

UNIVERSIDAD NACIONAL DEL COMAHUE
Maestría en Estadística Aplicada



TITULO DE LA TESIS

*Sensibilidad de Métodos de Estimación de
parámetros de la distribución de Gibbs
Autobinomial para clasificación de imágenes
satelitales*

Trabajo Presentado por:

Silvina Pistonesi

Lic. en Matemática
(Universidad nacional del Sur, 1996)

Neuquén, 2010.

COMISIÓN ASESORA DE TESIS

Director

Dr. Oscar H. Bustos

Co - Director

Mg. Mariana Santamaría

JURADO DE TESIS

Fecha de aprobación de la tesis: días de de 2010

Nombre del Director: Dr. Oscar. H Bustos

Certifica:

Que *Silvina Pistonesi* ha realizado bajo mi dirección la tesis “*Sensibilidad de Métodos de Estimación de parámetros de la distribución de Gibbs Autobinomial para clasificación de imágenes satelitales*” la cual se encuentra finalizada y en condiciones de ser presentada.

Y para que así conste a todos los efectos del interesado se expide el presente certificado en Neuquén a los días del mes de del año 2010.

AGRADECIMIENTOS

Principalmente a mi director de tesis, Dr. Oscar Bustos, que me alentó y me acompañó en todo momento y desde un principio para la concreción de este trabajo.

A mi familia: mi novio, Jorge, mis padres, Lucy y Héctor y mi hermano, Adrián, que sin su apoyo incondicional no lo hubiese logrado.

ÍNDICE DE TABLAS	IX
ÍNDICE DE FIGURAS	XI
GLOSARIO DE ABREVIATURAS	XIV
CAPÍTULO 1: INTRODUCCIÓN	1
1.1. Descripción del Problema	1
1.2. Disciplinas o teorías involucradas	2
1.3. Estado del conocimiento del tema	2
1.4. Fundamentación del tema elegido	6
1.5. Objetivos del trabajo	7
1.6. Estructura de la tesis	8
CAPÍTULO 2: CAMPOS ALEATORIOS DE MARKOV	11
2.1. Conceptos Generales	11
2.2. Sistemas de vecinos y cliques	13
2.3. Campos Aleatorios de Gibbs (CAG)	18
2.4. Equivalencia entre CAM y CAG	23
2.5. Modelos de Textura	26
2.5.1. Modelo de Ising	28
2.5.2. Modelo Autobinomial	33
CAPÍTULO 3: SIMULACIÓN DE DISTRIBUCIONES DE GIBBS	40
3.1. Algoritmo Metropolis	41
3.2. Algoritmo Gibbs Sampler (GS)	44
CAPÍTULO 4: ESTIMADORES DE LOS PARÁMETROS EN DISTRIBUCIONES DE GIBBS	53
4.1. Estimación de Parámetros	53
4.1.1. Estimador de Máxima Verosimilitud	55
4.1.2. Coding	57
4.1.3. Estimador de Máxima Pseudo-verosimilitud	58

4.1.4. Estimador de Mínimos Cuadrados Condicional	60
CAPÍTULO 5: ESTUDIO DE SENSIBILIDAD	64
5.1. Comparación y Análisis de sensibilidad	64
5.1.1. Visualización del efecto de contaminación	73
CAPÍTULO 6: CLASIFICACIÓN SUPERVISADA DE UNA IMAGEN SATELITAL	78
6.1. Teledetección espacial	78
6.2. Imágenes digitales	82
6.3. Clasificación de imágenes	83
6.4. Aplicación: Clasificación supervisada de una imagen satelital	90
6.4.1. Resultados	94
CAPÍTULO 7: CONCLUSIONES Y FUTURAS LÍNEAS DE INVESTIGACIÓN	102
7.1. Conclusiones	102
7.2. Aportes esperados de la Tesis	105
7.3. Líneas futuras de trabajo	105
BIBLIOGRAFÍA	108
APÉNDICES	115
A. NOCIONES MATEMÁTICAS DE LAS CADENAS DE MARKOV Y DE METODOS DE OPTIMIZACIÓN	116
A.1. Cadenas de Markov.	116
A.1.1. Conceptos Básicos	116
A.1.2. Coeficiente de Contracción	120
A.1.3. Convergencia de cadenas de Markov Homogéneas	122
A.2. Algoritmo Gibbs Sampler	124
A.3. Método de Maximización	129
A.3.1. Preliminares	129
A.3.2 Método de Newton	133
B. ALGORITMOS DE LOS MÉTODOS MONTE CARLO	136

B.1. Algoritmo de Metropolis	136
B.2. Algoritmo Gibbs Sampler	137
B.3. Algoritmo de Contaminación	140
B.4. Algoritmo del método de estimación Mínimos de Cuadrados Condicional	141
C. RESULTADOS DE LAS ESTIMACIONES	144
C.1. Modelo Autobinomial H-V	144
C.1.1. Estimador MCC	144
C.1.2. Estimador MP	145
C.2. Modelo Autobinomial I-D	145
C.2.1. Estimador MCC	145
C.2.2. Estimador MP	146

CAPÍTULO 3

3.1 Parámetros utilizados para generar las texturas de la Figura 3.1.	51
3.2 Parámetros utilizados para generar las texturas de la Figura 3.2.	51

CAPÍTULO 5

5.1. Las cuatro situaciones de valores específicos de los parámetros de los modelos.	64
5.2. Resultados de las estimaciones de los parámetros del Modelo Autobinomial H-V Puro.	66
5.3. Resultados de las estimaciones de los parámetros del Modelo Autobinomial I-D Puro.	67
5.4. Resultados de las estimaciones de los parámetros del Modelo Autobinomial H-V contaminado.	71
5.5. Resultados de las estimaciones de los parámetros del Modelo Autobinomial I-D contaminado.	72

CAPÍTULO 6

6.1 Matriz de confusión. Efectividad de la clasificación.	88
6.2 Estimaciones de los parámetros de las regiones de referencia.	94
6.3 Estimaciones de los coeficientes de Lin.	95
6.4 Matrices de confusión generadas.	99
6.5 Porcentajes de errores de omisión y comisión.	99
6.6 Estimaciones del coeficiente Kappa para cada región y modelo Autobinomial.	100

APÉNDICE C

Tabla C.1 Resultados de las estimaciones del vector de parámetros $\theta_{H-V} = (\beta_{11}, \beta_{12})$, del Modelo Autobinomial H-V Puro y Contaminado, en las cuatro situaciones consideradas.	144
Tabla C.2 Resultados de las estimaciones del vector de parámetros $\theta_{H-V} = (\beta_{11}, \beta_{12})$, del Modelo Autobinomial H-V Puro y Contaminado, en las cuatro situaciones consideradas.	145

Tabla C.3 Resultados de las estimaciones del vector de parámetros $\theta_{I-D} = (\beta_{21}, \beta_{22})$, del Modelo Autobinomial I-D Puro y Contaminado, en las cuatro situaciones consideradas. 145

Tabla C.4 Resultados de las estimaciones del vector de parámetros $\theta_{I-D} = (\beta_{21}, \beta_{22})$, del Modelo Autobinomial I-D Puro y Contaminado, en las cuatro situaciones consideradas. 146

CAPÍTULO 2

2.1. Ejemplo de una imagen. Arreglo rectangular de puntos (píxeles).	12
2.2. Notación matricial de una imagen. Cada x_{ij} representa un píxel de la imagen.	12
2.3. Vecindarios de órdenes 1, 2, 4, 5, 8 del sitio $s = \bullet$ y $t = \circ \in \partial(s)$.	15
2.4. Arreglo de sitios irregularmente espaciados. Sistema de vecinos de orden r del píxel del sitio s .	15
2.5. Tipos de cliques: (a) para el sistema de vecinos de orden uno. (b) para sistemas de vecinos de segundo orden.	16
2.6. Entramado en una dimensión.	29
2.7. Entramado bidimensional.	29
2.8. Entramado tridimensional.	29
2.9. Una configuración de los espines. Los espines orientados en la dirección positiva del eje Z corresponden al valor 1 de las variables X_s y los orientados en la dirección negativa de Z , a los valores -1 de las variables X_s .	30
2.10. Sistema de vecinos cercanos del sitio $s = (i, j)$.	31
2.11. Sistema de vecinos de 2º orden del sitio s .	34
2.12. Sistemas de vecinos considerados en cada modelo: (a) Autobinomial H-V (b) Autobinomial I-D.	37

CAPÍTULO 5

5.1. Modelo Autobinomial H-V. La textura de la izquierda está generada utilizando los valores específicos ($\beta_{11} = 0, \beta_{12} = 1$), la segunda presenta una contaminación del 1% de los píxeles, la siguiente un 10% y la de la derecha un 20% de los mismos.	65
5.2. Modelo Autobinomial I-D. La textura de la izquierda está generada utilizando los valores específicos ($\beta_{11} = 1, \beta_{12} = -1$), la segunda presenta una contaminación del 1% de los píxeles, la siguiente un 10% y la de la derecha un 20% de los mismos.	65
5.3. Modelo Autobinomial H-V. La textura de la izquierda está generada utilizando los valores específicos ($\beta_{11} = 1, \beta_{12} = -1$), la del medio usando los valores de los parámetros estimados a partir del método MCC y la de la derecha usando el método M P.	68

- 5.4. Modelo Autobinomial I-D. La textura de la izquierda está generada utilizando los valores específicos ($\beta_{21} = -1, \beta_{22} = 1$), la del medio usando los valores de los parámetros estimados a partir del método MCC y la de la derecha usando el método de MP. 68
- 5.5. Sesgos de las estimaciones de los parámetros de los Modelos Autobinomiales H-V e I-D, realizadas con los distintos métodos. Modelo H-V: **(a)** para los valores específicos ($\beta_{11} = 0, \beta_{12} = 1$), **(b)** para ($\beta_{11} = 1, \beta_{12} = 1$). Modelo I-D: **(c)** Para los valores específicos ($\beta_{21} = 1, \beta_{22} = -1$), **(d)** para ($\beta_{21} = -1, \beta_{22} = 1$). 69
- 5.6. Desvíos estándar de las estimaciones de los parámetros de los Modelos Autobinomiales H-V e I-D, realizadas con los distintos métodos. Modelo H-V: **(a)** para los valores específicos ($\beta_{11} = 0, \beta_{12} = 1$), **(b)** para ($\beta_{11} = 1, \beta_{12} = 1$). Modelo I-D: **(c)** para los valores específicos ($\beta_{21} = 1, \beta_{22} = -1$), **(d)** para ($\beta_{21} = -1, \beta_{22} = 1$). 70
- 5.7. Comparación de los métodos de estimación MCC y MP: a izquierda, el Diagrama de dispersión de las estimaciones del vector de parámetros $\vartheta_{H-V} = (\beta_{11}, \beta_{12})$ del Modelo Autobinomial H-V Puro, para el vector de valores especificados $\vartheta_{H-V} = (1, 1)$. A derecha, el Diagrama de dispersión de las estimaciones del vector de parámetros $\vartheta_{I-D} = (\beta_{21}, \beta_{22})$ del Modelo Autobinomial I-D Puro, para el vector de valores especificados $\vartheta_{I-D} = (1, 1)$. 74
- 5.8. Comparación de los métodos de estimación MCC y MP: elipses de confianza del 95% para estimar el vector $\vartheta_{H-V} = (\beta_{11}, \beta_{12})$ y $\vartheta_{I-D} = (\beta_{21}, \beta_{22})$, respectivamente. Considerando los casos: $\vartheta_{H-V} = (1, 1)$ y $\vartheta_{I-D} = (1, 1)$, respectivamente. 75
- 5.9. Comparación de los modelos H-V e I-D frente a un 20 % de contaminación de los píxeles de la textura: elipses de confianza del 95% para estimar el vector $\vartheta_{H-V} = (\beta_{11}, \beta_{12})$ y $\vartheta_{I-D} = (\beta_{21}, \beta_{22})$. Considerando el caso $(-1, 1)$, para los vectores de parámetros de ambos modelos. 75
- 5.10. Comparación de los estimadores MCC y MP frente a un 20 % de contaminación de los píxeles de la textura: elipses de confianza del 95% para estimar el vector $\vartheta_{H-V} = (\beta_{11}, \beta_{12})$. Considerando el caso $\vartheta_{H-V} = (1, -1)$. 76

CAPÍTULO 6

- 6.1. Los componentes de un sistema de Teledetección espacial: (A) Fuente de energía. (B) Atmósfera. (C) Cubierta terrestre. (D) Sensor. (E) Sistema de Transmisión, recepción y procesamiento. (F) Interprete. (G) Usuario. 80
- 6.2. Organización de datos de ocho bandas espectrales sobre una misma escena. 83
- 6.3. Imagen satelital elegida. 91
- 6.4. Imagen satelital y subimagen seleccionada para el estudio. 92

6.5. Marcadores de región: región I en verde y la región II en azul.	93
6.6. Regiones de referencia elegidas en la imagen. A la derecha, la región I y a la izquierda la región II.	93
6.7. Gráfico de la región I considerando: (a) el modelo Autobinomial H-V y (b) el modelo Autobinomial I-D.	96
6.8. Gráfico de la región II considerando: (a) el modelo Autobinomial H-V y (b) el modelo Autobinomial I-D.	97
6.9. Comparación de los clasificadores: (a) según región I, considerando el modelo Autobinomial H-V. (b) según región II, considerando el modelo Autobinomial I-D. En ambos casos, lo negro son coincidencias y los parches blancos representan las discordancias entre los dos clasificadores.	98

APÉNDICE A

A.3.1. Conjunto convexo y no convexo.	130
A.3.2. Función convexa y función cóncava.	132

GLOSARIO DE ABREVIATURAS

CAG: Campo Aleatorio de Gibss.

CAM: Campo Aleatorio de Markov.

ETM: Enhanced Thematic Mapper.

GS: Gibbs Sampler

H-V: Modelo Atobinomial Horizontal- Vertical.

I-D: Modelo Atobinomial Diagonal Izquierda- Derecha

ISODATA: Interactive Self-Organizing Data Analysis Technique Algorithm.

MP: Maxima Pseudo-verosimilitud.

MV: Máxima Verosimilitud.

MCC: Mínimos Cuadrados Condicional.

ND: Nivel digital.

SAR: Sintetic Aperture Radar.

Las imágenes satelitales contienen una enorme cantidad de información espacial. Para capturarla varios autores proponen, en el marco del modelado estocástico de una imagen, el uso de Campos Aleatorios Gibbs-Markov. Entre la variedad de modelos de Markov, diversos autores sugieren, en particular, el modelo Autobinomial como modelo apropiado para la caracterización de datos reales y generación de texturas sintéticas. Esta clase de modelos estocásticos provee una descripción intuitiva de la imagen a través del vector de parámetros de su función energía. La estimación de este vector de parámetros es el paso clave en la extracción de la información espacial y en la caracterización del contenido de una imagen.

El objetivo principal de este trabajo de tesis es evaluar la sensibilidad de los métodos de estimación del vector de parámetros de un modelo Gibbs - Autobinomial: Máxima Pseudo- verosimilitud y el estimador de Mínimos Cuadrados Condicional ante la presencia de “outliers”, desviaciones del modelo supuesto y/o distintos grados de contaminación de diferentes patrones homogéneos de textura. Para lograrlo se efectuó un estudio Monte Carlo en ciertas situaciones específicas del modelo Autobinomial. Se implementó un algoritmo para simular distintas texturas a partir de la asignación de diferentes valores al vector de parámetros del modelo Autobinomial y se efectuaron modificaciones necesarias a las rutinas de software disponibles para determinar las estimaciones de los parámetros de un modelo Gibbs - Autobinomial utilizando los estimadores mencionados. Por otra parte, se estudió la performance de reglas de clasificación basadas en la hipótesis de que en pequeños sectores de imágenes satelitales puede considerarse que la textura de la imagen está descrita razonablemente bien por modelos Autobinomiales con pocos parámetros.

En el análisis presentado, en general, los métodos MCC y MP emitieron resultados equivalentes. Sin embargo, el estimador MP presentó una mejor performance en las

estimaciones que el MCC. El método MCC resultó ser más sensible que el MP frente a los distintos grados de contaminación considerados.

Los resultados obtenidos de la clasificación de la imagen satelital, a través de los métodos CMC y MP mostraron nivel de concordancia “muy bueno”, por lo que, podría concluirse que serían clasificadores intercambiables.

PALABRAS CLAVES: Campos Aleatorios Gibbs-Markov, Texturas Sintéticas, Modelo Autobinomial, Estimador de Máxima Pseudo-verosimilitud, Estimador de Mínimos Cuadrados Condicional, Sensibilidad.

Satellite images contain an enormous amount of spatial information. In order to capture it, several authors suggest the use of Gibbs- Markov Random Fields, in the framework of the stochastic modelling of an image. Among the variety of Markov models, some authors particularly propose the Autobinomial one as the appropriate for the characterization of real data and the generation of synthetic textures. This type of stochastic models provides an intuitive description of the image through the parameters vector of its energy function. The estimate of this parameters vector is the key step in extracting spatial information and for image content characterization.

The main objective of this thesis is to assess the sensitivity of the methods of estimating the parameters vector of a Gibbs-Autobinomial model: Maximum Pseudo likelihood and Conditional Least Squares estimator in the presence of outliers, which are deviations from the assumed model and/or diverse degrees of contamination of different homogeneous texture patterns. To accomplish this, a Monte Carlo study was conducted in specific situations of the Autobinomial model. An algorithm was implemented for the purpose of simulating different textures on the basis of assigning different values to the parameters vector of the Autobinomial model. All necessary amendments were made to available software routines in order to determine estimates of the parameters of a Gibbs - Autobinomial model using the estimators mentioned.

On the other hand, the performance of classification rules was studied. These rules were based on the assumption that in small sections of satellite images it can be considered that the image texture is reasonably well described by Autobinomial models with a small number of parameters.

In the presented analysis, MCC and MP methods showed equivalent results in general. However, the MP estimator presented a better estimation performance than the MCC, whereas the latter method proved to be more sensible than the first in the face of the different degrees of contamination considered.

The results obtained from satellite image classification by means of CMC and MP methods revealed a very good concordance level. Therefore, it can be concluded that they would be interchangeable classifiers.

KEYWORDS: Gibbs-Markov Random Fields, Synthetic Textures, Autobinomial Model, Maximum Pseudolikelihood Estimator, Conditional Least Squares Estimator, Sensitivity.

1.1 Descripción del Problema

Las imágenes satelitales contienen una enorme cantidad de información espacial. Para capturar esta información Schroderl et al. [80] proponen, en el marco del modelado estocástico de una imagen, el uso de Campos Aleatorios Gibbs-Markov. Esta clase de modelos estocásticos provee una descripción intuitiva de parámetros de la imagen usando parámetros de una función energía. Tales autores sugieren, en particular, el modelo Autobinomial como modelo apropiado para la caracterización de datos reales y generación de texturas sintéticas. En este modelo la función de energía esta definida por la expresión:

$$\mathbf{H}(\mathbf{x}_s, \partial_s; \boldsymbol{\theta}) = -\ln \binom{M}{x_s} - \mathbf{x}_s \boldsymbol{\eta}_s, \quad (1.1)$$

donde x_s denota la realización de un píxel ubicado en la posición s , ∂_s representa el sistema de vecinos del píxel x_s , M indica el máximo valor de gris y $\binom{M}{n}$ denota los coeficientes de la distribución Binomial.

La influencia de los píxeles vecinos ponderada por los parámetros β_{ij} está representada por la cantidad:

$$\boldsymbol{\eta}_s = \mathbf{a} + \sum_{i,j} \beta_{ij} \frac{(x_{ij} + x_{ij}^*)}{M} \quad (1.2)$$

Todos los parámetros reunidos en un vector

$$\boldsymbol{\theta} = (\mathbf{a}, \beta_{11}, \beta_{12}, \beta_{21}, \beta_{22}, \beta_{31}, \beta_{32}, \dots)^T$$

determinan las propiedades estadísticas de las variables aleatorias $\{X_s\}$, esto es, la apariencia en la imagen que es una realización típica de un tal campo aleatorio.

La estimación del vector de parámetros θ de la función de energía es el paso clave en la extracción de la información espacial usando modelos de Gibbs. Si bien existen varios métodos de estimación de estos parámetros [39], según la literatura los más apropiados son: Estimador Máxima Pseudo-verosimilitud (MP) [6] y el de Mínimos Cuadrados Condicional (MCC) [58]. Este último método es el que se utiliza con más frecuencia para el modelo Autobinomial por sus propiedades.

En virtud de la relevancia de la estimación del vector θ , es fundamental conocer las propiedades de tales estimadores y evaluar su comportamiento ante distintos grados de contaminación a diferentes patrones homogéneos de textura.

1.2 Disciplinas o teorías involucradas

El tema desarrollado en este trabajo de tesis se vincula fundamentalmente con *la teoría Estadística* (criterios de estimación de parámetros de un modelo Autobinomial en el marco de la teoría de Campos Aleatorios Gibbs-Markov y herramientas para evaluar la sensibilidad de dichos estimadores frente a “outliers” o desviaciones del modelo supuesto en ciertas situaciones específicas del modelo Autobinomial).

Además, esta vinculado con la *Informática*, en la medida en que se deberá recurrir a algoritmos que permitan simular distintas texturas a partir de la asignación de diferentes valores al vector de parámetros del modelo Autobinomial y la estimación del mismo.

También se relaciona con las disciplinas vinculadas con *Teledetección de Imágenes*, dado que las técnicas estadísticas e informáticas habrán de ser aplicadas a cuestiones vinculadas a la interpretación de imágenes satelitales.

1.3 Estado del conocimiento del tema

A través de los años, los modelos probabilísticos aplicados al procesamiento de imágenes, han sido la temática de varias publicaciones científicas, conferencias, y talleres. En la actualidad existen numerosos trabajos respecto al modelado de

imágenes. La mayor parte de estos modelos probabilísticos han sido desarrollados y exitosamente empleados en otros campos científicos como por ejemplo, en la Estadística Física, el Procesamiento de Señales o la Teoría de las Mediciones.

Entre la variedad de modelos estocásticos que han sido usados para generar y representar texturas se encuentran los Campos Aleatorios de Markov (CAM) [43-17]. Hammersley and Clifford en 1971 [42], caracterizaron los modelos de Markov como distribuciones de Gibbs; proporcionando el uso de esta equivalencia como una simplificación en la especificación del modelo y permitiendo obtener realizaciones del mismo.

En 1974, Besag publica su trabajo más reconocido [6]. En él considera modelos definidos a partir de las características locales. Describe los CAM y Campos Aleatorios de Gibbs (CAG) a partir de funciones energéticas, potenciales y cliques. Introduce los modelos Autobinomiales, los Autopoissonianos y Autoexponenciales. Analiza los modelos Binarios y Gaussianos. Para estimar los parámetros de los modelos, define el método de codificación (Coding method). Proporciona una nueva demostración del teorema de equivalencia de Hammersley y Clifford usando una expansión de la función de energía que se dice única.

En 1975, Besag y Moran [8] sientan las bases para realizar estimaciones en campos Gaussianos. En este trabajo, se sugiere el método de Máxima Pseudo-verosimilitud (MP) para la estimación de parámetros en un contexto bastante general de CAM, como alternativa a la estimación por máxima verosimilitud. En 1977, Besag [9] comprueba la eficiencia del método de estimación MP para el caso de campos Gaussianos y realiza una comparación entre los métodos Máxima Verosimilitud (MV), Coding y MP.

Hassner y Slansky en 1980 [43] analizan el uso de CAM como modelos estocásticos de textura y presentan varios ejemplos particulares de síntesis de texturas, pero no discuten el ajuste de los CAM para describir texturas reales. El trabajo de Cross y Jain [17] constituyó uno de los primeros artículos referidos al modelado de texturas reales a partir de CAM. Sugieren al modelo Autobinomial, como modelo apropiado para la caracterización de datos reales y generación de texturas sintéticas. Con el propósito de sintetizar texturas es necesario estimar los parámetros del modelo. Cross y Jain utilizaron el algoritmo Metropolis (63) para la síntesis y el método de códigos (6) como proceso de estimación de parámetros. Las bases matemáticas de este trabajo se deben a Besag [6].

