \exp\{\vartheta^T H_\Lambda(x)\} = h_{\Lambda+s}(\theta_s(x)) \exp\{\vartheta^T H_{\Lambda+s}(\theta_s(x))\},$$

y por ende $\rho_{\Lambda+s}^\vartheta \circ \theta_s = \rho_\Lambda^\vartheta$. □

Dem. Lema 3.6.1:

Consideremos el conjunto $A = \{x \in E^{\mathcal{S}} : f_n(x) \rightarrow f(x)\}$ y probemos que (ii) vale sí, y sólo si, $P(A^c) = 0$. Como

$$A^c = \bigcup_{\varepsilon > 0} \bigcap_{m=1}^{\infty} \bigcup_{n=m}^{\infty} B_{n,\varepsilon}^c,$$

se tiene que

$$P(A^c) = 0 \Leftrightarrow P\left(\bigcup_{\varepsilon > 0} \bigcap_{m=1}^{\infty} \bigcup_{n=m}^{\infty} B_{n,\varepsilon}^c\right) = 0 \Leftrightarrow P\left(\bigcap_{m=1}^{\infty} \bigcup_{n=m}^{\infty} B_{n,\varepsilon}^c\right) = 0 \quad \forall \varepsilon > 0 \Leftrightarrow \lim_{m \rightarrow \infty} \pi\left(\bigcup_{n=m}^{\infty} B_{n,\varepsilon}^c\right) = 0 \quad \forall \varepsilon > 0$$

donde la última equivalencia vale por ser decreciente la sucesión $C_m := \bigcup_{n=m}^{\infty} B_{n,\varepsilon}^c$. □

Dem. Lema 3.6.2:

Sea $\varepsilon > 0$. Luego existe $i_0 \in \mathbb{N}$ tal que

$$\sum_{i=i_0}^{\infty} \pi\left(\bigcup_{n=r_i}^{r_{i+1}-1} B_{n,\varepsilon}^c\right) < \varepsilon,$$

pero entonces

$$\pi\left(\bigcup_{n=r_{i_0}}^{\infty} B_{n,\varepsilon}^c\right) = \pi\left(\bigcup_{i=i_0}^{\infty} \bigcup_{n=r_i}^{r_{i+1}-1} B_{n,\varepsilon}^c\right) \leq \sum_{i=i_0}^{\infty} \pi\left(\bigcup_{n=r_i}^{r_{i+1}-1} B_{n,\varepsilon}^c\right) < \varepsilon,$$

lo cual implica que vale (ii) del Lema 3.6.1, y por lo tanto finaliza la prueba. □

Dem. Lema 3.6.3:

Sea $A := \{x \in E^{\mathcal{S}} : \max_{1 \leq i \leq n} |S_i(x)| \geq \varepsilon\}$. Luego $A = \bigcup_{i=1}^n A_i$, donde los conjuntos A_i son disjuntos y vienen dados por

$$A_i := \{x \in E^{\mathcal{S}} : |S_1(x)| < \varepsilon, \dots, |S_{i-1}(x)| < \varepsilon, |S_i(x)| \geq \varepsilon\}.$$

Luego $1_A = \sum_{i=1}^n 1_{A_i}$ y $S_n^2 \geq S_n^2 1_A = \sum_{i=1}^n S_n^2 1_{A_i}$. Por lo que

$$\pi(S_n^2) \geq \sum_{i=1}^n \pi(S_n^2 1_{A_i}). \quad (6.1.5)$$

Para cada $1 \leq i < n$, sea $T_i := \sum_{j=i+1}^n f_j$. Luego

$$1 \leq i < n \Rightarrow S_n^2 I_{A_i} = (S_i + T_i)^2 I_{A_i} = S_i^2 I_{A_i} + T_i^2 I_{A_i} + 2S_i I_{A_i} T_i. \quad (6.1.6)$$

Como $2S_i I_{A_i}$ es una función \mathcal{F}_Λ -medible y la función T_i es \mathcal{F}_Δ -medible, con $\Lambda := \bigcup_{j=1}^i \Delta_j$ y $\Delta := \bigcup_{j=i+1}^n \Delta_j$ disjuntos, utilizando la Proposición 3.2.6(i), se tiene que

$$1 \leq i < n \Rightarrow \pi(2S_i I_{A_i} T_i) = \pi(2S_i I_{A_i}) \pi(T_i) = \pi(2S_i I_{A_i}) \sum_{j=i+1}^n \pi(f_j) = 0.$$

Luego por (6.1.6)

$$1 \leq i < n \Rightarrow \pi(S_n^2 I_{A_i}) = \pi(S_i^2 I_{A_i}) + \pi(T_i^2 I_{A_i}).$$

Luego por (6.1.5)

$$\pi(S_n^2) \geq \sum_{i=1}^n \pi(S_n^2 1_{A_i}) = \sum_{i=1}^n (\pi(S_i^2 I_{A_i}) + \pi(T_i^2 I_{A_i})) \geq \sum_{i=1}^n \pi(S_i^2 I_{A_i}).$$

Por la Proposición 3.2.6(i), se tiene que $\pi(f_i f_j) = \pi(f_i) \pi(f_j)$ para $i \neq j$, y por ende

$$\pi(S_n^2) \geq \sum_{i=1}^n \sum_{j=1}^i \pi(f_j^2 I_{A_i}) \geq \sum_{i=1}^n \pi(f_i^2 I_{A_i}) \geq \varepsilon^2 \sum_{i=1}^n \pi(A_i) = \varepsilon^2 \pi(A).$$

Esto implica que, utilizando nuevamente la Proposición 3.2.6(i),

$$\pi\left(\left\{x \in E^{\mathcal{S}} : \max_{1 \leq i \leq n} |S_i(x)| \geq \varepsilon\right\}\right) = \pi(A) \leq \frac{\pi(S_n^2)}{\varepsilon^2} = \frac{1}{\varepsilon^2} \sum_{i=1}^n \pi(f_i^2).$$

□

6.2. Especificaciones con simetrías.

En la presente sección desarrollaremos un tema que generaliza lo visto en la Sección 3.4, y que es imprescindible para probar la condición (H.17) del estimador de PMV basado en el modelo a posteriori. Por lo que, se desarrollará la teoría necesaria y suficiente para tal fin. Para más detalles del tema ver el Capítulo 5 de [27].