Acuna en 1992 [1] propone una modificación al modelo Autobinomial introducido por Besag en [7] y un método de estimación Pseudo-verosimilitud bivariado. En 1994 extiende la utilidad de esta técnica de estimación de parámetros mediante la flexibilización de un cierto supuesto [2].

En 1997, M. Schöder, K. Seidel, and M. Dactu [79] emplean los Campos Aleatorios de Gibbs-Markov para capturar información espacial de imágenes satelitales, y desarrollan en particular, el modelo Autobinomial [17], modelo apropiado para dicho propósito. Se demuestran las capacidades del mismo a través de ejemplos. En 1998 dichos autores [78] completan este trabajo estableciendo los métodos de estimación de parámetros del modelo con más detalle y aplicando este modelo a dos tipos de datos ópticos: Landsat TM y RESURS-01.

En 1999, Gimel' fard [12] centra su trabajo en texturas de imágenes y presenta nuevas metodologías para su simulación, recuperación y segmentación, utilizando Campos Aleatorios de Gibbs con múltiples interacciones de pares entre señales como modelos probabilísticos de imágenes.

A principios del año 2000, Schröder et al. [80] presentan a los modelos de Gibbs: Gaussiano, Autobinomial y el Exponencial, como una familia poderosa de modelos para la extracción de información espacial. Definiéndolos mediante funciones de energía parametrizadas que caracterizan a las interacciones locales entre los píxeles vecinos. En el trabajo, a través de ejemplos, exponen diferentes metodologías tendientes a la extracción de información espacial, que van desde la estimación de parámetros, mediante la selección del modelo que mejor describe los datos de imagen, hasta la segmentación de toda la imagen en regiones con propiedades uniformes del modelo. En el mismo año, Pujol et al [72] efectúan una revisión detallada de los modelos de Campos de Markov en visión; estudian los principales modelos de CAM dependiendo de las formas en que se definen los sistemas de vecinos y las funciones potenciales (“funciones cliqué”), y finalmente, presentan una aplicación concreta a un modelo de CAM para el problema de segmentación de imágenes.

Otras aplicaciones del modelo Autobinomial se presentan en [57] donde se desarrolla e investiga la performance de algoritmos para la detección de cara derivados de consideraciones efectuadas en los CAM. Este trabajo emplea el modelo

Autobinomial para determinar si la información presente en sitios especiales (como por ejemplo, la localización de ojos, nariz, contorno facial) es utilizada por el CAM en realidad para distinguir entre la cara y lo que no es cara. En el mismo año, Ozyildiz et al [67] emplean el mencionado modelo del Campo Aleatorio Gibbs-Markov para modelar textura y la distribución de Gauss bivariada para la modelización del color de una imagen. En la tesis doctoral [26] se presentan también, en forma exhaustiva, los conceptos más importantes concernientes a los Campos Aleatorios de Markov y se hace uso de los modelos de los CAM de textura y contorno para el diagnóstico precoz de malformaciones.

En el 2001, Johansson [50] presenta un estudio sobre el proceso de estimación de los parámetros del modelo Autobinomial, comparando el método de Códigos de Besag con el método de Máxima Pseudo-verosimilitud, como así también los métodos de maximización: Simulated Annealing y el Método de Newton-Raphson. En los métodos de estimación no observa diferencias significativas. Los dos métodos de maximización determinan el mismo máximo, siendo el método de Newton-Raphson el más rápido. Sin embargo este último, resulta imposible de aplicar si la ubicación del máximo global difiere mucho de los valores iniciales. En estos casos, propone la utilización del Simulated Annealing.

Finalmente, entre las publicaciones más recientes respecto a aplicaciones del modelo Autobinomial pueden citarse:

- [20] **Datcu, M., Daschiel, H., Pelizzari, A., Quartulli, M., Galoppo, A., Colapicchioni, A., Pastori, M., Seidel, K., Marchetti, P.G., D'Elia, S.** (2003). *"Information mining in remote sensing image archives - Part A: System Concepts"*. En este artículo se aplica el modelo estocástico Autobinomial para capturar características espaciales, espectrales, y las estructuras geométricas presentes en la imagen de estudio.
- [38] **Gimel' farb, G. and Farag, A. A.** (2005). *"Texture analysis by accurate identification of simple Markovian Models"*. En este trabajo se presenta un algoritmo que permite la identificación precisa (estimación de parámetros) de modelos simples de Gibbs - Markov (entre los que se encuentra el Autobinomial) para lograr una mejor segmentación de imágenes multimodal y la síntesis de algunas variedades de texturas naturales.

- [62] **Lippok, A. and Reulke, R.** (2006). “*Discrimination between Urban Area and Vegetation in high resolution Images Using Markov Random Fields (MRF)*”. En este trabajo se eligió el modelo Autobinomial con el propósito de segmentar imágenes sobre la base de las características de textura.
- [45] **Hebar, M., Gleich, D. and Cucej, Z.** (2009). “*Autobinomial Model for SAR Image Despeckling and Information Extraction*”. Este trabajo emplea el modelo Autobinomial para describir las distintas texturas presentes en imágenes de radar de apertura sintética (SAR). Estimando los parámetros de textura mediante el uso de la inferencia Bayesiana de segundo orden. Los resultados experimentales con el método desarrollado por Hebar y Cia. prueban que la textura se preserva bien y que el ruido es reducido en forma significativa tanto en regiones homogéneas como heterogéneas.

1.4 Fundamentación del tema elegido

Las potencialidades del análisis de una imagen radican esencialmente en la dinámica de innovaciones tecnológicas y teóricas permanentes, que permite el desarrollo de nuevas propuestas metodológicas dirigidas a proponer criterios concretos sobre el uso de estas herramientas adecuadas para cada propósito particular.

Resulta relevante destacar, el aporte continuo del análisis estadístico proveyendo modelos para describir en forma robusta información espacial de datos de teledetección de imágenes. Contribuyendo también a desarrollar algoritmos sistemáticamente de una manera analítica y no heurística, para una gran variedad de problemas.

Al margen de estas consideraciones de carácter teórico-metodológico, debe enfatizarse la existencia de demandas crecientes a cerca de la extracción de información espacial asociada al contenido de una imagen en las prácticas concretas de investigación y de asistencia de carácter profesional, especialmente en las áreas de Geografía, Geología, Agronomía.

Debe destacarse la importancia de los Campos Aleatorios Gibbs-Markov y el modelo Autobinomial para caracterizar el contenido de una imagen, fundamentalmente en

lo que respecta al análisis de la textura. Esta ampliamente demostrado el poder de este último modelo en la generación de distintos tipos de texturas.

1.5 Objetivos del Trabajo

1.5.1 Objetivo general

Evaluar la sensibilidad de los métodos de estimación de parámetros de un modelo Gibbs-Autobinomial: Máxima Pseudo-verosimilitud y el estimador de Mínimos Cuadrados Condicional ante la presencia de “outliers”, desviaciones del modelo supuesto y/o distintos grados de contaminación de diferentes patrones homogéneos de textura.

1.5.2 Objetivos específicos

El objetivo principal se logrará en base a la consecución de los siguientes subobjetivos:

- Desarrollar un algoritmo para simular distintas texturas a partir de la asignación de diferentes valores al vector de parámetros del modelo Autobinomial.
- Revisar las rutinas de software disponibles para determinar las estimaciones de los parámetros de un modelo Gibbs-Autobinomial utilizando los estimadores de Máxima Pseudo-verosimilitud y el estimador de Mínimos Cuadrados Condicional, e implementar modificaciones si es necesario, que permitan adecuarlas a las necesidades concretas.
- Estudiar la performance de reglas de clasificación basadas en la hipótesis de que en pequeños sectores de imágenes satelitales puede considerarse que la textura de la imagen está descrita razonablemente bien por modelos Autobinomiales con pocos parámetros.

1.6 Estructura de la Tesis

La memoria de la tesis se estructura en 7 Capítulos y 3 Apéndices. El contenido de éstos se detalla a continuación:

- El **capítulo 1** corresponde a esta introducción.
- En el **capítulo 2**, se efectúa una breve descripción de los procesos markovianos con grilla finita; se introducen los conceptos fundamentales de los Campos Aleatorios de Markov y de los Campos Aleatorios de Gibbs, y particularmente, se presenta un resultado importante que permitirá representar un Campo Aleatorio de Markov a través de la distribución de probabilidad de Gibbs. Finaliza el capítulo con el desarrollo del modelo de Ising y el Modelo Autobinomial. A partir de este último, se considerarán dos nuevos modelos: Modelo Autobinomial H-V y el I-D.
- En el **capítulo 3**, se desarrollan en detalle las técnicas de Simulación de distribuciones de Gibbs: Algoritmo de Metrópolis y de Gibbs.
- En el **capítulo 4**, se estudian los estimadores de los parámetros en distribuciones de Gibbs y en particular los distintos estimadores del modelo Autobinomial y sus respectivas propiedades.
- En el **capítulo 5**, se desarrolla el Estudio Monte Carlo para analizar la sensibilidad de los estimadores clásicos de los parámetros en el modelo Autobinomial ante distintos grados de contaminación a diferentes patrones homogéneos de textura.
- En el **capítulo 6**, se presenta el problema de clasificación supervisada en imágenes satelitales, basándose en los dos modelos Autobinomiales considerados y en los métodos de estimación: Máxima Pseudo-verosimilitud y el de Mínimos Cuadrados Condicional.
- En el **capítulo 7**, se presentan las conclusiones y aportes del presente trabajo de Magíster. Se plantearán también las líneas de investigación futuras.

▪ APÉNDICES

A. Se desarrolla la matemática asociada con las técnicas de simulación de distribuciones de Gibbs y el método de maximización empleado para determinar el MP. No se han incluido las demostraciones de aquellos resultados que están desarrollados en la referencia indicada.

B. Se presentan los algoritmos que permiten obtener una realización de un Campo Aleatorio de Gibbs, el algoritmo diseñado para la contaminación de texturas y los que logran la estimación del vector de parámetros de cada uno de los modelos Autobinomiales considerados mediante el Método de MCC.

C. Se exhiben las estimaciones del vector de parámetros de los modelos considerados, Puros y Contaminados, logradas a partir de los estimadores MCC y MP.

Este trabajo ha dado lugar a la presentación de un póster: “Aplicación de La Regresión de Deming Ponderada en la Comparación de dos Métodos de Estimación de Parámetros del Modelo de Gibbs – Autobinomial”, en el XXXVII Coloquio Argentino de Estadística, realizado en la ciudad de San Miguel del Valle de Catamarca, del 7 al 9 de octubre de 2009.

CAMPOS ALEATORIOS DE MARKOV

En este capítulo se presenta una introducción a los Campos Aleatorios de Markov (CAM) con grilla finita; se describen en profundidad todos los conceptos más importantes concernientes a esta temática, junto con los resultados intermedios demostrados desde un punto de vista matemático. Se realiza una exhaustiva descripción de los modelos de CAM de textura de uso frecuente: el Modelo de Ising y el Autobinomial.

2.1 Conceptos generales

La teoría de Campos Aleatorios Markovianos originalmente surgió de la Física Estadística. Es una rama de la teoría de probabilidad que analiza las dependencias espaciales o contextuales de fenómenos físicos. Es una generalización natural de un proceso de Markov, en donde el índice del tiempo se reemplaza por un índice espacial.

Los campos aleatorios de Markov (CAM), entre otras aplicaciones, se utilizan para modelar imágenes con el propósito de extraer medidas de texturas. Diversos autores han puesto de manifiesto la capacidad de este modelo de CAM sobre un sistema de vecinos estándar para modelar y sintetizar microtexturas. Mediante él pueden describirse texturas con una fuerte aleatoriedad en la distribución de los niveles de gris que caracterizan una imagen.

Una imagen puede definirse a través de una grilla o malla rectangular bidimensional constituida de $m \times n$ píxeles uniformemente espaciados (Figura 2.1).

Sea \mathbf{S} el conjunto finito de índices que representan las posiciones de los píxeles, $\mathbf{S} = \{ (i, j) : i = 1, \dots, m; j = 1, \dots, n \}$. En general, el sitio del píxel se denota mediante la abreviatura $\mathbf{s} \equiv (i, j)$. El conjunto \mathbf{S} puede caracterizarse mediante una estructura matricial.

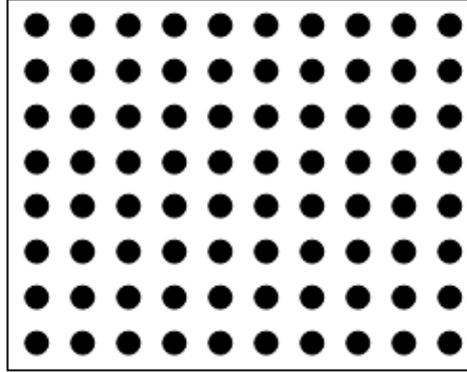


Figura 2.1 Ejemplo de una imagen. Arreglo rectangular de puntos (píxeles).

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_{11} & \mathbf{X}_{12} & \cdots & \mathbf{X}_{1n} \\ \mathbf{X}_{21} & \mathbf{X}_{21} & \cdots & \mathbf{X}_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ \mathbf{X}_{m1} & \mathbf{X}_{m2} & \cdots & \mathbf{X}_{mn} \end{bmatrix}$$

Figura 2.2 Notación matricial de una imagen. Cada x_{ij} representa un píxel de la imagen.

En cada píxel, ubicado en el sitio \mathbf{s} , está definida una variable aleatoria $X_{\mathbf{s}}$ (Figura 2.2). El conjunto completo de variables de todo el arreglo se denotará $\mathbf{X} = \{ X_{\mathbf{s}}, \mathbf{s} \in \mathbf{S} \}$. Se trabajará con campos aleatorios finitos, es decir, donde el conjunto de variables es finito. Cada variable $X_{\mathbf{s}}$ se le asocia un rango o espacio finito de estados $\Lambda_{\mathbf{s}} = \{ 0, \dots, M_{\mathbf{s}} \}$, que corresponde a los distintos niveles de gris (0 corresponde al negro, $M_{\mathbf{s}}$ al blanco y los valores entre ambos son tonos de gris). $x_{\mathbf{s}}$

denotará cada valor de gris del píxel \mathbf{s} . El producto $\Omega = \prod_{\mathbf{s} \in \mathbf{S}} \Lambda_{\mathbf{s}}$ es el espacio finito de

todas las configuraciones $\mathbf{x} = (x_{\mathbf{s}})_{\mathbf{s} \in \mathbf{S}}$ es decir, es el espacio imagen o textura. Se va

a considerar un espacio Ω homogéneo, todos los píxeles tienen igual rango de niveles de gris.

En el espacio de configuraciones Ω se define una medida de probabilidad o distribución de probabilidad. A cada configuración $\mathbf{x} \in \Omega$, se le asigna un valor $\Pi(\mathbf{x}) \geq 0$, tal que $\sum_{\mathbf{x} \in \Omega} \Pi(\mathbf{x}) = 1$.

Definición 2.1.1 Sea $(\mathbf{I}, \mathbf{S}, \Pi)$ un espacio de probabilidad. Una función \mathbf{X} definida sobre \mathbf{I} con valores en Ω se dirá que es un *campo aleatorio* si se verifica la condición de positividad, es decir, si $\Pi(\mathbf{X} = \mathbf{x}) > 0, \forall \mathbf{x} \in \Omega$.

Si la medida de probabilidad verifica la condición de positividad, significa que todas las configuraciones o texturas son posibles (con mayor o menor probabilidad).

La probabilidad conjunta se denotará $\Pi(\mathbf{x}) = \Pi(\mathbf{X} = \mathbf{x})$, donde $\mathbf{X} = (X_s)_{s \in \mathbf{S}}$ y $\mathbf{x} = (x_s)_{s \in \mathbf{S}}$.

2.2 Sistemas de vecinos y cliques

En \mathbf{S} existen interrelaciones entre los estados de ciertos píxeles. Estas dependencias van a ser, en general, locales. Esto significa que en una textura el estado de un píxel va a depender de los estados de los píxeles cercanos. Para cada posición $s \in \mathbf{S}$ se define un conjunto $\partial_s \subset \mathbf{S}$, cuyos elementos se denominan *vecinos de s*.

Definición 2.2.1 Un sistema de vecinos en el conjunto \mathbf{S} es la colección de conjuntos $\partial = \{ \partial_s, s \in \mathbf{S} \}$, que verifica:

$$1. s \notin \partial_s. \tag{2.1}$$

$$2. s \in \partial_t \Leftrightarrow t \in \partial_s. \tag{2.2}$$

$$3. \mathbf{S} = \bigcup_{s \in \mathbf{S}} \partial_s. \tag{2.3}$$

La primera, implica que un sitio no es vecino de si mismo, la segunda establece que la propiedad “ser vecino de” es simétrica, es decir, si \mathbf{s} es vecino de \mathbf{t} , éste último lo será del primero y la tercera, la unión de todos los vecinos conforman al conjunto \mathbf{S} .

En general, los sistemas de vecindario son *homogéneos*, significa que, conociendo los vecinos de una posición \mathbf{s} pueden determinarse los vecinos de otra posición \mathbf{t} sin más que desplazar a \mathbf{t} el sistema de vecinos de \mathbf{s} . Son invariantes en el espacio. El concepto de homogeneidad no es válido para los píxeles de borde; sin embargo, se va a seguir considerando al vecindario homogéneo, suponiendo implícitamente el efecto de bordes.

Los sistemas de vecinos son *isótropos*, si en todas las direcciones se comportan del mismo modo, es decir, son invariantes a rotaciones en las direcciones principales de la grilla.

Entre los elementos del conjunto \mathbf{S} puede definirse un *orden*, mediante el cual el conjunto de vecinos puede determinarse más explícitamente. Sean $\mathbf{s} = (i, j)$ y $\mathbf{t} = (k, l)$ dos píxeles de la imagen tal que $\mathbf{t} \in \partial_{\mathbf{s}}$, se define el *orden r del vecindario* (para el caso homogéneo e isótropo) como el menor entero que verifica:

$$r \geq d^2(\mathbf{s}, \mathbf{t}) = d^2((i, j), (k, l)) = (k - i)^2 + (l - j)^2 > 0, \quad (2.4)$$

para todos los vecinos $\mathbf{t} = (k, l)$ de $\mathbf{s} = (i, j)$. “ d ” representa la distancia Euclidea definida entre dos sitios \mathbf{t} y \mathbf{s} de la imagen.

El conjunto de vecinos de *orden r* del sitio \mathbf{s} corresponde a los píxeles próximos al sitio \mathbf{s} en un radio \sqrt{r} , es decir, es un conjunto constituido por todos los sitios (excepto el \mathbf{s}) que se encuentran a una distancia de a lo sumo \sqrt{r} .

$$\partial^{(r)}(\mathbf{s}) = \{ \mathbf{t} \in \mathbf{S} / 0 < d^2(\mathbf{s}, \mathbf{t}) \leq r \}.$$

El sistema de vecinos de primer orden del píxel del sitio $\mathbf{s} = (i, j)$, también llamado *sistema de vecinos cercanos* de \mathbf{s} , esta dado por el conjunto:

$$\partial^{(1)}(\mathbf{s}) = \{ \mathbf{t} \in \mathbf{S} / 0 < d^2(\mathbf{s}, \mathbf{t}) \leq 1 \} = \{ (i, j - 1), (i, j + 1), (i - 1, j), (i + 1, j) \}.$$

En la figura 2.2 se puede apreciar los sistemas de vecinos de los órdenes más comunes: 1, 2, 4, 5 y 8.

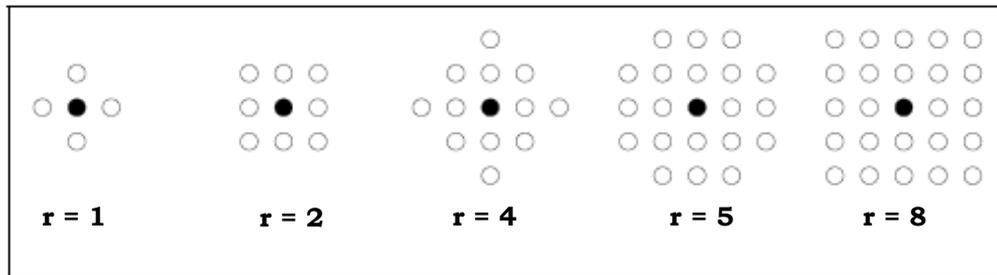


Figura 2.3 Vecindarios de órdenes 1, 2, 4, 5, 8 del sitio $s = \bullet$ y $t = \circ \in \partial(s)$.

En este tipo de sistemas de vecinos, el *orden del modelo* se define en función del número de vecinos que lo compongan.

Se ha asumido, en este trabajo, que los píxeles están uniformemente espaciados (la distancia horizontal y vertical entre dos píxeles adyacentes cualesquiera es la unidad); pero la expresión (2.4) es aplicable también a arreglos en que los sitios se encuentran irregularmente espaciados como muestra la figura 2.4.

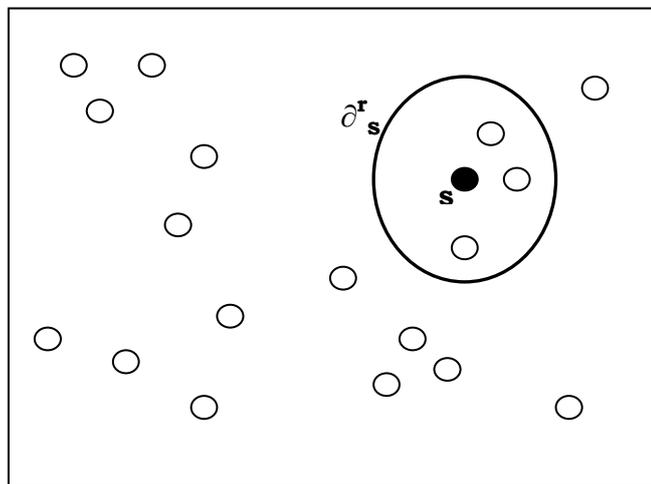


Figura 2.4 Arreglo de sitios irregularmente espaciados. Sistema de vecinos de orden r del píxel del sitio s .

El par $\{\mathbf{S}, \partial\}$ constituye un *grafo no dirigido*, donde el conjunto \mathbf{S} corresponde a los nodos y ∂ determina las relaciones entre los nodos de acuerdo con el sistema de vecinos definido. Todo grafo induce un sistema de vecindario.

Definición 2.2.2 Un *cliqué* C para la pareja $\{\mathbf{S}, \partial\}$ se define como un subconjunto C de \mathbf{S} en el que todo par de sitios son vecinos. Esto es, dados $s, t \in C$, $s \neq t$ implica $t \in \partial_s$. El subconjunto constituido por un sólo sitio es también un cliqué.

Al conjunto de todos los cliqués de \mathbf{S} se lo denotará \mathcal{C} , es decir, $\mathcal{C} = \bigcup_i C_i$, donde C_i

es el conjunto de todos los cliqués compuesto de i sitios, es decir, $C_i = \{\mathbf{s}_1, \dots, \mathbf{s}_i\}$.

El número y el tipo de cliqués que incluye a un cierto píxel \mathbf{s} esta vinculado a la estructura de vecinos considerada. Los cliqués asociados al sistema de vecinos cercanos del sitio $\mathbf{s} = (i, j)$, son los subconjuntos:

$$\{(i, j)\}; \{(i, j), (i, j+1)\}; \{(i, j), (i+1, j)\}; \{(i, j), (i, j-1)\} \text{ y } \{(i, j), (i-1, j)\}$$

Éstos se muestran en la figura 2.5 (a). El conjunto de cliqués, para el sistema de vecinos de segundo orden se muestran en la figura 2.5 (b).

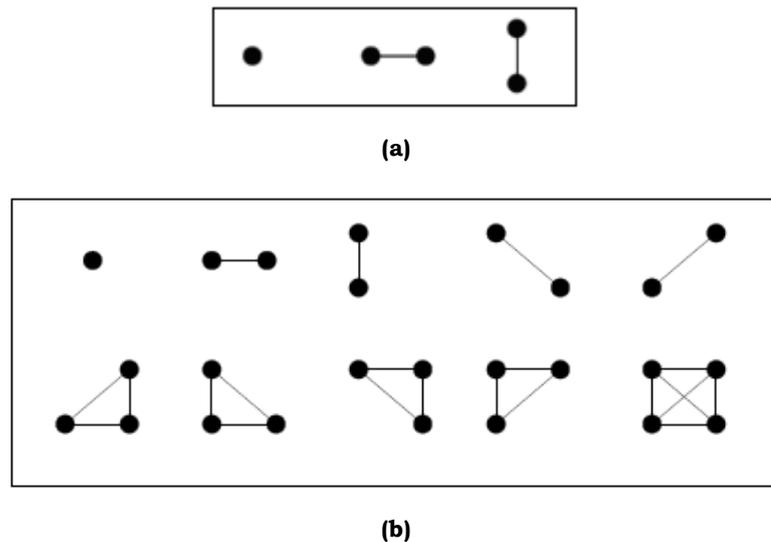


Figura 2.5 Tipos de cliqués: **(a)** para el sistema de vecinos de orden uno. **(b)** para sistemas de vecinos de segundo orden.

La variedad de tipos de cliqués crece exponencialmente a medida que aumenta el orden del sistema de vecinos.

Para un campo aleatorio puede definirse un tipo de probabilidad condicionada denominada *característica local del campo*, definida para $\mathbf{A} \subset \mathbf{S}$ como

$$\Pi_{\mathbf{A}}(\mathbf{x}) = \Pi(X_{\mathbf{A}} = \mathbf{x}_{\mathbf{A}} / X_{\mathbf{S}-\mathbf{A}} = \mathbf{x}_{\mathbf{S}-\mathbf{A}}).$$

A la familia de funciones $\{\Pi_{\mathbf{A}} / \mathbf{A} \subset \mathbf{S}\}$ se la denominará *familia de características locales asociadas a Π* .

Las características locales están bien definidas gracias a la propiedad de positividad que satisfacen los campos aleatorios. Si se considera, por ejemplo a X como cierta textura, representaría la probabilidad de que una subimagen de la textura tome un cierto valor, condicionada al resto de la textura.

Definición 2.2.3 Se dice que una familia de variables aleatorias $\mathbf{X} = \{X_s\}$ es un *Campo Aleatorio de Markov (CAM)* con respecto al sistema de vecinos ∂ si para todo $\mathbf{x} \in \Omega$ se satisfacen las siguientes dos propiedades:

1. $\Pi(\mathbf{x}) > 0 \forall \mathbf{x} \in \Omega.$ (2.5)

2. $\Pi(\mathbf{X}_s = \mathbf{x}_s / \mathbf{X}_t = \mathbf{x}_t, t \in \mathbf{S}, s \neq t) = \Pi(\mathbf{X}_s = \mathbf{x}_s / \mathbf{X}_t = \mathbf{x}_t, t \in \partial s).$ (2.6)

La condición 1. *Condición de positividad*, establece que $\mathbf{X} = \{X_s\}$ es un campo aleatorio. Esta propiedad refleja que cada instancia del campo aleatorio es probable. La condición 2., denominada *característica local*, expresa que la probabilidad de que el píxel X_s tome un nivel de gris x_s , condicionado a todos los niveles de gris $x_t, t \in \mathbf{S}, s \neq t$ de todos los píxeles restantes de la imagen, es equivalente a aquella probabilidad condicionada sólo a los valores de los niveles de gris x_t de la vecindad del píxel s . Esta última condición, también es llamada, *propiedad markoviana*.

Luego, para un campo aleatorio de Markov la *característica local del campo* puede definirse

$$\Pi_{\{s\}}(\mathbf{x}) = \Pi(\mathbf{X}_s = \mathbf{x}_s / \mathbf{X}_t = \mathbf{x}_t, t \in \mathbf{S}, s \neq t) = \Pi(\mathbf{X}_s = \mathbf{x}_s / \mathbf{X}_t = \mathbf{x}_t, t \in \partial s).$$

A la familia de funciones $\{\Pi_{\{s\}} / s \in \mathbf{S}\}$ se la llamará *familia de características locales asociadas a Π* .

Existen dos enfoques para caracterizar un Campo Aleatorio de Markov, en términos de probabilidades condicionales $\Pi_{\{s\}}(\mathbf{x}) = \Pi(x_s / x_t \in \partial s)$ o en términos de probabilidad conjunta $\Pi(\mathbf{x})$. Besag (1974) da a conocer las desventajas que surgen al utilizar la

aproximación mediante la probabilidad condicional respecto al uso de la aproximación mediante la probabilidad conjunta.

Pero estas dificultades pueden subsanarse gracias a un resultado teórico debido a Hammersley y Clifford. Este resultado establece la equivalencia entre campos aleatorios Markovianos y distribuciones Gibbsianas, que proporciona medios teóricos adecuados para calcular la probabilidad conjunta de un campo aleatorio Markoviano.

2.3 Campos Aleatorios de Gibbs (CAG)

Definición 2.3.1 Se dice que una familia de variables aleatorias $\mathbf{X} = \{X_S\}$ es un *Campo Aleatorio de Gibbs (CAG)* en \mathbf{S} con respecto al sistema ∂ si sus configuraciones siguen una distribución de Gibbs, es decir,

$$\Pi(\mathbf{x}) = \frac{1}{Z} e^{-H(\mathbf{x})}, \quad (2.7)$$

donde $Z = \sum_{z \in \Omega} e^{-H(z)}$ es la constante de normalización conocida como función de partición, y $H(\mathbf{x})$ es la función de energía que induce el CAG.

Proposición 2.3.1 Todo campo aleatorio Π es inducido por la función de energía

$$\mathbf{H}(\mathbf{x}) = -\ln \Pi(\mathbf{x}) - K, \quad (2.8)$$

siendo $K = \ln(Z)$.

Para que la función de energía por medio de Π sea única se fija una *configuración de referencia* o *de vacío*, $\mathbf{o} \in \Omega$, para la que $\mathbf{H}(\mathbf{o}) = 0$. En este caso,

$$-\ln \Pi(\mathbf{o}) - K = 0.$$

Luego $K = -\ln \Pi(\mathbf{o})$, lo que implica que $Z = \Pi(\mathbf{o})^{-1}$.

Resulta conveniente descomponer la función de energía en distintas contribuciones de las configuraciones de subconjuntos \mathbf{A} de \mathbf{S} .

Definición 2.3.2 Sea \mathbf{A} un subconjunto cualquiera de \mathbf{S} , el espacio de configuraciones $\mathbf{x}_A = \{x_s\}_{s \in A}$ en \mathbf{A} , se denotará $\Omega_A = \prod_{s \in A} \Lambda_s$. La aplicación:

$$X_A: \Omega \rightarrow \Omega_A,$$

que a cada configuración $x = \{x_s\}_{s \in \mathbf{S}}$ le hace corresponder una configuración $\mathbf{x}_A = \{x_s\}_{s \in A}$, se denomina *proyección* de Ω en Ω_A .

En adelante, $\{X_A = \mathbf{x}_A\}$ denotará $\{x \in \Omega : X_A(x) = \mathbf{x}_A\}$.

Definición 2.3.2 Un potencial \mathcal{V} es una familia $\{V_A / V_A: \Omega \rightarrow \mathfrak{R}, \mathbf{A} \subset \mathbf{S}\}$ de funciones que verifican:

1. $V_{\emptyset}(x) = 0$
2. $V_A(x) = V_A(y)$ si $x_A = y_A$

La *condición 1.* de la **Definición 2.3.2** significa que la función potencial para el conjunto vacío es la función nula y la *2.* que la función potencia V_A sólo va a depender de la configuración en $\mathbf{A} \subset \mathbf{S}$.

A cada función elemento de la familia \mathcal{V} se la denomina *función potencial*.

La energía del potencial se define:

$$\mathbf{H}_{\mathcal{V}}(x) = \sum_{\mathbf{A} \subset \mathbf{S}} V_{\mathbf{A}}(x). \quad (2.9)$$

Un campo Aleatorio de Gibbs viene inducido por el potencial \mathcal{V} , siempre que la energía del campo $\mathbf{H}_{\mathcal{V}}(x)$ se defina como muestra la expresión (2.9), es decir, sea suma de funciones potenciales. La forma de estas funciones puede ser arbitraria.

En caso que las dependencias son locales y dadas por un sistema de vecinos δ , el número de funciones potenciales no nulas se reduce notablemente. De todos los subconjuntos \mathbf{A} de \mathbf{S} , sólo es necesario considerar los dados por el conjunto de

¹ \emptyset representa el conjunto vacío

cliqués \mathcal{C} , lo que simplifica la definición de un CAG a través de funciones potenciales.

Definición 2.3.3 Dado un sistema de vecinos ∂ un potencial \mathcal{V} se dice que es un *potencial de vecindario* con respecto a ∂ si $V_{\mathbf{A}} = 0$ siempre que \mathbf{A} no sea un cliqué para ese sistema de vecinos, es decir, siempre que $\mathbf{A} \notin \mathcal{C}$. Si las funciones potenciales $V_{\mathbf{A}}$ son nulas para, es decir, para conjuntos de más de dos elementos, el potencial \mathcal{V} se denomina *potencial de pares*.

La energía del campo $\mathbf{H}(\mathbf{x})$, se define como suma de funciones potenciales no nulas, $V_c(\mathbf{x})$, de los cliqués c , sobre el conjunto de todos los cliqués posibles \mathcal{C} . El valor de $V_c(\mathbf{x})$ depende de la configuración local del cliqué.

Definición 2.3.4 Un campo aleatorio Π es un *Campo Aleatorio de Gibbs para el potencial \mathcal{V}* , si la función de energía que induce el campo es la energía del potencial $\mathbf{H}_{\mathcal{V}}(\mathbf{x})$. Si \mathcal{V} es un potencial de vecindario con respecto a ∂ , entonces Π es un CAG de vecindario para el potencial \mathcal{V} con respecto a ∂ .

Proposición 2.3.2 Sea Π es un campo aleatorio para algún potencial de vecindario \mathcal{V} para el sistema de vecinos ∂ , entonces

$$\Pi(\mathbf{x}) = \frac{e^{-\sum_{c \in \mathcal{C}} V_c(\mathbf{x})}}{\sum_{y \in \Omega} e^{-\sum_{c \in \mathcal{C}} V_c(y)}}, \quad (2.10)$$

siendo \mathcal{C} el conjunto de cliqués para sistema de vecinos ∂ . Luego las características locales están dadas por:

$$\Pi(X_s = x_s, s \in \mathbf{A} / X_s = x_s, s \in \mathbf{S} - \mathbf{A}) = \frac{\exp\left(-\sum_{c \in \mathcal{C}, c \cap \mathbf{A} \neq \emptyset} V_c(\mathbf{x})\right)}{\sum_{y_{\mathbf{A}} \in \Omega_{\mathbf{A}}} \exp\left(-\sum_{c \in \mathcal{C}, c \cap \mathbf{A} \neq \emptyset} V_c(y_{\mathbf{A}} x_{\mathbf{S}-\mathbf{A}})\right)} \quad (2.11)$$

donde $\mathbf{x} = y_{\mathbf{A}} x_{\mathbf{S}-\mathbf{A}} \in \Omega_{\mathbf{A}} \times \Omega_{\mathbf{S}-\mathbf{A}}$, $y_{\mathbf{A}} \in \Omega_{\mathbf{A}}$ y $x_{\mathbf{S}-\mathbf{A}} \in \Omega_{\mathbf{S}-\mathbf{A}}$.

Además, puede extenderse la expresión (2.11) para las características locales de cualquier subconjunto $\mathbf{A} \subset \mathbf{S}$.

$$\prod (X_S = x_S, s \in \mathbf{A} / X_S = x_S, s \in \mathbf{S} - \mathbf{A}) = \prod (X_{\mathbf{A}} = \mathbf{x}_{\mathbf{A}} / X_S = x_S, s \in \partial_{\mathbf{A}}), \quad (2.12)$$

donde $\partial_{\mathbf{A}} = \left[\bigcup_{s \in \mathbf{A}} \partial_s \right] - \mathbf{A}$, representa el vecindario para cada subconjunto $\mathbf{A} \subset \mathbf{S}$.

En particular, \prod es un CAM con respecto a ∂ .

Si $\mathbf{o} \in \Omega$ es la configuración de referencia, \mathbf{A} un subconjunto cualquiera de \mathbf{S} y $\mathbf{x} \in \Omega$ otra configuración, se denotará con $\mathbf{x}^{\mathbf{A}}$ la configuración que coincide con \mathbf{x} en \mathbf{A} y con \mathbf{o} fuera de \mathbf{A} . Es decir,

$$\mathbf{x}^{\mathbf{A}} = \left\{ \mathbf{x}_s^{\mathbf{A}}, s \in \mathbf{S} \right\}, \quad \mathbf{x}_s^{\mathbf{A}} = \begin{cases} \mathbf{x}_s & s \in \mathbf{A} \\ \mathbf{o} & s \notin \mathbf{A} \end{cases}$$

Definición 2.3.5 Un potencial \mathbf{V} tal que $V_{\mathbf{A}}(\mathbf{x}) = 0$ siempre que $\mathbf{x}_s = \mathbf{o}_s$, para algún $s \in \mathbf{A}$, se denomina *potencial normalizado*.

Teorema 2.3.1 Todo campo aleatorio \prod es un Campo Aleatorio de Gibbs para algún potencial. Puede elegirse un potencial \mathbf{V} con $V_{\emptyset}(\mathbf{x}) = 0$ y para $\mathbf{A} \neq \emptyset$ dado por

$$V_{\mathbf{A}}(\mathbf{x}) = - \sum_{\mathbf{B} \subset \mathbf{A}} (-1)^{|\mathbf{A}-\mathbf{B}|} \ln(\prod(\mathbf{x}^{\mathbf{B}})), \quad \forall \mathbf{x} \in \Omega, s \in \mathbf{S}. \quad (2.13)$$

Para todo subconjunto \mathbf{A} de \mathbf{S} y todo $a \in \mathbf{A}$,

$$V_{\mathbf{A}}(\mathbf{x}) = - \sum_{\mathbf{B} \subset \mathbf{A}} (-1)^{|\mathbf{A}-\mathbf{B}|} \ln(\prod(X_a = \mathbf{x}_a^{\mathbf{B}} / X_s = \mathbf{x}_s^{\mathbf{B}}, s \neq a)), \quad \forall \mathbf{x} \in \Omega, s \in \mathbf{S}, a \in \mathbf{A} \quad (2.14)$$

Se verifica además que, el potencial \mathbf{V} es el único potencial normalizado para \prod .

El **Teorema 2.3.1** cuya demostración se debe a G. R. Grimmett [40], se basa en la fórmula de inversión de Möbius [75] presentada a través del siguiente **Lema 2.3.1**.

$$\varphi(\mathbf{A}) = - \sum_{B \subset \mathbf{A}} (-1)^{|\mathbf{A}-B|} \psi(B), \quad \forall \mathbf{A} \subset \mathbf{S} \quad \text{si y sólo si} \quad \psi(\mathbf{A}) = \sum_{B \subset \mathbf{A}} \varphi(B), \quad \forall \mathbf{A} \subset \mathbf{S}$$

Demostración. Para la demostración del **Teorema 2.3.1** basta con probar que:

$$\varphi(\mathbf{A}) = - \sum_{B \subset \mathbf{A}} (-1)^{|\mathbf{A}-B|} \psi(B), \quad \forall \mathbf{A} \subset \mathbf{S} \quad \Rightarrow \quad \psi(\mathbf{A}) = \sum_{B \subset \mathbf{A}} \varphi(B), \quad \forall \mathbf{A} \subset \mathbf{S}$$

La fórmula de la derecha puede escribirse de la siguiente forma

$$\begin{aligned} \sum_{B \subset \mathbf{A}} \varphi(B) &= \sum_{B \subset \mathbf{A}} \sum_{D \subset B} (-1)^{|B-D|} \psi(D) \\ &= \sum_{D \subset \mathbf{A}} \sum_{C \subset \mathbf{A}-D} (-1)^{|C|} \psi(D) \\ &= \sum_{D \subset \mathbf{A}} \psi(D) \sum_{C \subset \mathbf{A}-D} (-1)^{|C|} = \psi(\mathbf{A}). \end{aligned}$$

Si se considera la suma interna de la última ecuación, se puede observar que si $\mathbf{A} - D = \emptyset$ vale 1 y si $\mathbf{A} - D \neq \emptyset$ entonces sea $m = |\mathbf{A} - D|$,

$$\begin{aligned} \sum_{C \subset \mathbf{A}-D} (-1)^{|C|} &= \sum_{i=1}^m \left| \{C \subset \mathbf{A}-D : |C|=i\} \right| (-1)^i \\ &= \sum_{i=1}^m \binom{m}{i} (-1)^i = (1-1)^m = 0. \end{aligned}$$

Si se supone que se verifica la segunda condición y se usa el mismo argumento que se empleó para la anterior, se tiene que

$$\begin{aligned} \sum_{B \subset \mathbf{A}} (-1)^{|\mathbf{A}-B|} \psi(B) &= \sum_{D \subset B \subset \mathbf{A}} (-1)^{|\mathbf{A}-B|} \psi(D) \\ &= \sum_{D \subset \mathbf{A}} \psi(D) \sum_{C \subset \mathbf{A}-D} (-1)^{|C|} = \varphi(\mathbf{A}). \end{aligned}$$

Lo que completa la demostración del lema. ♦

Corolario 2.3.1 Un CAG queda unívocamente determinado por sus características locales para los conjuntos unitarios $\{s\} \subset \mathbf{S}$, es decir, por

$$\prod (X_s = x_s / X_t = x_t, t \in \mathbf{S} - \{s\}).$$

Para el vecindario $\partial_s = \mathbf{S} - \{s\}$, $\forall s \in \mathbf{S}$ un campo aleatorio Π es un Campo Aleatorio de Markov. Este CAM es un CAG de vecindario con respecto a ∂_s . Esto también se verifica para cualquier otro sistema de vecinos, $\partial_{\mathbf{A}} = \bigcup_{s \in \mathbf{A}} \partial_s - \mathbf{A}$, con $\mathbf{A} \subset \mathbf{S}$, como lo demuestra el teorema que se presenta en la siguiente sección, conocido como Teorema de Equivalencia o Teorema de Hammersley y Clifford.

2.4 Equivalencia entre CAM y CAG

El teorema de Hammersley y Clifford, establece la equivalencia entre estos dos tipos de campos aleatorios, uno caracterizado por su propiedad local (la Markovianidad), y el otro, caracterizado por su propiedad global (la distribución Gibbsiana). Debido a este resultado teórico un Campo Aleatorio de Markov puede ser descrito como un Campo Aleatorio de Gibbs.

La importancia del teorema radica en que provee una fórmula explícita para la distribución de probabilidad conjunta en términos de la función de energía, especificando las funciones potenciales junto con el sistema de vecinos.

La versión más antigua del teorema de equivalencia es de Hammersley y Clifford (1968) pero existen otras publicaciones al respecto que datan de principios de los 70, como Grimmett (1975, basada en el teorema de inversión de Möbius) y Averintsev (1978) y de la década de los 80, como Georgii (1988).

Teorema 2.4.1 (Hammersley - Clifford)

Sea ∂ un sistema de vecindario dado en \mathbf{S} . Entonces se cumple siempre que:

- (a) Un Campo Aleatorio es un Campo Aleatorio de Markov para $\partial \Leftrightarrow$ es un Campo Aleatorio de Gibbs para ∂ .
- (b) Para un Campo Aleatorio de Markov Π con un sistema de vecinos ∂ ,

$$\prod (X_s = x_s , s \in \mathbf{A} / X_s = x_s , s \in \mathbf{S} - \mathbf{A}) = \prod (X_s = x_s , s \in \mathbf{A} / X_s = x_s , s \in \partial \mathbf{A}),$$

$$\forall \mathbf{A} \subset \mathbf{S}.$$

Demostración. Por medio de la **proposición 2.3.2** se deduce que un Campo Aleatorio Gibbs para ∂ es un Campo Aleatorio Markoviano para ∂ , como así también la implicación (b).

Basta probar entonces que un Campo Aleatorio Markoviano con respecto al vecindario ∂ es un Campo Aleatorio de Gibbs para ∂ . Para demostrarlo, es necesaria la unicidad de representación de un Campo Aleatorio de Gibbs, es decir, que la elección de funciones de potenciales para un CAM específico sea única. Debido a que varios potenciales diferentes pueden definir la misma distribución de Gibbs, el potencial se normaliza respecto a una configuración determinada, denominada configuración de *referencia* o *de vacío*, para la que todas las funciones potenciales son nulas. De esta manera la relación entre un CAG y un potencial normalizado es unívoca.

En la demostración se adoptará la notación $B + a$ para expresar $B \cup \{a\}$.

Sea Π un Campo Aleatorio de Markov respecto al vecindario ∂ y sea V un potencial tal que:

$$V_{\mathbf{A}}(\mathbf{x}) = \begin{cases} 0 & \text{si } \mathbf{A} = \emptyset \\ - \sum_{B \subset \mathbf{A}} (-1)^{|\mathbf{A}-B|} \ln(\Pi(\mathbf{x}^B)) & \text{si } \mathbf{A} \neq \emptyset \end{cases}$$

Basta mostrar que $V_{\mathbf{A}}$ es nulo cuando \mathbf{A} no es un cliqué. Supóngase entonces que \mathbf{A} no es un cliqué, entonces existe un $a \in \mathbf{A}$ y $b \in \mathbf{A} - \partial a$. Usando la expresión:

$$V_{\mathbf{A}}(\mathbf{x}) = - \sum_{B \subset \mathbf{A} - a} (-1)^{|\mathbf{A}-B|} [\ln(\Pi(\mathbf{x}^B)) - \ln(\Pi(\mathbf{x}^{B+a}))] \quad (2.15)$$

se describe la suma presentada en (2.14) de la siguiente forma:

$$\mathbf{V}_{\mathbf{A}}(\mathbf{x}) = - \sum_{B \subset \mathbf{A} / \{a, b\}} (-1)^{|\mathbf{A}-B|} * \ln \left[\frac{\prod (X_a = B_{x_a} / X_s = B_{x_s}, s \neq a)}{\prod (X_a = B+b_{x_a} / X_s = B+b_{x_s}, s \neq a)} \right] * \left[\frac{\prod (X_a = x_a^{B+a+b} / X_s = x_s^{B+a+b}, s \neq a)}{\prod (X_a = x_a^{B+a} / X_s = x_s^{B+a}, s \neq a)} \right].$$

Si se considera la primera fracción dentro del argumento del logaritmo natural, como $a \neq b$ entonces $\left\{ X_a = x_a^B \right\} = \left\{ X_a = x_a^{B+b} \right\}$. Además, como $b \notin \partial a$, por definición de CAM, el numerador y denominador de dicha fracción coinciden. Se aplica el mismo argumento a la segunda fracción. En consecuencia, como el logaritmo de 1 es cero, se anulan todos los términos de la suma, lo que implica que $\mathbf{V}_{\mathbf{A}} = 0$, cuando \mathbf{A} no es un cliqué.