Consideremos un conjunto finito E y el conjunto de sitios $\mathcal{S} = \mathbb{Z}^d$, $d \in \mathbb{N}$. Llamemos \mathcal{T} al conjunto de todas las simetrías sobre $E^{\mathcal{S}}$, esto es, de todas las funciones biyectivas $\tau : E^{\mathcal{S}} \rightarrow E^{\mathcal{S}}$ para la cual existen dos funciones biyectivas $\tau_1 : \mathcal{S} \rightarrow \mathcal{S}$ y $\tau_2 : E \rightarrow E$ tales que

$$\tau = \tau_2 \circ x \circ \tau_1,$$

recordando siempre que si $x \in E^S$, entonces x es una función de S en E . Un ejemplo de simetrías son las traslaciones θ_s , trabajadas en la Sección 3.4, donde $\tau_1(t) = t - s$ y τ_2 es la identidad en E . Otro ejemplo de simetría es la función B definida en la prueba del Lema 6.4.1, para la cual τ_1 es la identidad en S y τ_2 viene dada por la biyección b sobre E que solamente intercambia dos etiquetas.

Definición 6.2.1 Dada una simetría $\tau \in \mathcal{T}$ y una especificación $\gamma = \{\gamma_\Lambda\}_{\Lambda \in \mathfrak{G}}$, se define una nueva especificación $\tau(\gamma)$ dada por

$$\tau(\gamma)_\Lambda(A|x) = \gamma_{\tau_1^{-1}(\Lambda)}(\tau^{-1}(A)|\tau^{-1}(x)) \quad \forall x \in E^S \quad \forall A \in \mathcal{F} \quad \forall \Lambda \in \mathfrak{G}.$$

Además, dada una probabilidad $\pi \in \mathcal{P}(E^S, \mathcal{F})$, se define la probabilidad $\tau(\pi) \in \mathcal{P}(E^S, \mathcal{F})$ dada por

$$\tau(\pi)(A) = \pi(\tau^{-1}(A)) \quad \forall A \in \mathcal{F}.$$

Por último, dada una familia de funciones $f = \{f_\Lambda\}_{\Lambda \in \mathfrak{G}}$ de E^S en \mathbb{R} , se nota con $\tau(f) = \{\tau(f)_\Lambda\}_{\Lambda \in \mathfrak{G}}$ a la familia de funciones dadas por

$$\tau(f)_\Lambda(x) = f_{\tau_1^{-1}(\Lambda)}(\tau^{-1}(x)) \quad \forall x \in E^S \quad \forall \Lambda \in \mathfrak{G}.$$

Definición 6.2.2 Se dice que $\tau \in \mathcal{T}$ es una simetría de una especificación γ si

$$\tau(\gamma) = \gamma.$$

Además, se dice $\pi \in \mathcal{P}(E^S, \mathcal{F})$ preserva una simetría $\tau \in \mathcal{T}$ si

$$\tau(\pi) = \pi.$$

De igual modo, se dice que τ preserva a λ si $\lambda \circ \tau_2 = \lambda$. Se dice que una familia de funciones $f = \{f_\Lambda\}_{\Lambda \in \mathfrak{G}}$ es τ -invariante si

$$\tau(f) = f.$$

Teorema 6.2.1 Si $\mathcal{G}(\gamma) = \{\pi\}$. Entonces π preserva todas las simetrías de γ .

Dem. Corresponde al Corolario 5.11 de [27]. □

Como se trabaja con λ la medida que cuenta sobre E finito, cualquier $\tau \in \mathcal{T}$ preserva a λ . Si consideramos la familia de potenciales $\{\Phi^\vartheta\}_{\vartheta \in \Upsilon}$ considerados en la Sección 3.2 y sus respectivas especificaciones $\{\gamma^\vartheta\}_{\vartheta \in \Upsilon}$. Entonces se verifica el siguiente resultado

Teorema 6.2.2 Para cada $\vartheta \in \Upsilon$ y $\tau \in \mathcal{T}$ que preserva a λ , si la familia de funciones $\Phi^\vartheta = \{\Phi_\Lambda^\vartheta\}_{\Lambda \in \mathfrak{G}}$ es τ -invariante. Entonces τ es una simetría de la especificación γ^ϑ .

Dem. Corresponde al Corolario 5.9 de [27]. □

6.3. Consistencia de $\widehat{\beta}_{post}$

En la Sección 4.3 se probó que el modelo a priori satisface (H.1)-(H.15), y con ello, por el Teorema 3.8.1, resultó la consistencia del estimador a priori. Posteriormente se enunció el Teorema de consistencia del estimador a posteriori, el cual probaremos en la presente Sección. Supondremos válido entonces (H.11'), esto es, $\mathcal{G}(\tilde{\gamma}^\beta) = \{\pi^\beta\} \forall \beta \in \Upsilon$, siendo $\{\tilde{\gamma}^\beta\}_{\beta \in \Upsilon}$ las especificaciones del modelo a priori. Bajo este supuesto, en la Sección 6.4 del Apéndice se probará que vale (H.17).

Recordando definiciones, la función objetivo que maximiza el estimador a priori $\widehat{\beta}_{n,prior}$ basado en una muestra sobre Λ_n viene dada por

$$U_n^{prior}(\beta, x) = \frac{1}{\#\Lambda_n^\circ} \sum_{s \in \Lambda_n^\circ} \left(\beta H_s(x_s) - \ln \tilde{Z}_s^\beta(x) \right),$$

mientras que la función objetivo que maximiza el estimador a posteriori $\widehat{\beta}_{n,post}$ basado en una muestra sobre Λ_n viene dada por

$$U_n^{post}(\beta, x) = \frac{1}{\#\Lambda_n^\circ} \sum_{s \in \Lambda_n^\circ} \left(\ln f_{x_s}(I_s) + \beta H_s(x_s) - \ln Z_s^\beta(x) \right).$$

Luego

$$\widehat{\beta}_{n,prior}(x) \in \tilde{M}_{\Lambda_n}(x) := \left\{ \vartheta \in \Upsilon : U_n^{prior}(\beta, x) = \min_{\alpha \in \Upsilon} U_n^{prior}(\alpha, x) \right\},$$

y

$$\widehat{\beta}_{n,post}(x) \in M_{\Lambda_n}(x) := \left\{ \vartheta \in \Upsilon : U_n^{post}(\beta, x) = \min_{\alpha \in \Upsilon} U_n^{post}(\alpha, x) \right\}.$$

En el contexto planteado, para cada $s \in \mathcal{S}$ se tiene que

$$\begin{aligned} h_s(x) &= f_{x_s}(I_s), \quad H_s(x) = \#\{t \in \partial s : x_s = x_t\}, \\ \tilde{\gamma}_s^\beta(H_s|x) &= \frac{\sum_{\ell \in E} H_s(\ell) \exp\{\beta H_s(\ell)\}}{\sum_{\ell \in E} \exp\{\beta H_s(\ell)\}}, \quad \gamma_s^\beta(H_s|x) = \frac{\sum_{\ell \in E} H_s(\ell) f_\ell(I_s) \exp\{\beta H_s(\ell)\}}{\sum_{\ell \in E} f_\ell(I_s) \exp\{\beta H_s(\ell)\}}, \\ \tilde{\rho}_s(x) &= \frac{\exp\{\beta H_s(x_s)\}}{\sum_{\ell \in E} \exp\{\beta H_s(\ell)\}} \quad \text{y} \quad \rho_s(x) = \frac{f_{x_s}(I_s) \exp\{\beta H_s(x_s)\}}{\sum_{\ell \in E} f_\ell(I_s) \exp\{\beta H_s(\ell)\}}. \end{aligned}$$

Luego

Lema 6.3.1 Para cada $\beta \in \Upsilon$, $\nabla U_n^{prior}(\beta, \cdot) - \nabla U_n^{post}(\beta, \cdot) \rightarrow 0$ π^β -p.p.

Dem. Por el Lema 3.8.14 se tiene que

$$\nabla U_n^{prior}(\beta, \cdot) - \nabla U_n^{post}(\beta, \cdot) = \frac{1}{\#\Lambda_n^\circ} \sum_{s \in \Lambda_n^\circ} \left(\tilde{\gamma}_s^\beta(H_s|\cdot) - \frac{\tilde{\gamma}_s^\beta(h_s H_s|\cdot)}{\tilde{\gamma}_s^\beta(h_s|\cdot)} \right) = \frac{1}{\#\Lambda_n^\circ} \sum_{s \in \Lambda_n^\circ} \varepsilon_s,$$

donde para cada $s \in \mathcal{S}$, $\varepsilon_s := \tilde{\gamma}_s^\beta(H_s|\cdot) - \tilde{\gamma}_s^\beta(\tilde{h}_s H_s|\cdot)$, con $\tilde{h}_s := \frac{h_s}{\tilde{\gamma}_s^\beta(h_s|\cdot)}$. Por lo tanto sólo se debe probar que

$$\frac{1}{\#\Lambda_n^\circ} \sum_{s \in \Lambda_n^\circ} \varepsilon_s \rightarrow 0 \quad \pi^\beta\text{-p.p.} \quad (6.3.7)$$

Como $\tilde{\gamma}_s^\beta(\tilde{h}_s|\cdot) \equiv 1 \ \forall s \in \mathcal{S}$, se tiene que $\pi^\beta(\tilde{h}_s) = 1 \ \forall s \in \mathcal{S}$. Luego por ser $\pi^\beta \in \mathcal{G}(\tilde{\gamma}^\beta)$ y luego por (H.17),

$$\pi^\beta(\varepsilon_s) = \pi^\beta(\tilde{\gamma}_s^\beta(\tilde{h}_s H_s|\cdot) - \tilde{\gamma}_s^\beta(H_s|\cdot)) = \pi^\beta(\tilde{h}_s H_s - H_s) = \pi^\beta(\tilde{h}_s)\pi^\beta(H_s) - \pi^\beta(H_s) = 0 \quad \forall s \in \mathcal{S}.$$

Además, para cada $s \in \mathcal{S}$, la función ε_s es \mathcal{F}_{∂_s} -medible y acotada. Luego por el Corolario 3.6.1, vale (6.3.7), y finaliza la prueba. \square

Dem. Teorema 4.3.1

Veamos (i): por el Lema 3.8.5, se tiene que

$$\liminf_{n \rightarrow \infty} \frac{N_n(\zeta)}{\#\Lambda_n^\circ} > 0 \quad \pi^{\beta_0}\text{-p.p.} \quad \forall \zeta \in E^{\partial_s},$$