Esto completa la demostración del teorema. ♦

Debido a la equivalencia establecida entre un Campo Aleatorio de Markov con respecto al vecindario ∂ y un Campo Aleatorio de Gibbs para ∂ , con función de energía $\mathbf{H}(\mathbf{x}_S, \partial \mathbf{x}_S; \boldsymbol{\theta})$, donde $\boldsymbol{\theta}$ es el vector de parámetros ¹, la ecuación (2.6) puede expresarse:

$$\begin{aligned} \prod_{\{s\}}(\mathbf{x}) &= \prod (X_S = \mathbf{x}_S / X_t = \mathbf{x}_t, t \in \mathbf{S}, s \neq t) \\ &= \prod (X_S = \mathbf{x}_S / X_t = \mathbf{x}_t, t \in \partial s) \\ &= \prod (\mathbf{x}_S / \partial s; \boldsymbol{\theta}) = \frac{1}{Z_s} e^{-\mathbf{H}(\mathbf{x}_S, \partial s; \boldsymbol{\theta})}. \end{aligned} \tag{2.16}$$

La función de distribución de probabilidad condicional de \mathbf{x}_S se determina entonces, a partir de la función de energía $\mathbf{H}(\mathbf{x}_S, \partial s; \boldsymbol{\theta})$.

Con lo cual,

¹El vector de parámetros $\boldsymbol{\theta}$ representa las correlaciones entre el píxel \mathbf{x}_S y sus vecinos $\partial \mathbf{x}_S$.

La función de distribución de probabilidad condicional de \mathbf{x}_s se determina entonces, a partir de la función de energía $\mathbf{H}(\mathbf{x}_s, \partial_s; \boldsymbol{\theta})$.

Con lo cual,

$$\Pi(\mathbf{x}_s/\partial_s; \boldsymbol{\theta}) = \frac{e^{-\sum_{C \in C_s} V_C(\mathbf{x})}}{\sum_{z_s \in \Lambda_s} e^{-\sum_{C \in C_s} V_C(z_s \mathbf{x}_{C_s - s})}}, \quad (2.17)$$

donde $C_s = \{C \in \mathcal{C}, s \in C\}$, es decir, es el conjunto de cliques que contienen a s y Λ_s representa el conjunto de valores de la posición s .

Un conjunto importante para las aplicaciones lo constituyen las distribuciones de Gibbs con potenciales nulos fuera de cliques con tres o más elementos, los *potenciales de pares*. Esto es:

$$V_C(\mathbf{x}) = \sum_{s \in \mathbf{S}} V_1(\mathbf{x}_s) + \sum_{s \in \mathbf{S}} \sum_{t \in \partial(s)} V_2(\mathbf{x}_s, \mathbf{x}_t). \quad (2.18)$$

Considerando la segunda sumatoria de la ecuación (2.18), y siendo $\{i, j\}$ y $\{j, i\}$ dos cliques diferentes en C_2 (puesto que los sitios de los cliques están ordenados), la probabilidad condicional puede escribirse como:

$$\Pi_{\{s\}}(\mathbf{x}) = \Pi(\mathbf{x}_s/\partial_s; \boldsymbol{\theta}) = \frac{e^{-V_1(\mathbf{x}_s) + \sum_{t \in \partial_s} V_2(\mathbf{x}_s, \mathbf{x}_t)}}{\sum_{\mathbf{x}_s \in \Lambda_s} e^{-V_1(\mathbf{x}_s) + \sum_{t \in \partial_s} V_2(\mathbf{x}_s, \mathbf{x}_t)}}. \quad (2.19)$$

2.5 Modelos de textura

Existe una gran diversidad de cualidades esenciales que definen una *textura*, por tal motivo, que no existe una única metodología simple y unificada que permita representarla y analizarla. En la literatura las vías más populares sobre la temática pueden resumirse en: *Métodos basados en medidas estadísticas* (como matrices de concurrencia del nivel de gris, en inglés, GLCM), *Métodos basados en modelos*

estadísticos o geométricos, y métodos *basados en técnicas de Procesamiento de Señales*. Debido al objetivo fijado en esta tesis, la atención se centrará en los métodos basados en modelos estocásticos que persiguen tanto la descripción de texturas observadas como la síntesis de texturas artificiales. Entre los métodos basados en modelos, hay algunos que son válidos para caracterizar la textura en vecindades pequeñas en una única escala, es decir, sólo pueden utilizarse de manera eficiente para el análisis de micro-texturas [6] y otros tipos que son capaces de caracterizar de manera eficiente texturas a diferentes escalas, combinando características espaciales y frecuenciales de la imagen de textura, macrotexturas.

Según indican ciertos autores, la textura puede considerarse como un fenómeno de la distribución espacial de los niveles de gris sobre una determinada área y su análisis puede descomponerse en dos niveles: el primero relacionado con las propiedades locales, también denominadas *primitivas*, que constituyen la textura y el segundo con la extracción de la relación espacial entre ellas.

Las primitivas están compuestas por regiones de una imagen y se caracterizan por las propiedades relativas a la distribución de los niveles de gris en las regiones que la componen y por las medidas morfológicas de las mismas (área, perímetro, forma, etc). En consecuencia, una imagen de textura puede ser descrita en términos del tipo, número y organización espacial de sus primitivas. Esta organización espacial puede ser aleatoria, dependiente de primitivas vecinas o dependiente del tiempo.

Cuando en una textura el tamaño de las primitivas es pequeño y la interacción espacial está limitada localmente, ésta se denomina *micro-textura*. En caso contrario, cuando su tamaño y/o su interacción se extienden, se dice que se trata de *macro-textura*.

En las secciones siguientes se desarrollarán en detalle algunos modelos de Campos Aleatorios de Markov que permiten generar y representar texturas de una imagen. En estos modelos, los parámetros caracterizan los dos aspectos mencionados anteriormente: especifican las propiedades locales que determinan los niveles de gris que forman las primitivas y la relación existente entre los niveles de gris de las primitivas que forman la imagen. Se trata de los denominados *Auto-modelos* (Besag, 1974), son los más sencillos entre los modelos Aleatorios de Markov. Existen numerosas publicaciones científicas respecto este tipo de modelos [1-6-11-14-17-18-20-43-47-54-62-76-77].

Se asumen, principalmente, sistemas de vecinos de segundo orden teniendo en cuenta que para órdenes superiores aumenta considerablemente el volumen de cálculos y no produce prácticamente rédito alguno en variedad de textura.

2.5.1 Modelo de Ising

El Modelo de Ising es uno de los modelos que constituyó el punto de partida para el estudio de numerosas cuestiones acerca de los Campos Aleatorios de Markov [35-54]. Su nombre se debe a Ernest Ising [46], quien a principios de 1925 escribió su tesis doctoral referida a un nuevo modelo que trataba de explicar teóricamente uno de los fenómenos más notables en la física del estado sólido, el *ferromagnetismo*. Ising deseaba a través del mismo estudiar el comportamiento de materiales [ferromagnéticos](#), específicamente, simular la estructura de una porción de material ferromagnético. Intentaba demostrar que el sistema presentaba una transición de fase.

El ferromagnetismo es uno de los fenómenos que más ha interesado a los físicos durante el siglo XX, tanto por su complejidad teórica, como por el amplio campo de aplicaciones que posee dentro de la técnica moderna.

La principal virtud de este modelo, reside en el hecho que, aplicado a una red bidimensional sin campo externo [65] y con campo magnético [87] permite describir transiciones de fase con total solidez matemática (usando la matriz de transferencia), consolidando así definitivamente la Mecánica Estadística. Otros autores utilizaron diferentes técnicas de cálculo para resolver este modelo [81].

Ising demostró que el modelo unidimensional no presentaba tal transición de fase. Peierls, en 1936, fue el precursor en probar que el modelo de Ising de dos o más dimensiones presentaba transición de fase a temperaturas no nulas [71]. Del modelo en tres dimensiones no se han encontrado soluciones exactas, a pesar de la intensa búsqueda. Así mismo, tampoco se conocen soluciones exactas del modelo de Ising bidimensional cuando el campo externo aplicado es distinto de cero.

Este modelo fue aplicado, además, a otros problemas físicos para reproducir sistemas como la red de gas y aleaciones binarias, y a otros campos como la biología (para modelizar macromoléculas, la desnaturalización de ADN, y en redes

neuronales), la química, la economía y la sociología (para modelizar el aislamiento cultural).

El modelo de Ising consiste en tratar al sólido como un sistema compuesto de un conjunto finito, N , de puntos fijos, espaciados en forma regular, que constituyen un entramado D - dimensional.

En una dimensión ($D = 1$), se considera una sucesión de puntos, $1, 2, 3, \dots, N$ dispuestos en una línea.



Figura 2.6 Entramado en una dimensión.

Si $D = 2$, los N puntos están dispuestos sobre una red en un plano como se muestra en la figura 2.7.

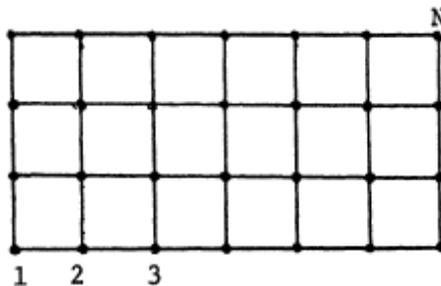


Figura 2.7 Entramado bidimensional.

En el modelo de dimensión 3, los N puntos están ubicados en los nudos de una red cúbica simple. Las unidades, en este caso, son cubos.

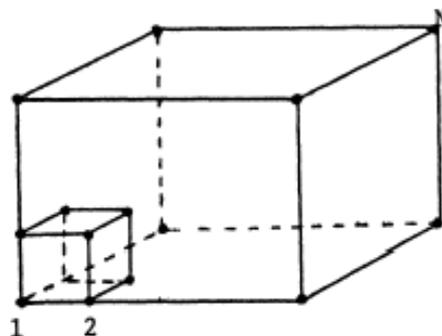


Figura 2.8 Entramado tridimensional.

En este caso se considera el modelo de Ising Bidimensional. Éste consiste, como se mencionó anteriormente, en un sistema de N puntos de un plano que forman un arreglo cuadrado (en inglés, lattice). Sea $\mathbf{S} = \{ (i, j) : i, j = 1, \dots, m \}$ el conjunto finito de índices que representan las posiciones de los N puntos. En general, el sitio del punto se denota mediante la abreviatura $\mathbf{s} \equiv (i, j)$. Cada punto o sitio posee un pequeño *dipolo* o *espin*¹, que puede encontrarse, en un momento dado, en una de dos posibles posiciones: “up²” (+) o “down³” (-).

Si se define al espacio Ω como espacio muestral formado por todas las secuencias $w = \{ w_1, w_2, \dots, w_N \}$ donde $w_s = +$ indica un espin up y $w_s = -$ indica un espin down, puede pensarse a cada espin como una función definida sobre Ω tal que:

$$X_s(w) = \begin{cases} 1 & \text{si } w_s = + \\ -1 & \text{si } w_s = - \end{cases}$$

Luego, a cada punto o sitio de la red hay asociada una variable de espin X_s , que puede tomar únicamente dos valores numéricos: +1 (espin up) y -1 (espin down).

Definición 2.5.1 La asignación de valores $\mathbf{x} = \{ x_1, x_2, \dots, x_N \}$ a las N variables espin del arreglo se denomina una *configuración* del sistema. Existen 2^N configuraciones distintas.

La figura 2.9 muestra un ejemplo de una configuración de los espines

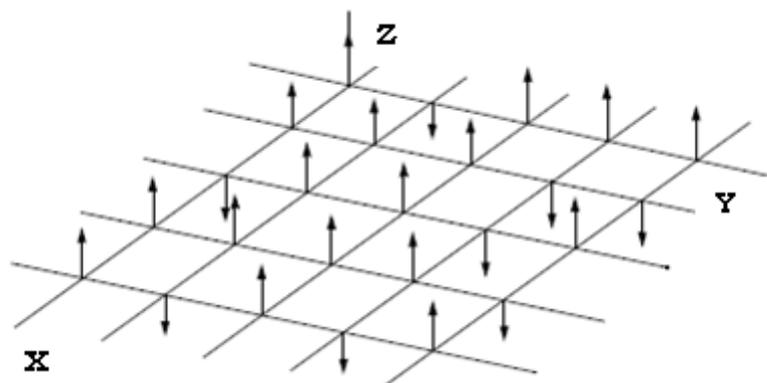


Figura 2.9 Una configuración de los espines. Los espines orientados en la dirección positiva del eje Z corresponden al valor 1 de las variables X_s y los

¹ Momento magnético

² Hacia arriba

³ Hacia abajo

orientados en la dirección negativa de Z, a los valores -1 de las variables X_s .

Definición 2.5.2 Dos sitios $\mathbf{s} = (i, j)$ y $\mathbf{t} = (k, f)$ se dirán *vecinos cercanos* si verifican:

$$i = k \text{ y } |j - f| = 1 \quad \text{ó} \quad j = f \text{ y } |i - k| = 1.$$

Se denotarán: $\mathbf{s} \sim \mathbf{t}$.

Luego, $\partial_{\mathbf{s}} = \{ \mathbf{t} \in \mathbf{S} / \mathbf{s} \sim \mathbf{t} \}$ y $\partial = \{ \partial_{\mathbf{s}}, \mathbf{s} \in \mathbf{S} \}$ es el sistema de vecindades sobre \mathbf{S} .

Excepto los sitios que pertenecen al borde del arreglo bidimensional, cada sitio posee 4 vecinos cercanos (Figura 2.10).

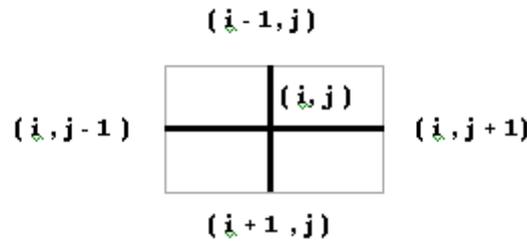


Figura 2.10 Sistema de vecinos cercanos del sitio $\mathbf{s} = (i, j)$.

Definición 2.5.2 Cada segmento que une los sitios vecinos del entramado se denomina “bond” (en castellano, cadena).

La energía $\mathbf{H}(\mathbf{x})^4$, asociada a cada configuración $\mathbf{x} = \{ x_1, x_2, \dots, x_N \}$ se define:

$$\mathbf{H}(\mathbf{x}) = \mathbf{H}(\{ x_1, x_2, \dots, x_N \}) = -\mathbf{J} \sum_{s=1}^N \sum_{s \sim j} x_s x_j - \mathbf{B} \mathbf{m} \sum_{s=1}^N x_s, \quad (2.20)$$

donde \mathbf{J} y \mathbf{B} son parámetros que representan las “energías” asociadas con las interacciones entre los vecinos cercanos y la interacción de un campo magnético externo, respectivamente. La primera suma se efectúa sobre todos los pares de vecinos cercanos (computados una sola vez). En el modelo de ferromagnetismo, la constante \mathbf{J} representa una propiedad del material considerado. Si \mathbf{J} es positiva ($\mathbf{J} > 0$) los espines se acoplan en orientaciones paralelas, es decir, la interacción

⁴ H es el hamiltoniano de Heisenberg que constituye la energía total del sistema y gobierna la dinámica del mismo.

provoca una alineación de los espines vecinos, lo que contribuye a un bajo nivel de energía (se favorecen los productos $x_i x_j = +1$). En este caso se trata de una fase *ferromagnética*. Si \mathbf{J} es negativa ($\mathbf{J} < 0$) los espines tienden a alinearse entre si en sentido antiparalelo (se favorecen los productos $x_i x_j = -1$), y se está en presencia de una fase *antiferromagnético*. El segundo término representa el efecto de un campo magnético externo de intensidad \mathbf{B} . La constante $\mathbf{m} > 0$ depende del material considerado. Este término minimiza la energía cuando todos los espines tienen la misma dirección que el campo externo. Para más detalle ver Kindermann and Snell [55].

La energía de una configuración viene dada por la expresión (2.20), si se considera la interacción homogénea ($\mathbf{B}_i = \mathbf{B} > 0 \quad \forall i$) e isótropa ($\mathbf{J}_{ij} = \mathbf{J} \quad \forall i, j$), donde el acoplamiento entre espines es independiente de las posiciones en la red.

En caso contrario se define:

$$\mathbf{H}(\mathbf{x}) = - \sum_{s=1}^N \sum_{j \in \partial_s} \mathbf{J}_{ij} x_s x_j - \mathbf{B}_s \mathbf{m} \sum_{s=1}^N x_s. \quad (2.21)$$

En el modelo de Ising la medida de probabilidad de la realización \mathbf{x} del campo aleatorio esta dada por:

$$\mathbf{P}(\mathbf{x}) = \frac{1}{Z} e^{-\beta \mathbf{H}(\mathbf{x})} \quad (2.22)$$

donde $\beta = 1/kT$, T representa la temperatura absoluta y k la constante de Boltzmann. $\mathbf{H}(\mathbf{x})$ es la función de energía definida anteriormente y Z es la función partición,

$$Z = \sum_{\mathbf{x}} e^{-\frac{1}{Tk} \mathbf{H}(\mathbf{x})} \quad (2.23)$$

$\sum_{\mathbf{x}}$ indica que la suma se establece sobre todas las configuraciones posibles de los N espines.

El signo negativo del exponente, otorga una alta probabilidad en estados de baja energía. Un pequeño valor de β (valor alto de temperatura) produce un achatamiento de la distribución, provocando que todas las configuraciones sean relativamente iguales, mientras que un alto valor de β (valor bajo de temperatura) tiende a acentuar los valores de probabilidad en los estados de baja energía.

2.5.2 Modelo Autobinomial

Para caracterizar el contenido de una imagen compuesta por microtexturas o generar texturas sintéticas, el modelo más apropiado [6-17] es el *modelo Autobinomial*.

Sea X_s una variable aleatoria asociada al píxel s de la imagen, con distribución condicional Binomial dependiente de los valores en los sitios vecinos, de parámetros \mathbf{M}_s y probabilidad de éxito ω_s , entonces:

$$\Pi(x_s/\partial_s; \boldsymbol{\theta}) = \binom{M_s}{x_s} \omega_s^{x_s} (1 - \omega_s)^{M_s - x_s}, \quad \text{con } x_s = 0, 1, 2, \dots, M_s$$

$$\text{donde } \omega_s = \frac{e^{a + \sum_{i=1}^T \beta_i \sum_{t \in C_s} x_t}}{1 + e^{a + \sum_{i=1}^T \beta_i \sum_{t \in C_s} x_t}} = \frac{e^{\eta_s}}{1 + e^{\eta_s}} = \frac{1}{1 + e^{-\eta_s}}, \quad (2.24)$$

$$\text{y } \eta_s = a + \sum_{i=1}^T \beta_i \sum_{\langle s, t \rangle_i} x_t.$$

El valor M_s representa el nivel máximo de gris del píxel s de la imagen. Se considerará $M_s = M$, $\forall s \in \mathbf{S}$. $C_s = \{ C \in \mathcal{C}, s \in C \}$, es el conjunto de cliques que contienen a s . T representa el número de parámetros β_i a considerar, relacionado al orden del sistema de vecinos utilizado.

Este trabajo focalizará su atención en el sistema de vecinos de *segundo orden*, es decir, vecinos que se encuentran del sitio s a una distancia menor o igual a $\sqrt{2}$. Sistema que se muestra en la figura 2.8.

v	u*	z*
t	s	t*
z	u	v*

Figura 2.11 Sistema de vecinos de 2° orden del sitio s .

Por lo tanto, $T = 8$. De aquí en adelante, los parámetros se denotarán β_{ij} , donde subíndice i indica el orden del sistema de vecinos y j el tipo de cliqué en dicho orden. Por ejemplo, el parámetro correspondiente al cliqué $\{s, t\}$ es el β_{11} (el píxel ubicado en el sitio t pertenece al sistema de vecinos de primer orden del píxel s y se trata del cliqué de tipo 1). El correspondiente al cliqué $\{s, u\}$ es el β_{12} (el píxel ubicado en el sitio u pertenece al sistema de vecinos de primer orden y se trata del cliqué de tipo 2), β_{21} es el parámetro correspondiente al cliqué $\{s, v\}$ y β_{22} para el $\{s, z\}$.

Debido a la traslación invariante, los parámetros β_{st} y β_{ts} , para el par de vecinos $\{s, t\}$ y $\{s, t^*\}$ respectivamente, coinciden, al igual que los parámetros correspondientes a los vecinos: $\{s, u\}$ y $\{s, u^*\}$, $\{s, v\}$ y $\{s, v^*\}$ y $\{s, z\}$ y $\{s, z^*\}$.

Luego, para el sistema de vecinos de segundo orden del píxel ubicado en el sitio s el exponente, η_s , se expresa:

$$\eta_s = a + \beta_{11} \frac{(x_t + x_{t^*})}{M} + \beta_{12} \frac{(x_u + x_{u^*})}{M} + \beta_{21} \frac{(x_v + x_{v^*})}{M} + \beta_{22} \frac{(x_z + x_{z^*})}{M}. \quad (2.25)$$

Para cada s , la función de energía local correspondiente al modelo *Autobinomial* esta definida por:

$$\mathbf{H}(x_s, \hat{\theta}_s; \boldsymbol{\theta}) = -\ln \binom{M_s}{x_s} - a x_s - \sum_{t \in C_s^2} \beta_{st} x_s x_t = -\ln \binom{M}{x_s} - x_s \boldsymbol{\eta}_s(\boldsymbol{\theta}, x_s). \quad (2.26)$$

con $\beta_{st} = 0$ fuera de los vecinos. C_s^2 es el conjunto de todos los cliques que contienen a s , compuestos de 2 sitios y $\boldsymbol{\theta} = (a, \beta_{11}, \beta_{12}, \beta_{21}, \beta_{22})^T$ es un vector de parámetros.

La influencia de los píxeles vecinos está representada por la ecuación lineal (2.25). Los coeficientes β_{st} pesan la interacción entre los píxeles vecinos y el pixel \mathbf{s} . Reflejan la influencia de los diferentes cliques. En consecuencia, el vector $\boldsymbol{\theta}$ establece las propiedades estadísticas del campo aleatorio $\{x_s\}$, parametriza la información espacial y caracteriza el contenido de la imagen

Por ejemplo, el parámetro β_{11} pesa la correlación entre el pixel central \mathbf{s} y sus vecinos horizontales t y t^* , β_{12} pesa la correlación con sus vecinos verticales u y u^* . Las interpretaciones de los demás parámetros son análogas. El parámetro \mathbf{a} representa una autointeracción.

Para un modelo Autobinomial, la *función de partición* Z se determina mediante la expresión:

$$Z = \sum_{z_s=0}^M e^{-H(z_s, \partial_s; \boldsymbol{\theta})} = (1 + e^{\boldsymbol{\eta}})^M, \quad (2.27)$$

Por lo tanto, la probabilidad condicional se define:

$$\begin{aligned} \Pi_{\{\mathbf{s}|\mathbf{x}\}} &= \Pi(x_s / \partial_s; \boldsymbol{\theta}) = \frac{e^{-\mathbf{H}(x_s, \partial_s; \boldsymbol{\theta})}}{\sum_{z_s=0}^M e^{-\mathbf{H}(z_s, \partial_s)}} = \\ &= \frac{e^{\ln \binom{M}{x_s} + x_s a + \sum_{t \in C_s^2} \beta_{st} x_s x_t}}{e^{\ln \binom{M}{z_s} + z_s a + \sum_{t \in C_s^2} \beta_{st} z_s x_t}} = \end{aligned}$$

$$\begin{aligned}
&= \frac{\binom{M}{x_s} e^{x_s (a + \sum_{t \in C_s^2} \beta_{st} x_t)}}{\sum_{z_s=0}^M \binom{M}{z_s} e^{z_s (a + \sum_{t \in C_s^2} \beta_{st} x_t)}} = \\
&= \binom{M}{x_s} \frac{e^{x_s (a + \sum_{t \in C_s^2} \beta_{st} x_t)}}{\left[1 + e^{a + \sum_{t \in C_s^2} \beta_{st} x_t} \right]^M} \\
&= \binom{M}{x_s} \frac{\left[e^{a + \sum_{t \in C_s^2} \beta_{st} x_t} \right]^{x_s}}{\left[1 + e^{a + \sum_{t \in C_s^2} \beta_{st} x_t} \right]^M} \\
&= \binom{M}{x_s} \omega_s^{x_s} (1 - \omega_s)^{M - x_s}.
\end{aligned}$$

donde $\omega_s = \frac{e^{a + \sum_{t \in C_s^2} \beta_{st} x_t}}{1 + e^{a + \sum_{t \in C_s^2} \beta_{st} x_t}} \cdot \blacklozenge$

Esto demuestra que X_s tiene una distribución probabilidad condicional $\Pi(x_s/\partial_s; \boldsymbol{\theta})$ Binomial, dependiente de los valores en los sitios vecinos.

El valor esperado y de la varianza de la variable aleatoria X_s con distribución condicional Binomial, de parámetros \mathbf{M} y probabilidad de éxito ω_s , se obtiene:

$$E[X_s] = \mathbf{M} * \omega_s = \frac{\mathbf{M}}{1 + e^{-\eta}} \quad \text{y} \quad V[X_s] = \mathbf{M} * \omega_s * (1 - \omega_s) = \mathbf{M} \frac{e^{-\eta}}{(1 + e^{-\eta})^2}. \quad (2.28)$$

Para $\eta = 0$, los valores de la media y la varianza son $E[X_s] = \frac{\mathbf{M}}{2}$ y $V[X_s] = \frac{\mathbf{M}}{4}$, respectivamente.

En este trabajo se considerarán dos tipos de modelos Autobinomiales, creados a partir de un sistema de vecinos de segundo orden: *el Modelo Autobinomial Horizontal-Vertical (H-V)* y *el Modelo Autobinomial Diagonal Izquierda-Derecha (I-D)*. El primero evalúa la influencia de los vecinos de primer orden sobre el píxel \mathbf{s} de la imagen, es decir, refleja la interacción entre los píxeles horizontales y verticales respecto al píxel \mathbf{s} ; el segundo modelo, valora la influencia que establecen los píxeles ubicados en las diagonales izquierda y derecha sobre el píxel \mathbf{s} (Figura 2.12). En ambos casos el parámetro que refleja la autointeracción se considera nulo ($a = 0$).

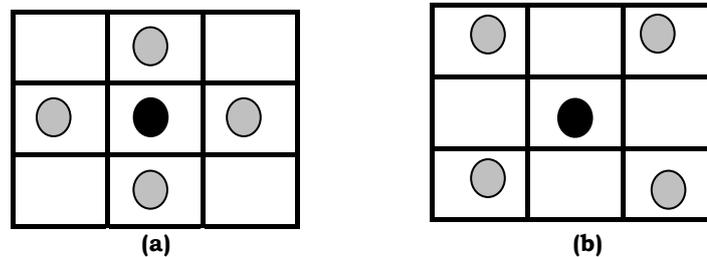


Figura 2.12 Sistemas de vecinos considerados en cada modelo: (a) Autobinomial H-V (b) Autobinomial I-D.

Luego, para el **Modelo Autobinomial H-V**, la función de energía se define:

$$\mathbf{H}(x_s, \partial_s; \theta) = - \ln \binom{M}{x_s} - x_s \boldsymbol{\eta}_s(\theta, x_s) = \ln \binom{M}{x_s} - x_s (\beta_{11} \frac{(x_t + x_{t^*})}{M} + \beta_{12} \frac{(x_u + x_{u^*})}{M}). \quad (2.29)$$

En este caso, el vector de parámetros es: $\theta = (0, \beta_{11}, \beta_{12}, 0, 0)^T$.

Para el **Modelo Autobinomial I-D**, se expresa:

$$\mathbf{H}(x_S, \partial_S; \boldsymbol{\theta}) = - \ln \binom{M}{x_S} - x_S \boldsymbol{\eta}_S(\boldsymbol{\theta}, x_S) = \ln \binom{M}{x_S} - x_S \left(\beta_{21} \frac{(x_V + x_{V^*})}{M} + \beta_{22} \frac{(x_Z + x_{Z^*})}{M} \right). \quad (2.30)$$

El vector de parámetros del modelo es: $\boldsymbol{\theta} = (0, 0, 0, \beta_{21}, \beta_{22})^T$

El trabajo de tesis se basó en estos dos modelos Autobinomiales para generar y clasificar texturas de una imagen.

SIMULACIÓN DE DISTRIBUCIONES DE GIBBS

En este capítulo se presentan muy brevemente dos de los métodos de simulación comúnmente utilizados en distribuciones de Gibbs. Los resultados de este capítulo han sido extraídos principalmente de [86]. Al final del capítulo se muestran algunas texturas generadas mediante el algoritmo GS a partir de diversos conjuntos de parámetros de los modelos Autobinomiales considerados.

Para poder evaluar visualmente las propiedades de un modelo específico o simplemente crear una imagen de textura sintética, es necesario obtener una realización de un Campo Aleatorio. Una textura es sintetizada a través de un modelo de Markov muestreando una realización $\mathbf{x} \in \Omega$ respecto a la distribución conjunta Π . En numerosos casos, debido a que el espacio de configuraciones $\Omega = \Lambda^{\mathbf{S}}$, es en general demasiado grande, resulta imposible determinar de un modo directo, una realización de un Campo Aleatorio en la malla bidimensional \mathbf{S} siguiendo la distribución de probabilidad Π . Por este motivo, es necesario implementar un algoritmo que permita obtener realizaciones de estos campos.

Los algoritmos que se desarrollarán a continuación operan generando una cadena de Markov de imágenes $\{ \mathbf{x}(0), \mathbf{x}(1), \dots, \mathbf{x}(n) \}$ (ver apartado A.1) que converge a una imagen $\mathbf{x} \in \Omega$ tal que,

$$\lim_{n \rightarrow \infty} P(X(n) = \mathbf{x} / X(0) = \mathbf{x}(0)) = \Pi(\mathbf{x}) \quad \forall \mathbf{x} \in \Omega \quad (3.1)$$

cualquiera sea la imagen inicial $X(0) = \mathbf{x}(0)$ elegida al comienzo de la cadena de Markov. Es decir, que la probabilidad de que se encuentre en el estado \mathbf{x} después de un número n suficientemente grande de iteraciones es independiente del estado de partida. Usualmente, la imagen inicial se considera una realización de un campo arbitrario sobre el espacio Ω y con distribución Uniforme. A la distribución de probabilidad $\Pi(\mathbf{x})$ se la denomina *distribución estable, estacionaria* o de *equilibrio*.

Stochastic relaxation (SR) es un método que genera cadenas de Markov [17-33]. Este término, es frecuentemente utilizado para denominar a algoritmos estocásticos que se basan en ir actualizando en sucesivos pasos cada píxel de una imagen. Los SR algoritmos más conocidos son: el Metropolis [63] y el Gibbs Sampler [31-33].

Estos algoritmos presentan las siguientes características:

- Una configuración dada es renovada en sucesivos pasos.
- La actualización en el n -ésimo paso se produce mediante cierto criterio probabilístico.
- El criterio probabilístico depende del número de paso que se trate y de la configuración actual.

A continuación se describen brevemente las características principales de cada uno de los algoritmos mencionados anteriormente.

3. 1 Algoritmo Metropolis

Entre los numerosos métodos Monte Carlo basados en Cadenas de Markov, el *algoritmo de Metrópolis* es el más popular y el más simple de aplicar. Este algoritmo fue ampliamente utilizado en problemas de la Mecánica Estadística.

El algoritmo fue desarrollado por Metropolis, Rosenbluth, Rosenbluth, Teller, y Teller (1953) para muestrear la distribución Boltzmann. En 1970, W. K. Hastings [44] lo generalizó a diferentes aplicaciones.

Se desea muestrear la distribución conjunta

$$\Pi(\mathbf{x}) = \frac{e^{-\frac{\mathbf{H}(\mathbf{x})}{kT}}}{\sum_{\mathbf{y} \in \Omega} e^{-\frac{\mathbf{H}(\mathbf{y})}{kT}}},$$

donde k es la constante de Boltzmann y T es la temperatura absoluta. Esta distribución, denominada distribución de Boltzmann-Gibbs, representa la probabilidad de que el sólido esté en el estado \mathbf{x} con energía $\mathbf{H}(\mathbf{x})$ a la temperatura T .

Este proceso parte de una configuración inicial \mathbf{x}_0 , generada mediante una distribución arbitraria ν definida sobre Ω . Esta configuración va sufriendo modificaciones utilizando realizaciones de las características locales del campo generándose así, una secuencia de configuraciones que constituyen una cadena de Markov, que bajo ciertas condiciones, converge a una realización del campo aleatorio de Boltzmann-Gibbs con distribución *de equilibrio* Π .

Esta cadena de Markov se caracteriza por la distribución inicial ν elegida, y por unas probabilidades de transición entre estados \mathbf{P}_s $s = 1, 2, \dots, |\mathbf{S}|$, denominadas *kernels de Markov*. Estas probabilidades de transición $\mathbf{P}_s(\mathbf{x}, \mathbf{x}')$ generan un nuevo estado \mathbf{x}' , a partir del valor actual \mathbf{x} , como se describe a continuación:

1. Se selecciona una configuración \mathbf{y} , donde todas sus componentes salvo la s -ésima son iguales a las de \mathbf{x} . El valor \mathbf{y}_s es seleccionado al azar a través de la función $\mathbf{G}_s(\mathbf{x}, \mathbf{y}_s)$ denominada *propuesta* o de *exploración*, que puede depender de la configuración \mathbf{x} .
2. Se acepta este nuevo estado con probabilidad $\alpha(\mathbf{x}, \mathbf{y})$; caso contrario es rechazado y se retorna al estado actual. Esto puede lograrse seleccionando al azar un número, \mathbf{u} , a través de la distribución Uniforme en $[0,1)$, y luego,

$$\mathbf{x}' = \begin{cases} \mathbf{y} & \text{si } \mathbf{u} < \alpha(\mathbf{x}, \mathbf{y}) \\ \mathbf{x} & \text{en caso contrario} \end{cases}$$

Este algoritmo permite generar un nuevo estado \mathbf{x}' de \mathbf{x} con probabilidad $\mathbf{P}(\mathbf{x}, \mathbf{x}')$ (el kernel de transición) que se define mediante la expresión:

$$\mathbf{P}_s(\mathbf{x}, \mathbf{x}') = \mathbf{G}(\mathbf{x}, \mathbf{x}'_s) \cdot \alpha(\mathbf{x}, \mathbf{x}') \prod_{i \neq k} \delta(\mathbf{x}_i, \mathbf{x}'_i) + \delta(\mathbf{x}, \mathbf{x}') \left[1 - \sum_{\bar{\mathbf{x}}} \mathbf{G}(\mathbf{x}, \bar{\mathbf{x}}) \alpha(\mathbf{x}, \bar{\mathbf{x}}) \prod_{i \neq k} \delta(\mathbf{x}_i, \bar{\mathbf{x}}_i) \right],$$

$$\text{donde } \delta(\mathbf{x}, \mathbf{x}') = \begin{cases} 1 & \mathbf{x} = \mathbf{x}' \\ 0 & \text{en otro caso} \end{cases}$$

El primer término representa la probabilidad de proponer un cambio en la componente s de la configuración \mathbf{x} , es decir, \mathbf{x}_s por \mathbf{x}'_s y luego de aceptar dicho cambio. El segundo término establece la posibilidad de rechazar el nuevo estado, por lo que permanece en el estado actual.

El algoritmo descrito anteriormente es el denominado algoritmo *Local Metropolis*, puesto que es actualizada sistemáticamente una sola componente, s , del estado \mathbf{x} , por vez, propuesta por la distribución $\mathbf{G}_s(\mathbf{x}, \mathbf{y}_s)$. En el algoritmo *Global Metropolis*, todas las componentes son actualizadas simultáneamente, usando una distribución de propuesta dada por una función simple $\mathbf{G}(\mathbf{x}, \mathbf{x}')$.

La función de propuesta, $\mathbf{G}_s(\mathbf{x}, \mathbf{y}_s)$, debe verificar las siguientes propiedades:

i) ser no negativa,

$$\text{ii) } \sum_{\bar{\mathbf{x}}_s} \mathbf{G}_s(\mathbf{x}, \bar{\mathbf{x}}_s) = 1,$$

En este caso, se va necesitar que sea simétrica, es decir,

$$\mathbf{G}_s(\mathbf{x}, \mathbf{y}_s) = \mathbf{G}_s(\mathbf{y}, \mathbf{x}_s), \text{ para } \mathbf{y}_i = \mathbf{x}_i, i \neq s.$$

En muchas aplicaciones, $\mathbf{G}_s(\mathbf{x}, \mathbf{y}_s)$ depende sólo de \mathbf{x}_s , no de \mathbf{x}_i con $i \neq s$.

La función de aceptación $\alpha(\mathbf{x}, \mathbf{y})$, puede definirse de varias maneras. Por ejemplo:

$$\alpha(\mathbf{x}, \mathbf{y}) = \min \left\{ 1, \frac{\Pi(\mathbf{y})}{\Pi(\mathbf{x})} \right\}.$$

También puede expresarse,

$$\alpha(\mathbf{x}, \mathbf{y}) = \min \left\{ 1, e^{(\mathbf{H}(\mathbf{y}) - \mathbf{H}(\mathbf{x}))/\mathbf{T}} \right\},$$

donde $\mathbf{H}(\mathbf{x})$ representa la función de energía. En este caso no es necesario el conocimiento de la función de partición Z .

Otro criterio de aceptación, propuesto por Barker (1965), ha sido

$$\alpha(\mathbf{x}, \mathbf{y}) = \frac{\Pi(\mathbf{y})}{\Pi(\mathbf{x}) + \Pi(\mathbf{y})}, \text{ para } \mathbf{x} \neq \mathbf{y}.$$

Para que la cadena de Markov converja a una realización de la distribución de Boltzmann Gibbs debe verificarse que el Kernel \mathbf{P}_s sea Π -irreducible y aperiódico, y Π es invariante bajo \mathbf{P}_s (ver Apéndice A.1.3).

La convergencia del algoritmo dependerá de la función de $\alpha(\cdot)$ elegida. Si se rechazan demasiados cambios, la convergencia será muy lenta.

Hastings [44] extendió el algoritmo de Metropolis para ampliar su uso considerando distribuciones de *propuesta* o de *exploración* que no satisfacen la condición de simetría. Por este motivo, se denominó algoritmo de Metropolis - Hastings.

3. 2 Algoritmo Gibbs Sampler (GS)

Este algoritmo denominado *Gibbs Sampler* (GS) (Muestreo de Gibbs, en castellano), fue introducido en el contexto del procesamiento de imágenes por Geman y Geman en 1984 [9]. Su nombre se debe a que se muestrea a partir de características locales de un Campo aleatorio de Gibbs. Las bases del método se remontan a N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. Teller y E. Teller (1953), con un desarrollo posterior por parte de Hastings (1970). Gelfand y Smith (1990) generaron un gran interés en revelar el amplio potencial del Gibbs Sampler para aplicarse a una vasta variedad de problemas estadísticos. Algunas de las aplicaciones se refieren a modelos lineales generalizados (Dellaportas y Smith (1990) quienes implementaron la metodología empleada por Gilks y Wild, y Zeger y Riuzaul Karim (1991)), a modelos mixtos (Diebolt y Robert 1990; Robert 1990); destinadas a la evaluación de algoritmos computacionales (Eddy y Schervish 1990); en inferencia

bayesiana en modelos para datos normales (Gelfand, Hill, y Lee 1992); al modelado de la infección de HIV (Lange, Carlin, y Gelfand 1990); al análisis bayesiano de problemas relacionados con la presencia de outliers (Verdinelli y Wasserman 1990); a inferencia bayesiana de series de tiempo Autorregresivas (Albert y Chib 1991); a la estimación de parámetros (Gelfand et al. 1992).

El algoritmo GS es un caso especial del Metropolis - Hastings Sampler y se basa en propiedades elementales de las cadenas de Markov.

El Gibbs Sampler se origina a partir de una configuración inicial \mathbf{x}_0 , obtenida al azar, que puede considerarse una realización de un campo aleatorio arbitrario definido sobre el espacio Ω y con distribución ν , que va a ir sufriendo modificaciones, utilizando realizaciones de las características locales del campo generándose así, una secuencia de configuraciones que bajo ciertas condiciones converge a una realización del CAG con distribución Π . Esta secuencia de configuraciones constituye la realización de una cadena de Markov. Esta cadena de Markov se caracteriza por una distribución inicial, y por unas probabilidades de transición entre estados \mathbf{P}_i , denominadas *kernels de Markov*. Estas probabilidades de transición son funciones $\mathbf{P}_i: \Lambda^{\mathbf{S}} \times \Lambda^{\mathbf{S}} \rightarrow [0, 1]$ que verifican

$$\text{i) } \mathbf{P}_i(\mathbf{x}, \mathbf{y}) \geq 0, \forall (\mathbf{x}, \mathbf{y}) \in \Lambda^{\mathbf{S}} \times \Lambda^{\mathbf{S}},$$

$$\text{ii) } \sum_{\mathbf{y} \in \Lambda^{\mathbf{S}}} \mathbf{P}_i(\mathbf{x}, \mathbf{y}) = 1, \forall \mathbf{x} \in \Lambda^{\mathbf{S}}.$$

Si $\mathbf{x}, \mathbf{y} \in \Omega$ son dos configuraciones, la probabilidad de transición o kernel $\mathbf{P}(\mathbf{x}, \mathbf{y})$, rige la transición de la configuración \mathbf{x} a la \mathbf{y} , es decir, determina la ley de probabilidad entre dos estados de la cadena de Markov. Los *kernels* de la cadena de Markov se determinan a partir de las características locales del campo aleatorio. Si estas probabilidades de transición, verifican ciertas propiedades, la cadena de Markov va a converger a la distribución Π del CAG deseada. Para probar la convergencia y la ergodicidad del algoritmo GS se considerarán los teoremas de convergencia y ergodicidad de las cadenas de Markov (ver apartados A.1.3 y A.2).

Si ν es la distribución de las configuraciones \mathbf{x} , entonces $\nu(\mathbf{x})\mathbf{P}(\mathbf{x}, \mathbf{y})$ es la probabilidad de seleccionar al azar a \mathbf{x} de ν y luego de elegir al azar la configuración \mathbf{y} de $\mathbf{P}(\mathbf{x}, \cdot)$. En el algoritmo GS todas las transiciones se regirán por el mismo kernel

\mathbf{P} , dando lugar a que la cadena de Markov creada sea *homogénea*. Luego, la distribución tras n iteraciones del algoritmo GS sería \mathbf{vP}^n .

La convergencia del algoritmo GS se establece si,

$$\lim_{n \rightarrow +\infty} \mathbf{vP}^n(\mathbf{x}) = \Pi(\mathbf{x}) \quad \mathbf{x} \in \Omega, \quad (3.2)$$

y uniformemente en todas las distribuciones iniciales \mathbf{v} , siendo Π el campo del que se desea obtener una realización.

Antes de determinar el kernel \mathbf{P} a partir de las características locales del campo Π , se especificarán primero los kernels de Markov definido para cada $\mathbf{N} \subset \mathbf{S}$, denotados $\Pi_{\mathbf{N}}$, que representan la probabilidad de transición de la configuración $\mathbf{x} \in \Omega$ a la configuración $\mathbf{y} \in \Omega$ y se denominan *características locales del campo*. Estos kernels se definen,

$$\Pi_{\mathbf{N}}(\mathbf{x}, \mathbf{y}) = \begin{cases} \frac{e^{-H(\mathbf{y}_{\mathbf{N}} \mathbf{x}_{\mathbf{S}/\mathbf{N}})}}{Z_{\mathbf{N}}} & \text{si } \mathbf{y}_{\mathbf{S}/\mathbf{N}} = \mathbf{x}_{\mathbf{S}/\mathbf{N}} \\ 0 & \text{en el resto} \end{cases}, \quad (3.3)$$

donde la función de partición viene dada por

$$Z_{\mathbf{N}} = \sum_{z_{\mathbf{N}} \in \Omega_{\mathbf{N}}} e^{-H(z_{\mathbf{N}} \mathbf{x}_{\mathbf{S}/\mathbf{N}})}. \quad (3.4)$$

Los kernels $\Pi_{\mathbf{N}}(\mathbf{x}, \mathbf{y})$ sólo producen cambios de la configuración \mathbf{x} en el subconjunto \mathbf{N} para la transición a la configuración \mathbf{y} . En otro caso la probabilidad de transición es nula.

En este trabajo se van a considerar subconjuntos $\mathbf{N} = \{s\}$, es decir, subconjuntos unitarios de \mathbf{S} . El kernel para pasar de la configuración de \mathbf{x} a \mathbf{y} será $\Pi_{\{s\}}$ y sólo modifica la configuración en la posición s .

Para obtener una realización de la distribución Π del CAG, será necesario ir modificando todas las posiciones $s \in \mathbf{S}$. Por tal motivo, es imperioso establecer lo que se denomina un *esquema de visita*. El esquema de visita considerado debe ser

tal que se vayan recorriendo todas las posiciones $s \in \mathbf{S}$ y que haga además que la cadena de Markov converja a la distribución deseada. En cada una de las sucesivas visitas se considera una nueva configuración que difiere de la previa en sólo una posición, es decir, la configuración $\mathbf{x}_i = \mathbf{x}_{i-1}$ para todas las posiciones que no coinciden con la posición seleccionada para el cambio.

El *esquema de visita* puede ser *determinista* o *aleatorio*. Entre los tipos de esquemas de visita determinísticos se encuentran el *secuencial*, que si bien es el más sencillo, en general, no proporciona mayor convergencia, y el denominado *parcialmente síncrono* o *esquema paralelo parcial*, que resulta ser más eficiente. El primero consiste en enumerar los elementos de la rejilla $\mathbf{S} = \{s_1, s_2, \dots, s_{|\mathbf{S}|}\}$ y recorrer secuencialmente, uno por uno, del primero al último, dando lugar a lo que se denomina un *barrido*. No existe restricción alguna respecto al orden elegido en que se visitan los distintos elementos. Este mismo esquema se repite para el resto de los barridos, es decir, si \mathbf{x} es una configuración para un barrido e \mathbf{y} lo es para el barrido siguiente, el kernel de Markov está definido,

$$\mathbf{P}(\mathbf{x}, \mathbf{y}) = \Pi_{\{s_1\}} \Pi_{\{s_2\}} \dots \Pi_{\{s_{|\mathbf{S}|-1}\}} \Pi_{\{s_{|\mathbf{S}|\}}(\mathbf{x}, \mathbf{y}). \quad (3.5)$$

donde los $\Pi_{\{s_i\}}$ representan las características locales del CAG Π , y s_i indica cada uno de los elementos del esquema de visita $\mathbf{S} = \{s_1, s_2, \dots, s_{|\mathbf{S}|}\}$. Es claro que, el kernel \mathbf{P} queda unívocamente determinado a partir de las características locales para cada posición $s \in \mathbf{S}$. En un esquema de visitas *determinístico* se dice que se ha efectuado un *barrido* cada vez que se recorren todas las posiciones de $s \in \mathbf{S}$ para pasar de una configuración \mathbf{x} a una \mathbf{y} .

El segundo esquema determinístico, se basa en establecer una partición $\mathbf{U}^8 = \{\mathbf{U}_1, \mathbf{U}_2, \dots, \mathbf{U}_L\}$ del conjunto \mathbf{S} en conjuntos disjuntos, \mathbf{U}_i de posiciones s que tengan características locales independientes (que sean condicionalmente independientes con respecto a la distribución de la ecuación (2.19)). En este esquema, todas las posiciones de cada partición se actualizan de forma simultánea, permitiendo que el algoritmo GS se ejecute de forma paralela, lo que disminuye considerablemente el tiempo de ejecución. Para más detalle puede consultarse [85]

⁸ Para que \mathbf{U} sea una *partición* de \mathbf{S} debe satisfacer: **1.** $\mathbf{U}_i \cap \mathbf{U}_j = \emptyset$ y **2.** $\bigcup_{i=1}^L \mathbf{U}_i = \mathbf{S}$.