Sea $x \in E^{\mathcal{S}}$ en este conjunto. Luego por el Lema 3.8.14(iv), $U_n^{\text{post}}(\cdot, x)$ es cóncava por verificarse

$$\nabla^2 U_n^{\text{post}}(\beta_0, x) = -\frac{1}{\#\Lambda_n^\circ} \sum_{s \in \Lambda_n^\circ} \gamma_s^{\beta_0} \left(H_s (H_s - \gamma_s^{\beta_0}(H_s|x))^T |x \right) \leq 0.$$

Sea $\beta \neq \beta_0$ y supongamos que $\nabla^2 U_n^{\text{post}}(\beta, x) = 0$. Entonces,

$$H_s(\xi x^s) = \gamma_s^\beta(H_s|x) \quad \forall \xi \in E \ \forall s \in \mathcal{S}.$$

En particular, $H_s(\cdot x^s)$ es constante $\forall s \in \mathcal{S}$, y esto implica que

$$\tilde{\rho}_s^\beta(\xi x^s) = \frac{\exp\{\beta H_s(\xi x^s)\}}{\sum_{\tilde{\xi} \in E} \exp\{\beta H_s(\tilde{\xi} x^s)\}} = \frac{1}{\#E} \quad \forall \xi \in E \ \forall s \in \mathcal{S}, \quad (6.3.8)$$

expresión que no depende de β . Luego, (6.3.8) contradice (H.15) por un argumento análogo al utilizado en la prueba del Lema 3.8.8. Concluyendo que $U_n^{\text{post}}(\cdot, x)$ es estrictamente cóncava $\forall n \gg 1$, y en consecuencia valen (i) y (ii).

Veamos ahora (iii): consideremos $x \in E^{\mathcal{S}}$ en el conjunto de probabilidad 1 al cual pertenecen los elementos que verifican las condiciones dadas en (i), en el Teorema 3.8.1, y en los Lemas 3.8.12 y 6.3.1. Sea B una bola cerrada centrada en β_0 . Luego por la consistencia de $\hat{\beta}_{n, \text{prior}}$ se tiene que

$$\|\hat{\beta}_{n, \text{prior}}(x) - \beta_0\| \rightarrow 0, \quad (6.3.9)$$

y por ende $\hat{\beta}_{n, \text{prior}}(x) \in B \ \forall n \gg 1$.

Veamos que

$$\nabla U_n^{\text{prior}}(\beta_0, x) \rightarrow 0. \quad (6.3.10)$$

Supongamos que no, luego existe $\varepsilon_0 > 0$, tal que para cada $n \in \mathbb{N}$, existe $k_n \geq n$, tal que $|\nabla U_{k_n}^{\text{prior}}(\beta_0, x)| \geq \varepsilon_0$.

Luego como $|\nabla U_{k_n}^{\text{prior}}(\hat{\beta}_{k_n, \text{prior}}(x), x)| = 0$, se tiene que

$$|\nabla U_{k_n}^{\text{prior}}(\hat{\beta}_{k_n, \text{prior}}(x), x) - \nabla U_{k_n}^{\text{prior}}(\beta_0, x)| \geq \varepsilon_0. \quad (6.3.11)$$

Por otro lado, como

$$\begin{aligned} \nabla^2 U_n^{\text{prior}}(\beta, x) &= -\frac{1}{\#\Lambda_n^\circ} \sum_{s \in \Lambda_n^\circ} \gamma_s^\beta \left(H_s(H_s - \gamma_s^\beta(H_s|x))^T |x \right) \leq \\ &\leq K := \max_{\beta \in B, s \in \mathcal{S}} \gamma_s^\beta \left(H_s(H_s - \gamma_s^\beta(H_s|x))^T |x \right) \quad \forall n \gg 1 \quad \forall \beta \in B, \end{aligned}$$

cota que existe por (H.1) y por ser B compacto. Luego

$$|\nabla U_n^{\text{prior}}(\widehat{\beta}_{n,\text{prior}}(x), x) - \nabla U_n^{\text{prior}}(\beta_0, x)| \leq K \|\widehat{\beta}_{n,\text{prior}}(x) - \beta_0\| \quad \forall n \gg 1.$$

Por lo que, usando (6.3.11), se tiene que

$$\frac{\varepsilon_0}{K} \leq \|\widehat{\beta}_{k_n,\text{prior}}(x) - \beta_0\| \quad \forall n \gg 1,$$

lo cual contradice (6.3.9), y por lo tanto vale (6.3.12). Implicando que, por el Lema 6.3.1

$$|\nabla U_n^{\text{post}}(\widehat{\beta}_{k_n,\text{post}}(x), x) - \nabla U_n^{\text{post}}(\beta_0, x)| = |\nabla U_n^{\text{post}}(\beta_0, x)| \rightarrow 0. \quad (6.3.12)$$

Como sólo debemos probar que

$$\|\widehat{\beta}_{n,\text{post}}(x) - \beta_0\| \rightarrow 0, \quad (6.3.13)$$

supongamos que esto no vale. Luego existe $\varepsilon > 0$ tal que para cada $n \in \mathbb{N}$, existe $k_n \geq n$ tal que $\|\widehat{\beta}_{k_n}(x) - \beta_0\| \geq \varepsilon$. Teniendo en cuenta (H.6), considerando

$$\bar{\beta}_{n,\zeta} := \beta_0 + \zeta \frac{\widehat{\beta}_{n,\text{post}}(x) - \beta_0}{\|\widehat{\beta}_{n,\text{post}}(x) - \beta_0\|} \in \Upsilon,$$

para $\zeta \in (0, \varepsilon)$ fijado arbitrariamente. Luego por (i), se tiene que $\nabla U_n^{\text{post}}(\cdot, x)$ es decreciente, y por lo tanto

$$0 = |\nabla U_n^{\text{post}}(\widehat{\beta}_{n,\text{post}}(x), x)| \leq |\nabla U_n^{\text{post}}(\bar{\beta}_{n,\zeta}, x)| \leq |\nabla U_n^{\text{post}}(\beta_0, x)| \rightarrow 0 \quad \forall \zeta \in (0, \varepsilon),$$

lo cual implica que $U_n^{\text{post}}(\cdot, x)$ tiende a ser una cte, lo cual es un absurdo por (H.10), que provino de suponer que no vale (6.3.13).

6.4. Independencia entre h_s y H_s en el modelo de Potts

Un resultado clave en la prueba de la Consistencia del estimador $\widehat{\beta}_{\text{post}}$ es la independencia de las variables aleatorias $h_s(X) = f_{X_s}(I_s)$ y $H_s(X) = \#\{t \in \partial s : X_s = X_t\}$ para cada $s \in \mathcal{S}$. Esta independencia se refiere a la independencia entre variables aleatorias definidas sobre el espacio de probabilidad (Ω, \mathcal{A}, P) definido en la Sección 3.5. Para ello alcanza con probar la independencia entre las variables X_s y $H_s(X)$ para cada $s \in \mathcal{S}$. Supondremos que la distribución π^β de X es la única medida de Gibbs asociada a la especificación del modelo a priori $\tilde{\gamma}^\beta$, esto es

$$(H.11') \quad \mathcal{G}(\tilde{\gamma}^\beta) = \{\pi^\beta\},$$

en lugar de suponer (H.11). Esta suposición implica la suposición (H.11), por lo que todos los resultados probados hasta el momento continúan valiendo. Por el Teorema 6.2.1 de la Sección 6.2, (H.11') implica que π^β preserva cualquier simetría de $\tilde{\gamma}^\beta$. En caso de no querer hacer esta suposición de unicidad, se debería probar alguna de las condiciones del Teorema 5.15 de [27], para luego agregar una suposición de que π^β es una de las probabilidades de Gibbs extremales que preserva determinadas simetrías, pero escapa los objetivos de la tesis.

Lema 6.4.1 *Si X tiene distribución π^β , probabilidad cuya existencia y unicidad asegura (H.11'). Entonces para cada $s \in \mathcal{S}$, las funciones $h_s(X) = f_{X_s}(I_s)$ y $H_s(X) = \#\{t \in \partial s : X_s = X_t\}$ son independientes como variables aleatorias definidas sobre el espacio de Probabilidad (Ω, \mathcal{A}, P) definido en la Sección 3.5.*

Dem. Dado $s \in \mathcal{S}$, alcanza con ver que para cada $k \in \{0, 1, \dots, \#\partial s\}$, la función

$$\ell \mapsto \pi^\beta(\{x \in E^{\mathcal{S}} : x_s = \ell, H_s(x) = k\})$$

es constante, ya que si esto sucede, entonces la función

$$\ell \mapsto P(X_s = \ell | H_s(X) = k) = \frac{\pi^\beta(\{x \in E^{\mathcal{S}} : x_s = \ell, H_s(x) = k\})}{\pi^\beta(\{x \in E^{\mathcal{S}} : H_s(x) = k\})}$$

también sería constante, y esto implicaría que conocer el valor de $H_s(X)$ no modifica la probabilidad de que X_s tome alguna etiqueta en particular.

Sean entonces $k \in \{0, 1, \dots, \#\partial s\}$ un valor en el rango de la variable $H_s(X)$ y $\ell_1, \ell_2 \in E$ en el rango de X_s . Luego consideremos la función biyectiva $b : E \rightarrow E$ dada por

$$b(\ell) := \begin{cases} \ell_2 & \text{si } \ell = \ell_1 \\ \ell_1 & \text{si } \ell = \ell_2 \\ \ell & \text{caso contrario.} \end{cases}$$

Sea ahora $B : E^{\mathcal{S}} \rightarrow E^{\mathcal{S}}$ la función biyectiva dada por $b(x) := b \circ x$. Por la Sección 6.2 del Apéndice, la función B es una simetría.

Luego $(B^{-1}(x))_t = (B(x))_t = (b \circ x)_t = b^{-1}(x_t) = b(x_t)$, y por lo tanto

$$\begin{aligned} \Phi_\Lambda^\beta(B^{-1}(x)) &= \begin{cases} -\beta & \text{si } \Lambda = \{s, t\}, t \in \partial s, (B^{-1}(x))_s = (B^{-1}(x))_t \\ 0 & \text{caso contrario} \end{cases} \\ &= \begin{cases} -\beta & \text{si } \Lambda = \{s, t\}, t \in \partial s, b(x_s) = b(x_t) \\ 0 & \text{caso contrario} \end{cases} \\ &= \begin{cases} -\beta & \text{si } \Lambda = \{s, t\}, t \in \partial s, x_s = x_t \\ 0 & \text{caso contrario} \end{cases} = \Phi_\Lambda^\beta(x). \end{aligned}$$

Luego por el Teorema 6.2.2, B es una simetría para $\tilde{\gamma}^\beta$. Luego por el Teorema 6.2.1, π^β preserva la simetría B .