En el esquema de *visita aleatorio*, las posiciones $s \in \mathbf{S}$ a ser visitadas son seleccionadas al azar, regidas por una distribución de probabilidad \mathbf{G} . Esta distribución \mathbf{G} se denomina distribución de *exploración* o de *propuesta*. Usualmente \mathbf{G} es la distribución uniforme en \mathbf{S} . Una vez que se ha determinado s se procede igual que en el esquema determinista. En este caso un barrido significa $|\mathbf{S}|$ visitas aleatorias para pasar de una configuración \mathbf{x} a una \mathbf{y} . Este fue el esquema de visita elegido en este trabajo de tesis.

El kernel de Markov $\tilde{\mathbf{P}}$ de un barrido se define,

$$\tilde{\mathbf{P}}(\mathbf{x}, \mathbf{y}) = \left[\sum_{s \in \mathbf{S}} \mathbf{G}(s) \Pi_{\{s\}}(\mathbf{x}, \mathbf{y}) \right]^{|\mathbf{S}|}. \quad (3.6)$$

Es importante resaltar que para que la cadena de Markov generada por el kernel $\tilde{\mathbf{P}}$ converja a la distribución Π del CAG deseada, $\mathbf{G}(s)$ debe ser *estrictamente positiva*, es decir, $\mathbf{G}(s) > 0 \forall s \in \mathbf{S}$. Todas las posiciones tienen probabilidad de ser visitadas.

En resumen, el algoritmo GS puede generarse como sigue:

Se parte de una configuración arbitraria \mathbf{x}_0 ; siguiendo un esquema de visita ya sea determinístico o aleatorio, la configuración va sufriendo modificaciones en las posiciones $s \in \mathbf{S}$ obteniéndose una realización \mathbf{x}_1 de las características locales de la ecuación (2.17) hasta completar un barrido. Operando de forma similar para los demás barridos, se genera la secuencia de configuraciones $\{\mathbf{x}_i\}_{i \geq 0}$, que converge a la realización a la distribución Π del CAG deseada (ver Algoritmo en B.2).

A continuación se presentan algunos ejemplos de texturas generadas mediante diversos conjuntos de parámetros de los modelos Autobinomiales H-V e I-D, a través de los algoritmo GS (ver Algoritmos B.2.2 y B.2.3). La implementación de los distintos algoritmos GS se efectuaron mediante el software Matlab versión 7.2.0.232 (R2006a) y a través de una computadora WTC modelo EXPLORER, con Microprocesador INTEL C2D E7300, Motherboard ASRock, 3Gb de memoria RAM DDR2 667 mhz, Disco Rígido de 320Gb SATA.

En este caso se han utilizado imágenes de 64 x 64 píxeles con 256 niveles de gris. El número de configuraciones distintas de Ω es 256^{4096} . En cada uno de los

ejemplos considerados, se partió de una imagen con distribución uniforme en la escala de los 256 niveles de gris. Se ha empleado un esquema de visita aleatorio para ambos modelos. Se considera cumplido un barrido cuando se hayan analizado los 3844 posibles cambios. En este trabajo no se han tenido en cuenta los píxeles del borde de la imagen (puesto que las aplicaciones se realizarán sobre imágenes satelitales terrestres). Se dará por finalizado este proceso cuando se hayan completado 200 barridos, criterio propuesto por Jan - Johansson (2001), para la estabilización del algoritmo.

La figura 3.1 muestra cuatro texturas homogéneas, con características isotrópicas, generadas a partir de distintos conjuntos de parámetros del modelo Autobinomial H-V. Puede observarse que las imágenes **II**, **III** y **IV** muestran texturas con una apariencia visual determinada por la dirección en la que el parámetro de primer orden β_{1j} tiene un valor positivo.

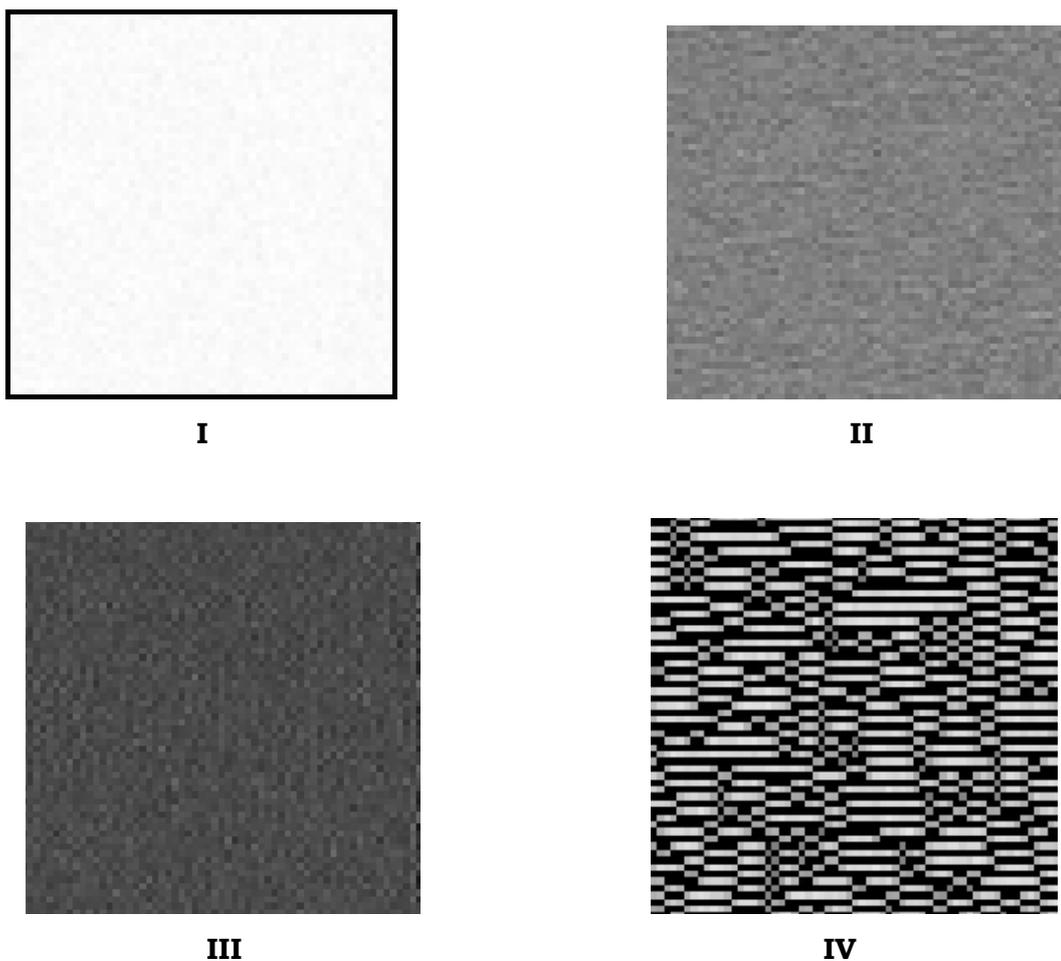


Figura 3.1 Texturas generadas con el algoritmo GS. Ejemplos de cuatro realizaciones del modelo Autobinomial H-V.

En la figura 3.2 se exhiben cuatro ejemplos de texturas, en escala de grises, de tamaño 64 x 64 píxeles, generadas a partir de ciertos valores específicos de los parámetros del modelo Autobinomial I-D. Puede apreciarse que las texturas **VI**, **VII** y **VIII** se caracterizan por la dirección en la cual el parámetro β_{2j} de segundo orden toma el valor positivo. Se trata de imágenes con características isotrópicas, con texturas homogéneas. La imagen 3.2.**V** es totalmente equivalente a ruido, ya que la probabilidad de que un píxel de ella presente un nivel de gris dado es estadísticamente independiente de los niveles de gris de sus vecinos debido a todos los parámetros del modelo considerado son nulos.

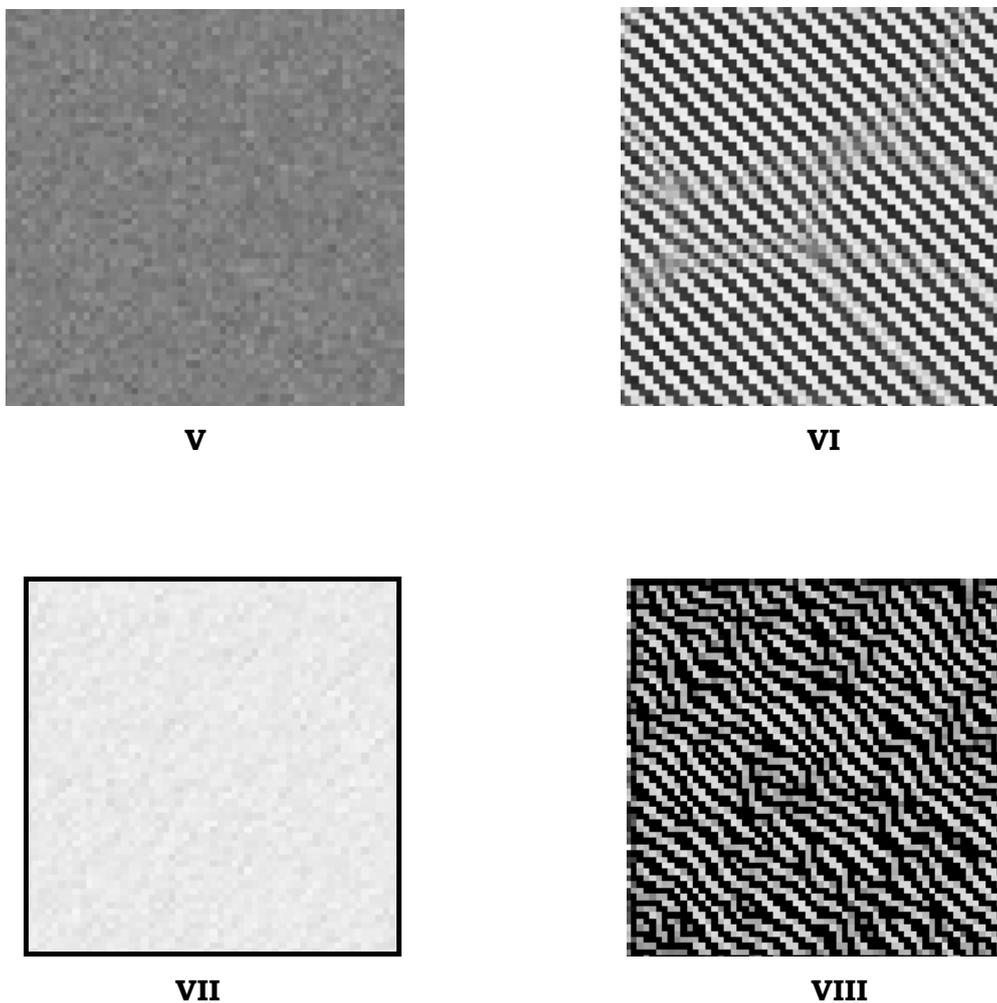


Figura 3.2 Texturas generadas con el algoritmo GS. Ejemplos de cuatro realizaciones del modelo Autobinomial I-D.

En la tabla 3.1 y 3.2 se exhiben los valores de los parámetros de los modelos Autobinomiales H-V y I-D utilizados para generar cada una de las texturas previas (figura 3.1 y 3.2).

Tabla 3.1 Parámetros utilizados para generar las texturas de la Figura 3.1.

Imágenes	Modelo H-V	
	β_{11}	β_{12}
I	1.0	1.0
II	0.7	-0.7
III	-1.5	0.0
IV	1.0	-10.0

Tabla 3.2 Parámetros utilizados para generar las texturas de la Figura 3.2.

Imágenes	Modelo I-D	
	β_{11}	β_{12}
V	0.0	0.0
VI	-1.1	1.5
VII	2.5	-1.2
VIII	-10.0	1.0

ESTIMADORES DE LOS PARÁMETROS EN DISTRIBUCIONES DE GIBBS

En este capítulo se presenta muy brevemente algunos de los estimadores de los parámetros en distribuciones de Gibbs y en particular se detallan los distintos estimadores del modelo Autobinomial: el estimador de *Máxima Pseudo-Verosimilitud* y el estimador de *Mínimos Cuadrados Condicional*. En el apéndice B se desarrollan los algoritmos de ambos estimadores, que permiten la estimación del vector de parámetros de la función de energía de un modelo de Gibbs Autobinomial H-V e I-D, respectivamente.

En el capítulo anterior se desarrolló el algoritmo Gibbs Sampler que permite, fijados ciertos valores de los parámetros de la función de energía, generar diferentes texturas, es decir, proporcionar una realización \mathbf{x} de un Campo Aleatorio de Gibbs. Ahora el objetivo es inverso, dada una textura real o sintética se desea determinar cuáles son los valores de los parámetros que la generan, es decir, cuál es el Campo Aleatorio de Gibbs que modela la textura considerada.

La determinación del Campo Aleatorio de Gibbs a partir de una realización particular \mathbf{x} , puede reformularse como un “*problema de estimación de parámetros*”.

4.1 Estimación de parámetros

En los modelos de Gibbs, la estimación del vector de parámetros $\theta = (a, \beta^T)^T = (a, b_{11}, b_{12}, b_{21}, b_{22}, \dots)^T$ de la función de energía $\mathbf{H}(\mathbf{x}; \theta)$ es el punto

clave en la extracción de información espacial como así también, la caracterización de las propiedades locales de las microtexturas presentes en una imagen. Representa la influencia que ejercen sobre los x_s los diferentes cliques $\{s, t\}$. Es por ello, que resulta de vital importancia el uso de estimadores confiables y robustos. Existen varios métodos de estimación de parámetros de los modelos de Gibbs: Maximum Likelihood Method (ML), Coding Method [6], Maximum Pseudo-likelihood Method [6], el Least Squares Error Method (24), el Condicional Least Squares Method [59], el Logit Model Fit Method (13) y el Minimum Logit Chi-Square Method (24). Este último método está basado en el procedimiento de Berkson (4), pero sólo se aplica a variables binarias.

Estos procedimientos de estimación no necesariamente conducen a una idéntica estimación del vector de parámetros.

Se pretende determinar los valores de las componentes del vector de parámetros $\theta = (a, \beta^T)^T = (a, b_{11}, b_{12}, b_{21}, b_{22}, \dots)^T$ en base a una única realización \mathbf{x} del sistema.

Un estimador del vector de parámetros θ es una función $\hat{\theta}(\mathbf{x})$, que proporciona para cada observación \mathbf{x} una aproximación para el vector de parámetros óptimo.

Como se indicó anteriormente se considerará que el CAG Autobinomial Π definido en Ω , depende de un vector de parámetros a estimar θ , constituido por 5 componentes (sistema de vecinos de segundo orden), es decir, $\theta = (a, \beta^T)^T = (a, b_{11}, b_{12}, b_{21}, b_{22})^T$ pertenece al espacio de parámetros $\Theta \in \mathfrak{R}^5$. Por lo tanto,

$$\Pi(\mathbf{x}; \theta) = \frac{1}{Z(\theta)} e^{-\mathbf{H}(\mathbf{x}; \theta)}, \quad (4.1)$$

siendo $\mathbf{H}(\mathbf{x}; \theta)$, la función de energía definida:

$$\mathbf{H}(\mathbf{x}; \theta) = \sum_{s \in \mathbf{S}} V_1(x_s) + \sum_{s \in \mathbf{S}} \sum_{t \in \partial s} V_2(x_s, x_t) = \sum_{s \in \mathbf{S}} \ln \binom{M}{x_s} + \sum_{s \in \mathbf{S}} x_s \eta_s.$$

$$\text{con } \eta_s = a + \beta_{11} \frac{(x_t + x_{t^*})}{M} + \beta_{12} \frac{(x_u + x_{u^*})}{M} + \beta_{21} \frac{(x_v + x_{v^*})}{M} + \beta_{22} \frac{(x_z + x_{z^*})}{M}, \quad (4.2)$$

y $Z(\boldsymbol{\theta}) = \sum_{\mathbf{z} \in \Omega} e^{-\mathbf{H}(\mathbf{z}; \boldsymbol{\theta})}$, la función de partición.

4.1.1 Estimador de Máxima Verosimilitud

El objetivo de la técnica de estimación de Máxima Verosimilitud o MV (Maximum likelihood) es determinar los valores de las componentes del vector de parámetros $\boldsymbol{\theta} = (\alpha, \beta^T)^T$ tal que el valor de $\Pi(\mathbf{x}; \boldsymbol{\theta})$ sea máximo para una única realización \mathbf{x} . En el caso discreto, el estimador de Máxima Verosimilitud es un estimador que maximiza la probabilidad de ocurrencia de la realización observada \mathbf{x} .

La *función de verosimilitud* se define:

$$\mathcal{L}(\mathbf{x}; \boldsymbol{\theta}) = \Pi(\mathbf{x}; \boldsymbol{\theta}) = \frac{1}{Z(\boldsymbol{\theta})} e^{-\mathbf{H}(\mathbf{x}; \boldsymbol{\theta})}. \quad (4.3)$$

En la función $\mathcal{L}(\mathbf{x}, \boldsymbol{\theta})$, \mathbf{x} representa la configuración observada, por lo tanto es constante, las únicas incógnitas son las componentes del vector de parámetros $\boldsymbol{\theta}$. Cualquier valor que maximice $\mathcal{L}(\mathbf{x}; \cdot)$ se denomina estimador de *Máxima Verosimilitud* del vector de parámetros $\boldsymbol{\theta}$. Este valor se obtiene calculando la derivada de $\mathcal{L}(\mathbf{x}; \boldsymbol{\theta})$ respecto de $\boldsymbol{\theta}$, igualando a cero y despejando $\boldsymbol{\theta}$. Generalmente, se utiliza el logaritmo natural de la función de verosimilitud para determinar de una manera más práctica y más sencilla la derivada de $\mathcal{L}(\mathbf{x}; \boldsymbol{\theta})$ con respecto a $\boldsymbol{\theta}$.

$$\mathbf{L}(\mathbf{x}; \boldsymbol{\theta}) = \ln \mathcal{L}(\mathbf{x}; \boldsymbol{\theta}) = \ln \Pi(\mathbf{x}; \boldsymbol{\theta}) = -\mathbf{H}(\mathbf{x}; \boldsymbol{\theta}) - \ln Z(\boldsymbol{\theta}). \quad (4.4)$$

Sea $\mathbf{y} \in \Omega$. Se define $Y(\mathbf{y}) \in \mathbb{R}^5$ por:

$$Y(\mathbf{y}) = \left(\sum_{s \in \mathbf{S}} y_s, - \sum_{s \in \mathbf{S}} y_s \frac{(y_t + y_{t^*})}{M}, - \sum_{s \in \mathbf{S}} y_s \frac{(y_u + y_{u^*})}{M}, - \sum_{s \in \mathbf{S}} y_s \frac{(y_v + y_{v^*})}{M}, - \sum_{s \in \mathbf{S}} y_s \frac{(y_z + y_{z^*})}{M} \right).$$

Entonces, el logaritmo natural de la función de verosimilitud, $\mathbf{L}(\mathbf{x}; \boldsymbol{\theta})$, dada la observación \mathbf{x} puede escribirse:

$$\mathbf{L}(\mathbf{x}; \boldsymbol{\theta}) = \boldsymbol{\theta}^T \mathbf{Y}(\mathbf{x}) - b(\boldsymbol{\theta}) - \mathbf{H}(\mathbf{x}; \mathbf{0}), \quad (4.5)$$

donde $\boldsymbol{\theta}$ es el vector columna $\boldsymbol{\theta} = (a, \boldsymbol{\beta}^T)^T = (a, b_{11}, b_{12}, b_{21}, b_{22})^T$

En efecto,

$$\mathbf{H}(\mathbf{x}; \boldsymbol{\theta}) = \sum_{s \in \mathbf{S}} \ln \binom{M}{x_s} + \sum_{s \in \mathbf{S}} x_s \boldsymbol{\eta}_s = \sum_{s \in \mathbf{S}} \ln \binom{M}{x_s} + \boldsymbol{\theta}^T \mathbf{Y}(\mathbf{x}).$$

Luego:

$$\mathbf{H}(\mathbf{x}; \mathbf{0}) = \sum_{s \in \mathbf{S}} \ln \binom{M}{x_s},$$

donde $\boldsymbol{\theta} = \mathbf{0} = (0, 0, 0, 0, 0)$.

En consecuencia:

$$\begin{aligned} \mathbf{L}(\mathbf{x}; \boldsymbol{\theta}) &= \ln \mathcal{L}(\mathbf{x}; \boldsymbol{\theta}) = -\mathbf{H}(\mathbf{x}; \boldsymbol{\theta}) - \ln Z(\boldsymbol{\theta}) \\ &= -[\mathbf{H}(\mathbf{x}; \boldsymbol{\theta}) - \mathbf{H}(\mathbf{x}; \mathbf{0})] - \ln Z(\boldsymbol{\theta}) - \mathbf{H}(\mathbf{x}; \mathbf{0}) \\ &= \boldsymbol{\theta}^T \mathbf{Y}(\mathbf{x}) - b(\boldsymbol{\theta}) - \mathbf{H}(\mathbf{x}; \mathbf{0}). \end{aligned}$$

Si $\hat{\boldsymbol{\theta}}$ es el estimador de MV de $\boldsymbol{\theta}$ debe satisfacer:

$${}^9 \nabla \mathbf{L}(\mathbf{x}; \hat{\boldsymbol{\theta}}) = \mathbf{0}, \quad (4.6)$$

donde se deduce que $\hat{\boldsymbol{\theta}}$ debe verificar:

⁹ Se llama gradiente de una función $z = f(x_1, \dots, x_n)$ en un punto $P(x_1, \dots, x_n)$ al vector normal a la superficie $f(x_1, \dots, x_n)$ en el punto P y sus componentes son las derivadas parciales de la función en dicho punto:

$$\text{grad}(z) = \nabla z = \left(\frac{\partial z}{\partial x_1}, \dots, \frac{\partial z}{\partial x_n} \right) = f_{x_1}(x_1, \dots, x_n) \hat{i}_1 + f_{x_2}(x_1, \dots, x_n) \hat{i}_2 + \dots + f_{x_n}(x_1, \dots, x_n) \hat{i}_n.$$

$$\mathbf{Y}(\mathbf{x}) = \sum_{y \in \Omega} \mathbf{Y}(y) \frac{e^{\hat{\boldsymbol{\theta}}^T \mathbf{Y}(y) - H(y,0)}}{\sum_{z \in \Omega} e^{\hat{\boldsymbol{\theta}}^T \mathbf{Y}(z) - H(z,0)}}. \quad (4.7)$$

Para despejar $\hat{\boldsymbol{\theta}}$ de la ecuación (4.7), es necesario recurrir a un proceso de optimización numérica como por ejemplo, el *Método de Newton - Raphson*¹⁰.

Si bien se ha demostrado la existencia, unicidad y consistencia del estimador MV [76] no resulta ser un estimador útil a efectos prácticos debido a que la función de partición Z, representa un problema inviable computacionalmente.

4.1.2 Coding

Este método fue propuesto por Besag en 1974. Consiste en definir unos códigos en la malla para estimar los parámetros de un modelo. El número de estos códigos o particiones depende del sistema de vecinos considerado.

Besag [6] comprobó que el método de estimación *Coding* resultaba ineficiente, debido a que sólo se utiliza un porcentaje de los datos para cada estimación y así las distintas estimaciones efectuadas en cada uno de los “codings” definidos pueden ser muy diferentes debido a la independencia existente entre los diferentes “codings”, lo que conlleva a la obtención de valores medios inadecuados. En consecuencia, propuso, en 1975 un nuevo método de estimación, denominado el Máximo Pseudo-verosimilitud.