Por otro lado,

$$\begin{aligned} \{x \in E^{\mathcal{S}} : x_s = \ell_1, H_s(x) = k\} &= \{x \in E^{\mathcal{S}} : \#\{t \in \partial s : x_t = \ell_1\} = k\} = \{x \in E^{\mathcal{S}} : \#\{t \in \partial s : (B(x))_t = \ell_2\} = k\} = \\ &= B^{-1}(\{x \in E^{\mathcal{S}} : \#\{t \in \partial s : x_t = \ell_2\} = k\}) = B^{-1}(\{x \in E^{\mathcal{S}} : x_s = \ell_2, H_s(x) = k\}) \end{aligned}$$

Luego

$$\pi^\beta(\{x \in E^{\mathcal{S}} : x_s = \ell_1, H_s(x) = k\}) = \pi^\beta(B^{-1}(\{x \in E^{\mathcal{S}} : x_s = \ell_2, H_s(x) = k\}))$$

$$= \pi^\beta (\{x \in E^S : x_s = \ell_2, H_s(x) = k\}),$$

lo cual prueba que la función

$$\ell \mapsto \pi^\beta (\{x \in E^S : x_s = \ell, H_s(x) = k\})$$

es constante para cada k , finalizando así la prueba. □

Lema 6.4.2 *Dado el espacio de probabilidad (Ω, \mathcal{A}, P) ; un campo aleatorio $X : \Omega \rightarrow E^S$ con distribución π^β ; y una función $f \in \mathcal{L}^1(E^S, \mathcal{E}^S, \pi^\beta, \mathbb{R})$. Entonces $P(f \circ X) = \pi^\beta(f)$.*

Dem. Supongamos primero que $f = 1_A$ con $A \in \mathcal{F}$. Luego

$$\pi^\beta(1_A) = \pi^\beta(A) = P(X \in A) = P(1_A(X) = 1) = 1P(1_A(X) = 1) + 0P(1_A(X) = 0) = P(1_A \circ X)$$

Si f es una función simple no negativa, el resultado vale por la linealidad de la integral, y si f es una función integrable no negativa, el resultado vale por el Teorema de Convergencia Dominada. Por último si f es simplemente una función integrable, el resultado vale considerando las partes positiva y negativa por separado. □

De estos dos últimos Lemas, resulta el siguiente resultado, que será identificado como la no correlación entre h_s y H_s .

Lema 6.4.3 *Si vale (H.11'), entonces $\pi^\beta(h_s H_s) = \pi^\beta(h_s) \pi^\beta(H_s) \quad \forall s \in \mathcal{S}$.*

Dem. Sea $s \in \mathcal{S}$. Como independencia implica no correlación, por el Lema 6.4.1 se tiene que

$$P((h_s H_s) \circ X) = P(h_s(X) H_s(X)) = P(h_s(X)) P(H_s(X)).$$

Luego por el Lema 6.4.2, resulta

$$\pi^\beta(h_s H_s) = P((h_s H_s) \circ X) = P(h_s(X)) P(H_s(X)) = \pi^\beta(h_s) \pi^\beta(H_s).$$

□

Conclusión

En la tesis se estudiaron los procesos de segmentación de imágenes digitales bajo modelos Markovianos y de estimación de los parámetros del mismo. Se mostró la fuerte interrelación que existe entre los mismos, y la sensibilidad de cada uno de ellos.

En cuanto al proceso de estimación, cada modelo Markoviano posee su estimador de Pseudo-Máxima Verosimilitud, y la consistencia del mismo depende de las características de las especificaciones o leyes condicionales que lo definen.

En la tesis se probó la consistencia fuerte del estimador de Pseudo-Máxima Verosimilitud correspondiente a una familia exponencial de distribuciones de Gibbs, generadas por especificaciones no invariantes por traslaciones equivalentes microscópicamente a un modelo invariante. Para tal prueba, se mostró una nueva versión de la ley fuerte de los grandes números que funciona independientemente de suposiciones de ergodicidad y de esperanzas condicionales respecto a la σ -álgebra cola.

En un contexto Bayesiano, se mostró que bajo determinados supuestos clásicos el modelo a posteriori para el mapa de clases es un modelo generado por especificaciones no invariantes. En la tesis se definió y estudió un nuevo estimador del parámetro de suavidad del modelo de Potts isotrópico, el cual es el estimador de Pseudo-Máxima Verosimilitud correspondiente al modelo a posteriori del mapa de clases de una imagen digital. Se probó su consistencia y se mostró bajo simulación su buena performance a muestra finita. Por otro lado, se mostró evidencia bajo simulación de que la utilización de la verosimilitud de la información radiométrica hace más sensibles las estimaciones.

En cuanto al proceso de segmentación, se trabajó principalmente con el clasificador contextual ICM, se mostró la influencia de la metodología de selección de modelo en el mismo, y se observó la necesidad de ajustar el modelo iteración tras iteración del mismo. Todos los códigos computacionales fueron escritos en el lenguaje Matlab, y se encuentran a disposición de los interesados en la página de Investigación reproducible <http://www.famaf.unc.edu.ar/~flesia/>

Como trabajo futuro se pretende culminar con el estudio de la Normalidad Asintótica del estimador de PMV, que permite asignar intervalos de confianza a las estimaciones propuestas y que proporciona una herramienta necesaria para la realización de test de hipótesis pertinentes a la selección del modelo adecuado a la clase de

configuraciones con las que se trabaja. Esta selección de modelo puede analizarse desde diversas perspectivas, de las cuales son de interés el orden de la vecindad considerado y la isotropía o no del mismo.

Los trabajos de investigación en consideración editorial y en prensa correspondientes a esta tesis son:

1. J. Gimenez, A. C. Frery, and A. G. Flesia, "Inference strategies for the smoothness parameter in the Potts Model". To appear in Proceedings IGARSS 2013, IEEE Press.
2. J. Baumgarter, A. Flesia, J. Gimenez, and J. Pucheta, "A new approach to image segmentation with two dimensional hidden Markov models". To appear Proceeding BRICS 2013, IEEE press.
3. A. Flesia, J. Baumgarter, J. Gimenez, and J. Martinez, "On classification accuracy of image segmentation methods with Markovian prior laws". Submitted to Pattern Recognition, 2013, Elsevier press.
4. J. Gimenez, A.C. Frery, A. G. Flesia. "When Data do not Bring Information: A Case Study in Markov Random Fields Estimation". Submitted to JSTARS, IEEE press, 2013.
5. J. Gimenez, A.G. Flesia. "On consistency of Pseudolikelihood estimators under non-invariant specifications". Preprint.