Puesto que el propósito de este trabajo es extraer información espacial, mediante los modelos Autobinomiales H-V e I-D, y esto es posible, a través de la estimación del vector de parámetros de sus respectivas funciones de energía, Schöder, Rehrauer, Siedel y Dactu (1998) establecen que los estimadores más adecuados para tal fin son: el Máximo Pseudo-Verosimilitud (Maximum Pseudo-Likelihood Estimator) y de Mínimos Cuadrados Condicional (Conditional Least-Squares Estimator).

¹⁰ Es un método numérico iterativo utilizado para encontrar aproximaciones de los ceros o raíces de una función real y también es ampliamente usado para resolver sistemas de ecuaciones no lineales.

A continuación se desarrollará una breve descripción de estos dos métodos de estimación.

Para la presentación de estos dos métodos de estimación, se considerará la expresión de la probabilidad condicional definida en (2.25),

$$\Pi(x_s/\partial_s; \boldsymbol{\theta}) = \binom{M}{x_s} \varpi^{x_s} (1 - \varpi)^{M - x_s}, \quad (4.8)$$

donde $\varpi = \frac{1}{1 + e^{-\boldsymbol{\eta}_s}}$ y $\boldsymbol{\eta}_s$, definida en (4.2).

4.1.3 Estimador de Máxima Pseudo-verosimilitud

El estimador de *Máxima Pseudo-verosimilitud* (MP), como se mencionó anteriormente, fue desarrollado por Besag [7-8] como una alternativa del estimador de *Máxima Verosimilitud*.

Para aplicar este método de estimación, se va asumir que la probabilidad conjunta de una imagen es producto de todas las probabilidades individuales de los píxeles de sitio $s \in \mathbf{S}$ dados los valores correspondientes de sus píxeles vecinos:

$$\Pi(\mathbf{x}; \boldsymbol{\theta}) \cong \prod_{s \in \mathbf{S}} \Pi(x_s/\partial_s) = \prod_{s \in \mathbf{S}} \frac{e^{-\mathbf{H}(x_s/\partial_s; \boldsymbol{\theta})}}{Z_s(\boldsymbol{\theta})}. \quad (4.9)$$

Es decir, la probabilidad conjunta es producto de probabilidades locales de los sitios en \mathbf{S} .

Con esta aproximación el estimador de *Máxima Pseudo-Verosimilitud* o MP de los parámetros se obtiene maximizando la pseudo probabilidad, ecuación (4.9). Es decir,

$$\hat{\boldsymbol{\theta}}_{\text{MP}} = \arg \max_{\boldsymbol{\theta}} \ln \left[\prod_{\boldsymbol{\theta}} (\{x_s\}) \right] = \arg \max_{\boldsymbol{\theta}} \ln \left[\prod_{s \in \mathbf{S}} \Pi_{\boldsymbol{\theta}}(x_s/\partial_{x_s}) \right]. \quad (4.10)$$

Este estimador en lugar de maximizar $\Pi(\mathbf{x}; \boldsymbol{\theta})$ con respecto a $\boldsymbol{\theta} \in \Theta$ dada una observación $\mathbf{x} \in \Omega$, maximiza el producto de las características locales. Esta función, sólo depende de $\boldsymbol{\theta}$, puesto a que la configuración \mathbf{x} es considerada una constante.

$$\begin{aligned}\hat{\boldsymbol{\theta}}_{\text{MP}} &= \arg \max_{\boldsymbol{\theta}} \ln \left[\prod_{s \in \mathbf{S}} \binom{M}{x_s} \varpi^{x_s} (1 - \varpi)^{M - x_s} \right] \\ &= \arg \max_{\boldsymbol{\theta}} \ln \left[\prod_{s \in \mathbf{S}} \binom{M}{x_s} \frac{e^{\boldsymbol{\eta}_s x_s}}{(1 + e^{\boldsymbol{\eta}_s})^M} \right].\end{aligned}\quad (4.11)$$

Como puede apreciarse en la expresión (4.11), para determinar el estimador MP se aplica, por conveniencia, el logaritmo natural de la pseudo probabilidad. Luego, si se define

$$\mathbf{L}(\mathbf{x}; \boldsymbol{\theta}) = \ln \Pi(\mathbf{x}; \boldsymbol{\theta}) \cong \sum_{s \in \mathbf{S}} \left[\ln \binom{M}{x_s} + \boldsymbol{\eta}_s x_s - M \ln(1 + e^{\boldsymbol{\eta}_s}) \right].\quad (4.12)$$

Para una imagen dada, la expresión $\sum_{s \in \mathbf{S}} \ln \binom{M}{x_s}$ es constante y puede ser eliminada.

Por lo tanto, la función a maximizar se reduce a:

$$\mathbf{L}(\mathbf{x}; \boldsymbol{\theta}) = \sum_{s \in \mathbf{S}} \left[\ln \binom{M}{x_s} + \boldsymbol{\eta}_s x_s - M \ln(1 + e^{\boldsymbol{\eta}_s}) \right]\quad (4.13)$$

donde, si se considera el modelo Autobinomial H-V, la expresión de $\boldsymbol{\eta}_s$ es:

$$\boldsymbol{\eta}_s = \beta_{11} \frac{(x_t + x_{t^*})}{M} + \beta_{12} \frac{(x_u + x_{u^*})}{M},\quad (4.14)$$

y si se trata del modelo Autobinomial I-D, la expresión de $\boldsymbol{\eta}_s$ es:

$$\boldsymbol{\eta}_s = \beta_{21} \frac{(x_v + x_{v^*})}{M} + \beta_{22} \frac{(x_z + x_{z^*})}{M}.\quad (4.15)$$

Para poder maximizar el producto de las características locales con respecto a $\theta \in \Theta$, dada una observación $\mathbf{x} \in \Omega$, se empleó el *Método de Newton* (desarrollado en el Apéndice A).

La ventaja principal que presenta este estimador es que las funciones de partición, para cada una de los sitios, son computacionalmente viables.

En 1989, Geman y Graffigne, demostraron la existencia, unicidad y consistencia de este estimador bajo condiciones generales. F. Comets, en 1992, demostró la consistencia asintótica del estimador MP. Seymour, a través de experimentos, dio a conocer ciertas limitaciones del estimador MP; mostró que como optimiza sólo propiedades locales, no es aplicable a imágenes con un amplio rango de dependencia. Sin embargo, diversos autores, como Guyon (1986), (1987), Jensen y Moller (1989) coinciden en que el MP es un estimador que brinda resultados razonables y consistentes.

4.1.4 Estimador de Mínimos Cuadrados Condicional

El estimador de *Mínimos Cuadrados Condicional* o MCC (*Conditional least squares estimator*), intenta minimizar el error entre la imagen dada y la esperada. Se define como:

$$\hat{\theta}_{\text{MCC}} = \arg \min_{\theta} \sum_{\mathbf{s}} (x_{\mathbf{s}} - \mathbf{E}(x_{\mathbf{s}}))^2, \text{ donde } \mathbf{E}(x_{\mathbf{s}}) = \sum_{\mathbf{s} \in \mathbf{S}} x_{\mathbf{s}} \Pi(x_{\mathbf{s}} / \partial_{\mathbf{s}}; \theta) \quad (4.16)$$

Empleando el valor esperado del Modelo Autobinomial, la ecuación (4.16) puede escribirse:

$$\hat{\theta}_{\text{MCC}} = \arg \min_{\theta} \sum_{\mathbf{s} \in \mathbf{S}} \left(x_{\mathbf{s}} - \frac{M}{1 + e^{-\eta_{\mathbf{s}}}} \right)^2 \quad (4.17)$$

$\eta_{\mathbf{s}}$ representa la influencia de los píxeles vecinos del píxel \mathbf{s} .

En [78] se muestra que, mediante el uso de la aproximación del campo medio, $\eta_{\mathbf{s}}$ puede ser aproximado por:

$$\eta_{\mathbf{s}} \approx -\log \left[\frac{M}{x_{\mathbf{s}}} - 1 \right] + \kappa_{\mathbf{s}}, \quad (4.18)$$

donde κ_s es un término Gaussiano de ruido de media cero y varianza δ . Con la definición de η_s , ecuación (2.18), el proceso de estimación de parámetros se reduce a resolver un sistema de ecuaciones lineales:

$$\mathbf{G} \boldsymbol{\theta} = \mathbf{d} + \mathbf{n}, \quad (4.19)$$

donde $\boldsymbol{\theta}$ es el vector de parámetros desconocidos, \mathbf{d} el vector de valores transformados de los píxeles x_s

$$\mathbf{d}_s = -\log \left[\frac{M}{x_s} - 1 \right] \quad (4.20)$$

y \mathbf{n} un vector Gaussiano de ruido. \mathbf{G} es una matriz de contaminación de los valores de los píxeles vecinos,

$$\mathbf{G}_{\mathbf{s},\mathbf{n}} = \begin{cases} 1 & \text{si } n = 0 \\ \frac{x_{st(n)} + x'_{st(n)}}{M} & \text{en caso contrario} \end{cases} \quad (4.21)$$

Los valores x_{st} y x'_{st} denotan los dos vecinos del píxel x_s descritos en el cliqué $\langle s, t \rangle_n$. La s -ésima fila especifica los niveles de gris $\{x_t\}$ de todos los vecinos del píxel s y la n -ésima columna contiene al vecino n de todos los píxeles de la imagen.

El parámetro \mathbf{a} no pondera ningún vecino, por ende vale 1.

Por ejemplo, si se tiene una imagen de $n \times n$ y se trabaja con el sistema de vecinos definidos para el modelo Autobinomial H-V, la matriz \mathbf{G} , tendrá un orden de $N \times 2$, con $N=n \times n$ filas. Los elementos de las dos columnas de \mathbf{G} se determinan mediante las siguientes expresiones:

$$\mathbf{g}_{s1} = \frac{(x_t + x_{t^*})}{M} \quad \text{y} \quad \mathbf{g}_{s2} = \frac{(x_u + x_{u^*})}{M},$$

para cada píxel \mathbf{s} de la imagen, donde x_t y x_{t^*} son los niveles de gris de sus vecinos ubicados horizontalmente y x_u y x_{u^*} son los de sus vecinos ubicados verticalmente.

Si se considera el modelo Autobinomial I-D, la matriz \mathbf{G} , tendrá el mismo orden y los elementos de las dos columnas quedarán determinados mediante las expresiones:

$$\mathbf{g}_{s1} = \frac{(x_v + x_{v^*})}{M} \quad \text{y} \quad \mathbf{g}_{s2} = \frac{(x_z + x_{z^*})}{M},$$

para cada píxel \mathbf{s} de la imagen, en este caso, x_v y x_{v^*} representan los niveles de gris de sus vecinos ubicados en la diagonal izquierda y x_z y x_{z^*} son los de sus vecinos ubicados en la diagonal derecha.

Considerando que el término de ruido \mathbf{n} posee distribución normal, el estimador de máxima verosimilitud de θ esta dado por:

$$\hat{\theta}_{\text{CMC}} = (\mathbf{G}^T \mathbf{G})^{-1} \mathbf{G}^T \mathbf{d}. \quad (4.22)$$

Esto provee una solución para los parámetros de textura, permitiendo para un espacio de parámetros de dimensión pequeña lograr una mayor eficiencia en el MAP de información texturada de la imagen.

Una ventaja importante de este estimador es su cálculo resulta significativamente rápido y sencillo, por este motivo, es ampliamente usado. Esto no implica que sea el más recomendado.

Los algoritmos para determinar el estimador de MCC del vector de parámetros del modelo Autobinomial H-V y el del modelo I-D, fueron desarrollados en el Apéndice B, Algoritmo B.4.1 y B.4.2 respectivamente.

En el próximo capítulo, se evaluarán y compararán los resultados emitidos por estos dos últimos métodos de estimación.

ESTUDIO DE SENSIBILIDAD

En este capítulo se compararán los métodos de estimación de *Máxima Pseudo-verosimilitud* y el de *Mínimos Cuadrados Condicional* y se discutirá la sensibilidad de los mismos frente a distintos grados de contaminación de diferentes patrones homogéneos de textura.

5.1 Comparación y Análisis de sensibilidad

De cada uno de los dos modelos Autobinomiales considerados, el Modelo Horizontal-Vertical (H-V) y el Modelo Diagonal Izquierda-Derecha (I-D), se establecieron cuatro situaciones diferentes para el par de parámetros (Tabla 5.1).

Tabla 5.1. Las cuatro situaciones de valores específicos de los parámetros de los modelos.

Modelo H - V		Modelo I - D	
β_{11}	β_{12}	β_{21}	β_{22}
0.00	1.00	0.00	1.00
1.00	1.00	1.00	1.00
1.00	-1.00	1.00	-1.00
-1.00	1.00	-1.00	1.00

Se generaron, para cada situación, 100 texturas de 32 x 32 píxeles con 200 barridos del algoritmo Gibbs Sampler. Mediante el software MATLAB (versión 7.2.0.232 (R2006a)) se confeccionaron tres algoritmos para generar tres grados de contaminación diferentes sobre las 800 texturas generadas. Los grados de

contaminación considerados fueron: el 1% (10 de los píxeles de la imagen), el 10% (102 píxeles) y el 20% (205 píxeles). El proceso de contaminación consistió en sustituir al azar el 1%, el 10% y el 20%, respectivamente, de los niveles de gris de los píxeles de las imágenes previamente generadas, por el nivel máximo de gris. Creando de esta forma, tres modelos contaminados para cada modelo Autobinomial H-V e I-D. En las figuras 5.1 y 5.2 se muestran dos ejemplos de los efectos que producen los distintos grados de contaminación, en texturas generadas por los modelos Autobinomiales H-V e I-D, respectivamente.



Figura 5.1 Modelo Autobinomial H-V. La textura de la izquierda está generada utilizando los valores específicos ($\beta_{11} = 0$, $\beta_{12} = 1$), la segunda presenta una contaminación del 1% de los píxeles, la siguiente un 10% y la de la derecha un 20% de los mismos.

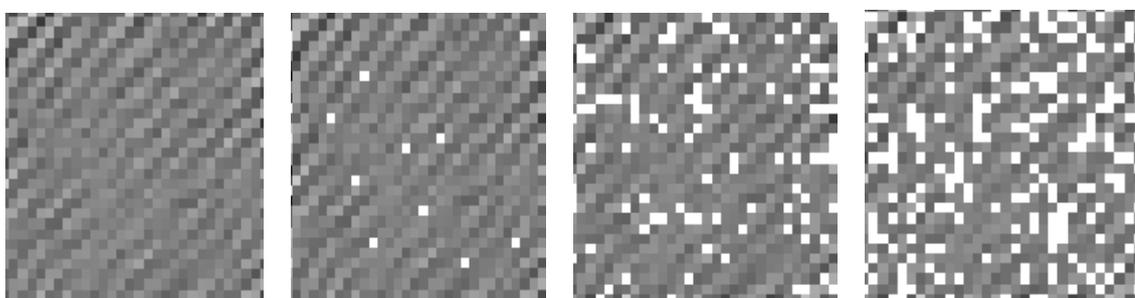


Figura 5.2 Modelo Autobinomial I-D. La textura de la izquierda está generada utilizando los valores específicos ($\beta_{11} = 1$, $\beta_{12} = -1$), la segunda presenta una contaminación del 1% de los píxeles, la siguiente un 10% y la de la derecha un 20% de los mismos.

Se estimaron los valores de los parámetros de cada una de los modelos, los puros y los contaminados, para cada una de las cuatro situaciones anteriormente mencionadas, a través del método de Máxima Pseudo-verosimilitud (MP) y del método de Mínimos Cuadrados Condicional (MCC).

Con el propósito de evaluar el desempeño de los métodos de estimación, MP y MCC, se exhibe en la tabla 5.2 y 5.3, el valor medio, el sesgo y desvío estándar de las estimaciones obtenidas de cada uno de los parámetros de los modelos H-V e I-D puros, respectivamente.

Tabla 5.2. Resultados de las estimaciones de los parámetros del Modelo Autobinomial H-V Puro.

Parámetro	β_{11}	β_{12}
Valores especificados	0.00	1.00
Método de Mínimos Cuadrados Condicional		
Media	0.00064	1.00412
Sesgo	0.00064	0.00412
D.S.	0.02791	0.02186
Método de Máxima Pseudo-verosimilitud		
Media	0.00055	1.00225
Sesgo	0.00055	0.00225
D.S.	0.01949	0.01811
Valores especificados	1.00	1.00
Método de Mínimos Cuadrados Condicional		
Media	1.00545	1.00590
Sesgo	0.00545	0.00590
D.S.	0.02404	0.03892
Método de Máxima Pseudo-verosimilitud		
Media	1.00335	1.00204
Sesgo	0.00335	0.00204
D.S.	0.01389	0.01203
Valores especificados	1.00	-1.00
Método de Mínimos Cuadrados Condicional		
Media	1.00292	-1.00283
Sesgo	0.00292	-0.00283
D.S.	0.01494	0.01119
Método de Máxima Pseudo-verosimilitud		
Media	1.00154	-1.00189
Sesgo	0.00154	-0.00189
D.S.	0.01066	0.00908
Valores especificados	-1.00	1.00
Método de Mínimos Cuadrados Condicional		
Media	-1.00389	1.00273
Sesgo	-0.00389	0.00273
D.S.	0.01077	0.01279
Método de Máxima Pseudo-verosimilitud		
Media	-1.00261	1.00198
Sesgo	-0.00181	0.00198
D.S.	0.01062	0.01171

En las figuras 5.3 y 5.4 se presentan imágenes, en escala de grises, correspondientes a valores específicos de los parámetros de los modelos puros y

texturas generadas usando los valores estimados de los parámetros mediante ambos métodos.

Tabla 5.3. Resultados de las estimaciones de los parámetros del Modelo Autobinomial I-D Puro.

Parámetro	β_{21}	β_{22}
Valores especificados	0.00	1.00
Método de Mínimos Cuadrados Condicional		
Media	- 0.00168	1.00802
Sesgo	- 0.00168	0.00802
D.S.	0.03151	0.03172
Método de Máxima Pseudo-verosimilitud		
Media	- 0.00128	1.00765
Sesgo	- 0.00128	0.00765
D.S.	0.02721	0.02337
Valores especificados	1.00	1.00
Método de Mínimos Cuadrados Condicional		
Media	1.00355	1.00115
Sesgo	0.00355	0.00115
D.S.	0.00441	0.01798
Método de Máxima Pseudo-verosimilitud		
Media	1.00328	1.00105
Sesgo	0.00328	0.00105
D.S.	0.00396	0.00810
Valores especificados	1.00	-1.00
Método de Mínimos Cuadrados Condicional		
Media	1.00309	-1.00296
Sesgo	0.00309	-0.00296
D.S.	0.01264	0.01300
Método de Máxima Pseudo-verosimilitud		
Media	1.00232	-1.00223
Sesgo	0.00232	-0.00223
D.S.	0.00849	0.01007
Valores especificados	- 1.00	1.00
Método de Mínimos Cuadrados Condicional		
Media	-1.00206	1.00233
Sesgo	- 0.00206	0.00233
D.S.	0.00888	0.00847
Método de Máxima Pseudo-verosimilitud		
Media	- 1.00201	1.00224
Sesgo	- 0.00201	0.00224
D.S.	0.00827	0.00801

No se vislumbran diferencias significativas respecto a los resultados numéricos obtenidos a través de ambos métodos, en los distintos modelos. De las imágenes de 32 x 32 píxeles, puede advertirse, que las texturas presentan una estructura bastante similar.

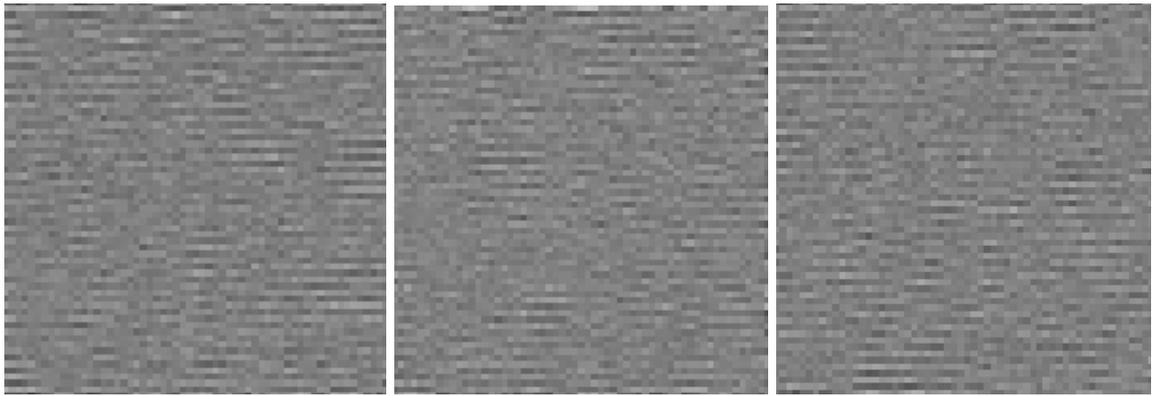


Figura 5.3 Modelo Autobinomial H-V. La textura de la izquierda está generada utilizando los valores específicos ($\beta_{11} = 1$, $\beta_{12} = -1$), la del medio usando los valores de los parámetros estimados a partir del método MCC y la de la derecha usando el método MP.

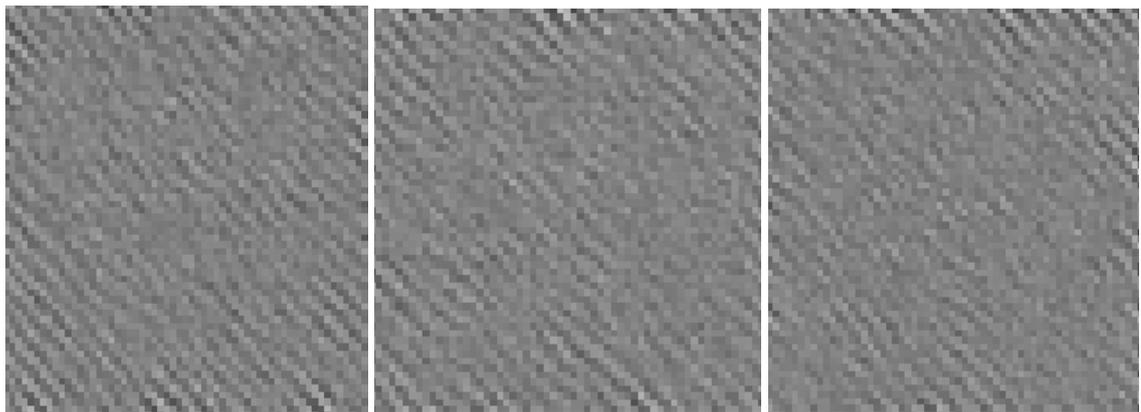


Figura 5.4 Modelo Autobinomial I-D. La textura de la izquierda está generada utilizando los valores específicos ($\beta_{21} = -1$, $\beta_{22} = 1$), la del medio usando los valores de los parámetros estimados a partir del método MCC y la de la derecha usando el método de MP.

El comportamiento de los sesgos y las desviaciones estándar de las estimaciones queda reflejado en las figuras 5.5 y 5.6, respectivamente. Las magnitudes de los mismos, en algunas de las situaciones consideradas, son relativamente altas para ambos métodos. Esto no es debido a que los estimadores sean incorrectos, sino es posible que correspondan a texturas generadas a una iteración del algoritmo GS donde todavía no se ha llegado a la convergencia o al hecho de que los distintos modelos estén compuestos únicamente por dos parámetros para caracterizar las texturas. A partir de los gráficos puede vislumbrarse que las estimaciones obtenidas a partir del método MCC presentan más sesgo y mayor dispersión que las

logradas por el de MP; esto indicaría que éste último resulta ser levemente más eficiente que el método MCC.