Bibliografía

- [1] J. Besag, “Statistical analysis of non-lattice data,” *Journal of the Royal Statistical Society. Series D (The Statistician)*, vol. 24, no. 3, pp. 179–195, 1975.
- [2] D. Strauss, “Clustering on colored lattice,” *Journal of Applied Probability*, vol. 14, pp. 135–143, 1977.
- [3] M. Janzura, “Asymptotic results in parameter estimation for gibbs random fields,” *Kybernetika*, vol. 33, pp. 135–159, 1997.
- [4] S. Geman and C. Graffigne, “Markov random field image models and their applications to computer vision,” *Proc. Int’l Congress of Mathematicians 1986. Providence, R.I.: Am. Math. Soc.*, 1987.
- [5] F. Comets, “On consistency of a class of estimators for exponential families of Markov random fields on the lattice,” *The Annals of Statistics*, vol. 20, no. 1, pp. 455–468, 1992.
- [6] S. Geman and D. Geman, “Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 6, no. 6, pp. 721–741, 1984.
- [7] X. Guyon, *Random Fields on a Network: Modeling, Statistics, and Applications*, ser. Graduate Texts in Mathematics. Springer, 1995.
- [8] C. Gaetan and X. Guyon, “Spatial statistics and modeling,” *Springer Series in Statistics*, 2010.
- [9] X. Descombes, R. D. Morris, J. Zerubia, and M. Berthod, “Estimation of Markov random field prior parameters using Markov chain Monte Carlo maximum likelihood,” *IEEE Transactions on Image Processing*, vol. 8, no. 7, pp. 954–963, Jul. 1999.
- [10] I. Murray and Z. Ghahramani, “Bayesian learning in undirected graphical models: Approximate MCMC algorithms,” in *Proceedings Conference Uncertainty in Artificial Intell. (UAI). Arlington, Virginia: AUAI Press*, pp. 392–399, 2004.
- [11] L. Risser, J. Idier, P. Ciuciu, and T. Vincent, “Fast bilinear extrapolation of 3D Ising field partition function. application to fMRI image analysis,” in *ICIP. IEEE*, 2009, pp. 833–836.
- [12] L. Risser, T. Vincent, F. Forbes, J. Idier, and P. Ciuciu, “Min-max extrapolation scheme for fast estimation of 3D Potts field partition functions. Application to the joint detection-estimation of brain activity in fMRI,” *J. Signal Process. Syst.*, vol. 65, no. 3, pp. 325–338, dec 2011.
- [13] M. Pereyra, N. Dobigeon, H. Batatia, and J. Tourneret, “Estimating the Granularity parameter of a Potts-Markov random field within an MCMC algorithm,” *IEEE Transactions on Image Processing*, vol. 22, no. 6, pp. 2385–2397, 2013.

- [14] J. Besag, "On the statistical analysis of dirty pictures," *Journal of the Royal Statistical Society*, vol. B-48, no. 3, pp. 259–302, 1986.
- [15] A. C. Frery, A. H. Correia, and C. C. Freitas, "Classifying multifrequency fully polarimetric imagery with multiple sources of statistical evidence and contextual information," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 10, pp. 3098–3109, 2007.
- [16] A. L. M. Levada, N. D. A. Mascarenhas, and A. Tannús, "Pseudolikelihood equations for the Potts MRF model parameters estimation on higher order neighborhood systems," *IEEE Geoscience and Remote Sensing Letters*, vol. 5, no. 3, pp. 522–526, 2008.
- [17] R. C. Gonzalez and R. E. Woods, *Digital Image Processing. 2nd edition*. Prentice-Hall, 2001.
- [18] J. P. Alvarado Moya, "Notas de clase: Procesamiento y análisis de imágenes digitales," Instituto Tecnológico de Costa Rica. Escuela de Ingeniería Electrónica, Costa Rica, Tech. Rep., 2013.
- [19] J. G. Flesia and A. G. Flesia, "The influence of processing in the accuracy of measurements in indirect digitalized intra-oral radiographic imaging for forensic applications." *The Forensic Oral Pathology Journal*, no. 4, pp. 20–24, 2011.
- [20] P. Fontanetti, P. Mandalunis, and N. Vermouth, "Response of bone tissue associated with eruption of rat first mandibular molar of pups born from mothers subjected to constant light during pregnancy." *Bone*, vol. 49, no. 6, pp. 1380–1385, 2011.
- [21] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society, serie B*, vol. 39, no. 1, pp. 1–38, 1977.
- [22] S. Borman, "The expectation maximization algorithm: A short tutorial," Tech. Rep., 2004.
- [23] Chin-Wei and B. Rajeswari, "Multiobjective optimization approaches in image segmentation - the directions and challenges," *International Journal on Advance Soft Computing and its Applications*, vol. 2, no. 1, pp. 40–65, 2010.
- [24] L. Li and R. Gray, *Image Segmentation and Compression Using Hidden Markov Models*, ser. The Kluwer International Series in Engineering and Computer Science. Kluwer Academic Publishers, 2000.
- [25] J. Baumgarter, A. Flesia, J. Gimenez, and J. Pucheta, "A new approach to image segmentation with two-dimensional hidden Markov models," Tech. Rep., 2013.
- [26] A. Flesia, J. Baumgarter, J. Gimenez, and J. Martinez, "On classification accuracy of image segmentation methods with Markovian prior laws," FaMAF-UNC, Argentina, Tech. Rep., 2013.
- [27] H. Georgii, *Gibbs Measures and Phase Transitions*, ser. De Gruyter Studies in Mathematics, 9. Walter De Gruyter Incorporated, 1988, no. v. 9.
- [28] J. Besag, "Spatial interaction and the statistical analysis of lattice systems," *Journal of the Royal Statistical Society*, vol. B-36, no. 2, pp. 192–236, 1974.
- [29] O. Barndorff-Nielsen, *Information and Exponential Families: In Statistical Theory*, ser. Wiley Series in Probability and Mathematical Statistics Series. J. Wiley, 1978.

- [30] B. Gidas, “Consistency of maximum likelihood and pseudo-likelihood estimators for gibbs distributions,” *The IMA Volumes in Mathematics and Its Applications*, vol. 10, pp. 129–145, 1988.
- [31] F. Comets and B. Gidas, “Parameter Estimation for Gibbs Distributions from Partially Observed Data,” *The Annals of Applied Probability*, vol. 2, no. 1, pp. 142–170, 1992.
- [32] D. K. Pickard, “Inference for discrete markov fields: the simplest nontrivial case,” *Journal of the American Statistical Association*, vol. 82, no. 397, pp. 90–96, 1987.
- [33] J. Gimenez, A. C. Frery, and A. G. Flesia, “Inference strategies for the smoothness parameter in the Potts Model. In proceedings IGARSS 2013 IEEE Xplore,” 2013.
- [34] V. Yohai, *Notas de Clase - Probabilidad y Estadística - Universidad de Buenos Aires*.
- [35] R. H. Swendsen and J. S. Wang, “Nonuniversal, critical dynamics in Monte Carlo simulations,” *Physical Review Letters*, 1987.
- [36] R. D. Morris, “Auxiliary variables for Markov random fields with higher order interactions,” Tech. Rep., 1998.
- [37] C. A. Mcgrory, D. M. Titterington, R. Reeves, and A. N. Pettitt, “Variational Bayes for estimating the parameters of a hidden Potts model,” *Statistics and Computing*, vol. 19, no. 3, pp. 329–340, Sep. 2009.
- [38] L. Younes, “Estimation and annealing for gibbsian fields,” *Annales de l’institut Henri Poincaré (B) Probabilités et Statistiques*, vol. 24, no. 2, pp. 269–294, 1988.
- [39] J. Liu, L. Wang, and S. Li, “MRF parameter estimation by MCMC method,” *Pattern Recognition*, vol. 33, no. 11, pp. 1919–1925, Nov. 2000.
- [40] M. V. Ibáñez and A. Simó, “Parameter estimation in Markov random field image modeling with imperfect observations: a comparative study,” *Pattern Recognition Lett.*, vol. 24, no. 14, pp. 2377–2389, Oct. 2003.
- [41] S. Okabayashi, L. Johnson, and C. J. Geyer, “Extending pseudo-likelihood for potts models,” *Statistica Sinica*, vol. 21, no. 1, pp. 331–347, 2011.
- [42] N. Friel, “Bayesian inference for gibbs random fields using composite likelihoods,” in *Winter Simulation Conference*, 2012, p. 28.
- [43] H. Derin and H. Elliott, “Modeling and segmentation of noisy and textured images using gibbs random fields,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 9, no. 1, pp. 39–55, 1987.
- [44] A. M. Ali, A. A. Farag, and G. L. Gimel Farb, “Analytical method for MGRF Potts model parameter estimation,” in *Proceedings of Pattern Recognition, 2008. ICPR 2008. 19th International Conference on*, pp. 1–4, 2008.
- [45] R. Brent, “Algorithms for minimization without derivatives,” *Prentice Hall, New York*, 1973.
- [46] A. C. Frery, S. Ferrero, and O. H. Bustos, “The influence of training errors, context and number of bands in the accuracy of image classification,” *International Journal of Remote Sensing*, vol. 30, no. 6, pp. 1425–1440, 2009.

- [47] Q. Jackson and D. A. Landgrebe, "Adaptive Bayesian contextual classification based on Markov random fields," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 40, no. 11, pp. 2454–2463, 2002.
- [48] G. M. Arbia, R. Benedetti, and G. Espa, "Contextual classification in image analysis: an assessment of accuracy of ICM," *Computational Statistics & Data Analysis*, vol. 30, no. 4, pp. 443–455, 1999.
- [49] F. Melgani and S. B. Serpico, "A Markov random field approach to spatio-temporal contextual image classification." *IEEE Transactions on Geoscience and Remote Sensing*, vol. 41, no. 11, pp. 2478–2487, 2003.
- [50] J. Besag, "Digital image processing: Towards bayesian image analysis," *Journal of Applied Statistics*, vol. 16, no. 3, pp. 395–407, 1989.
- [51] Arnold and Strauss, "Pseudolikelihood estimation: some examples." *Sankhya: The Indian Journal of Statistics Series*, vol. 53, pp. 233–243, 1991.
- [52] R. Morris, X. Descombes, and J. Zerubia, "The ising/potts model is not well suited to segmentation tasks," *Proc. Digital Signal Processing Workshop*, pp. 263–266, 1996.
- [53] C. Ji and L. Seymour, "A consistent model selection procedure for markov random fields based on penalized pseudolikelihood," *The Annals of Applied Probability*, vol. 6, no. 2, pp. 423–443, 1996.
- [54] A. Grelaud, C. Robert, and J. T. J.M. Marin, F. Rodolphe, "Abc likelihood-free methods for model choice in gibbs random fields," *Brazilian Journal of Probability and Statistics*, vol. 23, pp. 120–140, 2009.
- [55] J. Besag, "Discussion of paper by P. Switzer," *Bull. Int. Statist. Inst.*, vol. 50, no. 3, pp. p.p. 422–425, 1983.
- [56] B. C. K. Tso and P. M. Mather, "Classification of multisource remote sensing imagery using a genetic algorithm and Markov random fields." *IEEE Transactions on Geoscience and Remote Sensing*, vol. 37, no. 3, pp. 1255–1260, 1999.