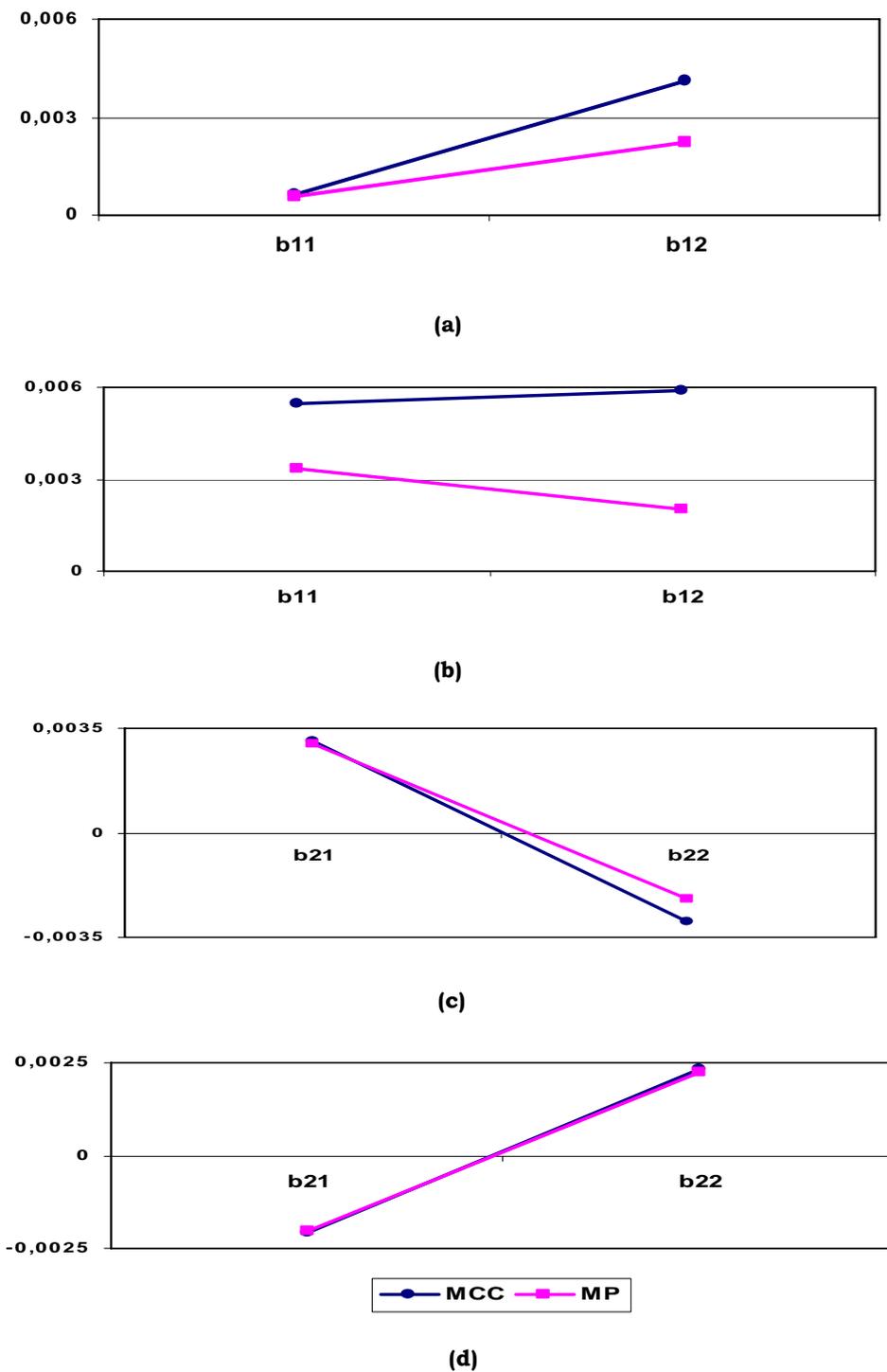


Figura 5.5 Sesgos de las estimaciones de los parámetros de los Modelos Autobinomiales H-V e I-D, realizadas con los distintos métodos. Modelo H-V: **(a)** para los valores específicos ($\beta_{11} = 0$, $\beta_{12} = 1$), **(b)** para ($\beta_{11} = 1$, $\beta_{12} = 1$). Modelo I-D: **(c)** Para los valores específicos ($\beta_{21} = 1$, $\beta_{22} = -1$), **(d)** para ($\beta_{21} = -1$, $\beta_{22} = 1$).

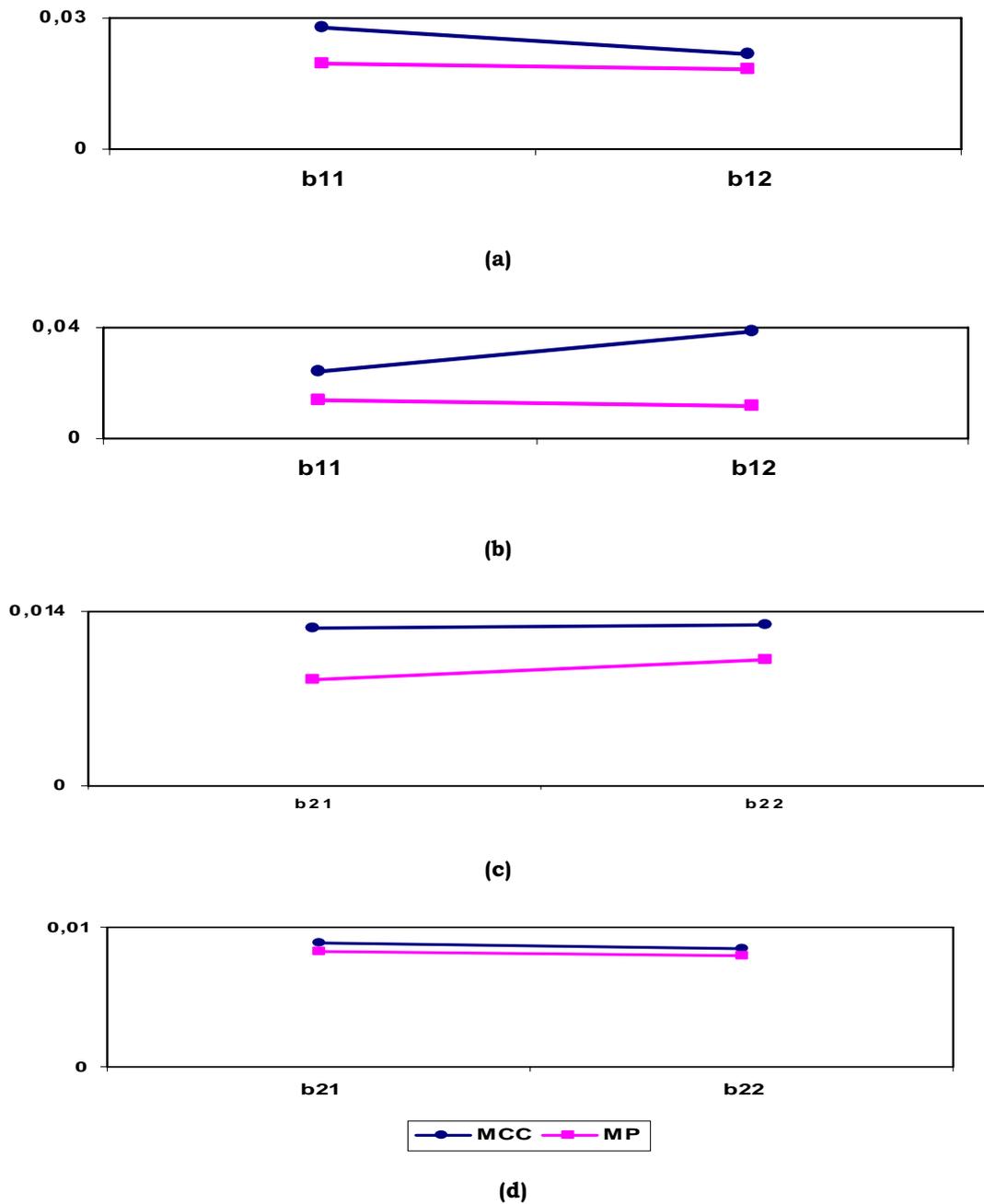


Figura 5.6 Desvíos estándar de las estimaciones de los parámetros de los Modelos Autobinomiales H-V e I-D, realizadas con los distintos métodos. Modelo H-V: **(a)** para los valores específicos ($\beta_{11} = 0, \beta_{12} = 1$), **(b)** para ($\beta_{11} = 1, \beta_{12} = 1$). Modelo I-D: **(c)** para los valores específicos ($\beta_{21} = 1, \beta_{22} = -1$), **(d)** para ($\beta_{21} = -1, \beta_{22} = 1$).

En lo que respecta a los modelos contaminados algunos de los resultados fueron los siguientes:

Tabla 5.4. Resultados de las estimaciones de los parámetros del Modelo Autobinomial H-V contaminado.

Parámetro	β_{11}	β_{12}
Contaminación al 1%		
Valores especificados	0.00	1.00
Método de Mínimos Cuadrados Condiciona		
Media	0.00786	1.00407
Sesgo	0.00786	0.00407
D.S.	0.04587	0.04627
Método de Máxima Pseudo-verosimilitud		
Media	0.00423	1.00285
Sesgo	0.00423	0.00285
D.S.	0.04509	0.03320
Contaminación al 10%		
Valores especificados	1.00	1.00
Método de Mínimos Cuadrados Condiciona		
Media	1.05999	1.03546
Sesgo	0.05999	0.03546
D.S.	0.13629	0.13566
Método de Máxima Pseudo-verosimilitud		
Media	1.01866	1.01041
Sesgo	0.01866	0.01041
D.S.	0.08871	0.08326
Contaminación al 20%		
Valores especificados	1.00	-1.00
Método de Mínimos Cuadrados Condiciona		
Media	0.78409	0.21789
Sesgo	-0.21591	1.21789
D.S.	0.17162	0.17198
Método de Máxima pseudo-verosimilitud		
Media	0.82868	-0.21906
Sesgo	- 0.17132	0.78094
D.S.	0.04410	0.28438

En la tabla 5.4 como en la 5.5, puede advertirse que a medida que aumenta el porcentaje de píxeles contaminados en la imagen de 32 x 32, las estimaciones se van distorsionando. Al 1%, las estimaciones de los parámetros de ambos modelos, no presentan diferencias significativas respecto a los valores especificados. Al 10%, los sesgos de las estimaciones comienzan a incrementarse levemente. Al 20%, las magnitudes de las estimaciones fueron notablemente diferentes respecto de los valores especificados. Revelando una alteración substancial.

Pudo comprobarse, en ambos modelos, que el contaminar el 20% de los píxeles de las texturas generadas con los valores de parámetros $\beta_{i1} = 1$, $\beta_{i2} = -1$ ó $\beta_{i1} = -1$, $\beta_{i2} = -1$ $i = 1, 2$, las estimaciones se vieron más afectadas que para los casos $\beta_{i1} = 0$, $\beta_{i2} = 1$ ó $\beta_{i1} = 1$, $\beta_{i2} = 1$, $i = 1, 2$.

Tabla 5.5. Resultados de las estimaciones de los parámetros del Modelo Autobinomial I-D contaminado.

Parámetro	β_{21}	β_{22}
Contaminación al 1%		
Valores especificados	0.00	1.00
Método de Mínimos Cuadrados Condicional		
Media	0.00545	0.99941
Sesgo	0.00545	-0.00059
D.S.	0.02331	0.03907
Método de Máxima Pseudo-verosimilitud		
Media	0.00305	1.00152
Sesgo	0.00305	0.00152
D.S.	0.02041	0.01327
Contaminación al 10%		
Valores especificados	1.00	1.00
Método de Mínimos Cuadrados Condicional		
Media	1.09570	1.03989
Sesgo	0.09570	0.03989
D.S.	0.15579	0.18268
Método de Máxima Pseudo-verosimilitud		
Media	1.05143	1.02508
Sesgo	0.05143	0.02508
D.S.	0.10117	0.14774
Contaminación al 20%		
Valores especificados	1.00	-1.00
Método de Mínimos Cuadrados Condicional		
Media	0.70979	0.14910
Sesgo	-0.29021	-1.14910
D.S.	0.06814	0.18521
Método de Máxima Pseudo-verosimilitud		
Media	0.81230	-0.23232
Sesgo	-0.1877	0.76768
D.S.	0.03654	0.11383

Al evaluar las estimaciones realizadas de los parámetros de los modelos Autobinomiales contaminados, se advierte que el método MCC resulta ser más sensible que el MP a la contaminación a diferentes patrones homogéneos de textura, debido a que las estimaciones obtenidas a partir del primero se ven fuertemente afectadas sobre todo cuando se contamina el 20% de los píxeles de la textura original. Si bien ambos manifiestan incrementos en los sesgos de las estimaciones a medida que aumenta el grado de contaminación, las estimaciones obtenidas por el método MP en los modelos contaminados, mantienen un sesgo y un desvío estándar menor que las alcanzadas con el MCC.

5.1.1 Visualización del efecto de contaminación

Si se considera ahora al vector de parámetros θ . De las 100 estimaciones del vector θ , de cada una de las situaciones de los distintos modelos, puros y contaminados, se determinaron los vectores de valores medios y las matrices de varianza-covarianza (Apéndice C).

A modo de comparar visualmente el desempeño de los dos métodos de estimación y la influencia de los distintos grados de contaminación en la estimación de los parámetros de los modelos H-V e I-D, se construyeron regiones de confianza de 95%, cuya expresión esta dada por:

$$n \underset{\approx}{(\bar{\mathbf{x}} - \boldsymbol{\theta})}^T \underset{\approx}{\mathbf{S}}^{-1} \underset{\approx}{(\bar{\mathbf{x}} - \boldsymbol{\theta})} \leq \frac{(n-1)p}{(n-p)} F_{p, n-p}(\alpha) \quad (5.1)$$

donde $\boldsymbol{\theta} = (\beta_{i1}, \beta_{i2})$ $i = 1, 2$, es el vector de parámetros del modelo, p = cantidad de parámetros, $\bar{\mathbf{x}} = (\bar{x}_{i1}, \bar{x}_{i2})$ $i = 1, 2$ y \mathbf{S} , son el vector de valores medios y la matriz de varianza covarianza, respectivamente. $F_{p, n-p}(\alpha)$ es el percentil ($\alpha 100$) superior de la distribución F con p y $n - p$ grados de libertad.

En la determinación de las elipses de confianza para estimar los vectores $\boldsymbol{\theta}_{\text{H-V}} = (\beta_{11}, \beta_{12})$ y $\boldsymbol{\theta}_{\text{I-D}} = (\beta_{21}, \beta_{22})$, se aplicó el Teorema Central del Límite Multivariado a las texturas muestreadas para cada situación y para cada modelo. Los resultados más relevantes se muestran en las siguientes figuras.

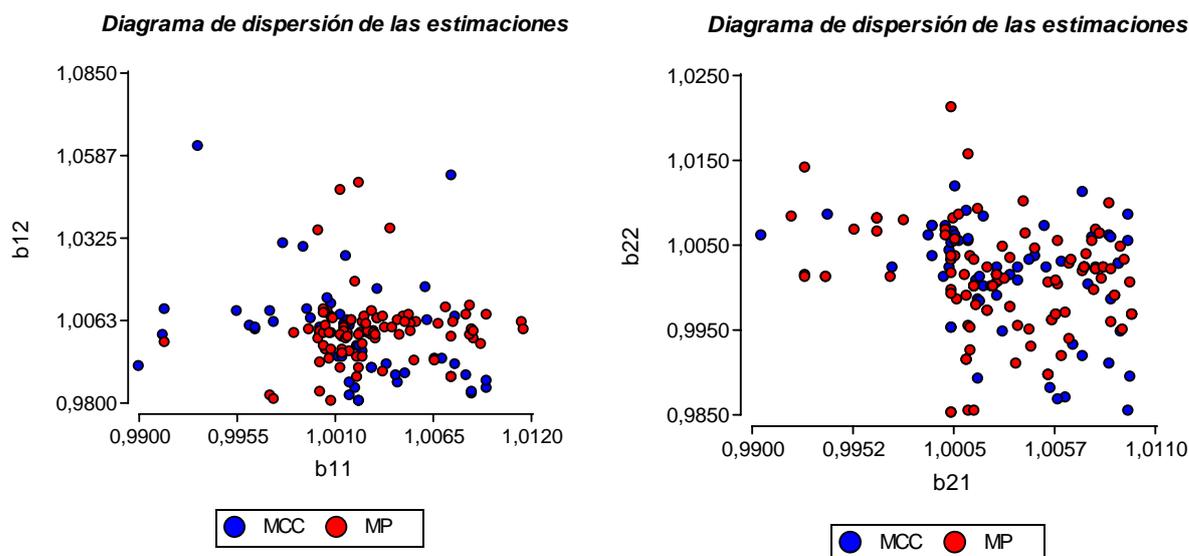


Figura 5.7 Comparación de los métodos de estimación MCC y MP: a izquierda, el Diagrama de dispersión de las estimaciones del vector de parámetros $\theta_{H-V} = (\beta_{11}, \beta_{12})$ del Modelo Autobinomial H-V Puro, para el vector de valores especificados $\theta_{H-V} = (1, 1)$. A derecha, el Diagrama de dispersión de las estimaciones del vector de parámetros $\theta_{I-D}=(\beta_{21}, \beta_{22})$ del Modelo Autobinomial I-D Puro, para el vector de valores especificados $\theta_{I-D}=(1, 1)$.

En la figura 5.7 de la izquierda puede advertirse que las estimaciones logradas a partir del método MP presentan una menor variabilidad que las obtenidas a través del Método MCC. Situación que se confirma al construir las elipses de confianza del 95% (figura 5.8 de la izquierda) para estimar el vector de parámetros $\theta_{H-V} = (\beta_{11}, \beta_{12})$, del modelo Autobinomial H-V Puro, considerando el caso del vector $\theta_{H-V} = (1, 1)$. Evidenciando (según los tamaños de las elipses) que el estimador MP alcanza una mayor precisión en las estimaciones del vector que las logradas por el MCC. Mientras que en el caso de las estimaciones del vector $\theta_{I-D}=(\beta_{21}, \beta_{22})$ del modelo Autobinomial I-D Puro, las estimaciones logradas por ambos estimadores presentan una variabilidad similar. Si bien en el diagrama de dispersión no se perciben diferencias significativas entre MCC y MP, en las elipses de confianza (figura 5.8 derecha) queda evidenciado que el método MP alcanza una precisión levemente superior.

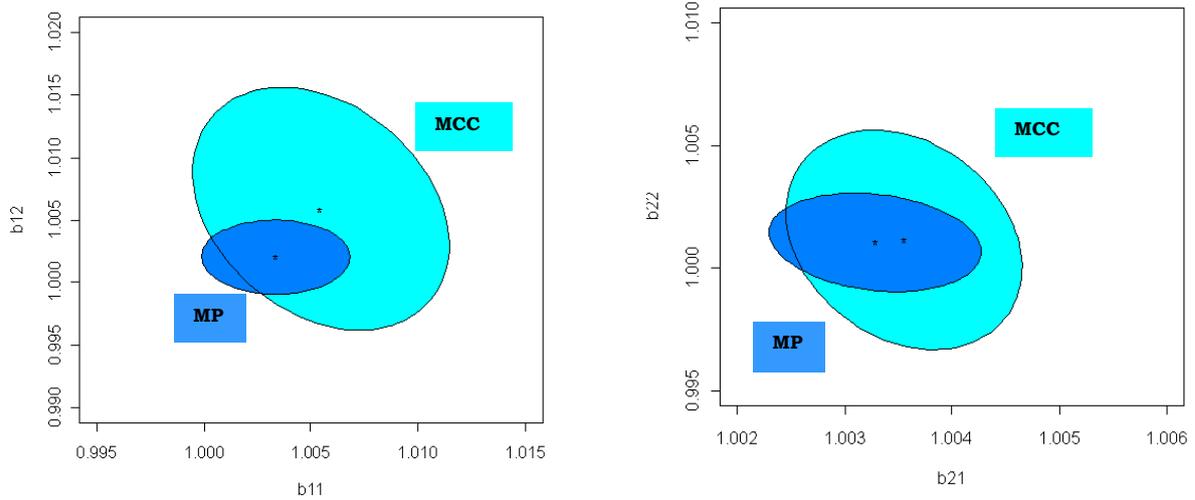


Figura 5.8 Comparación de los métodos de estimación MCC y MP: elipses de confianza del 95% para estimar el vector $\theta_{H-V} = (\beta_{11}, \beta_{12})$ y $\theta_{I-D} = (\beta_{21}, \beta_{22})$, respectivamente. Considerando los casos: $\theta_{H-V} = (1, 1)$ y $\theta_{I-D} = (1, 1)$, respectivamente.

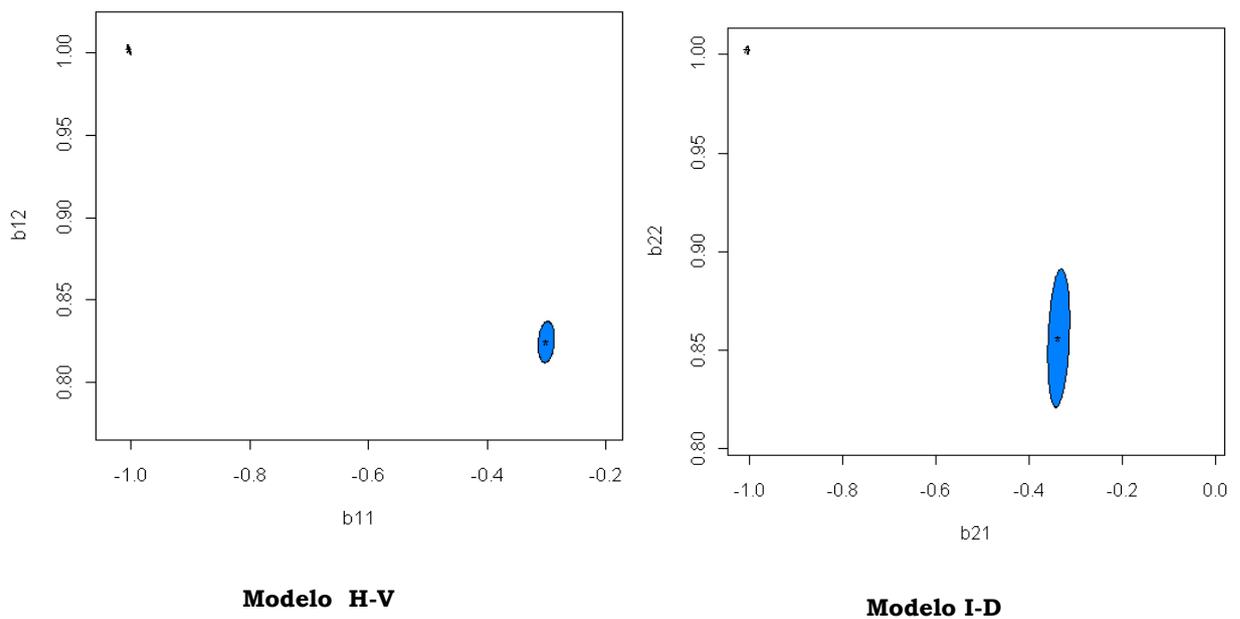


Figura 5.9 Comparación de los modelos H-V e I-D frente a un 20 % de contaminación de los píxeles de la textura: elipses de confianza del 95% para estimar el vector $\theta_{H-V} = (\beta_{11}, \beta_{12})$ y $\theta_{I-D} = (\beta_{21}, \beta_{22})$. Considerando el caso $(-1, 1)$, para los vectores de parámetros de ambos modelos.

A partir de las elipses de la figura 5.9 no se vislumbran diferencias significativas en lo que respecta a los dos modelos considerados frente a un 20 % de contaminación. Si se percibe que en el modelo I-D, las estimaciones de los parámetros de las

texturas contaminadas presentan una mayor variabilidad que las logradas a partir del modelo H-V. Situación que se ve reflejada en la forma y tamaño de las elipses. En todos los casos exhibidos en la figura 5.9, se utilizó el método MP para determinar las estimaciones, por presentar una mejor performance.

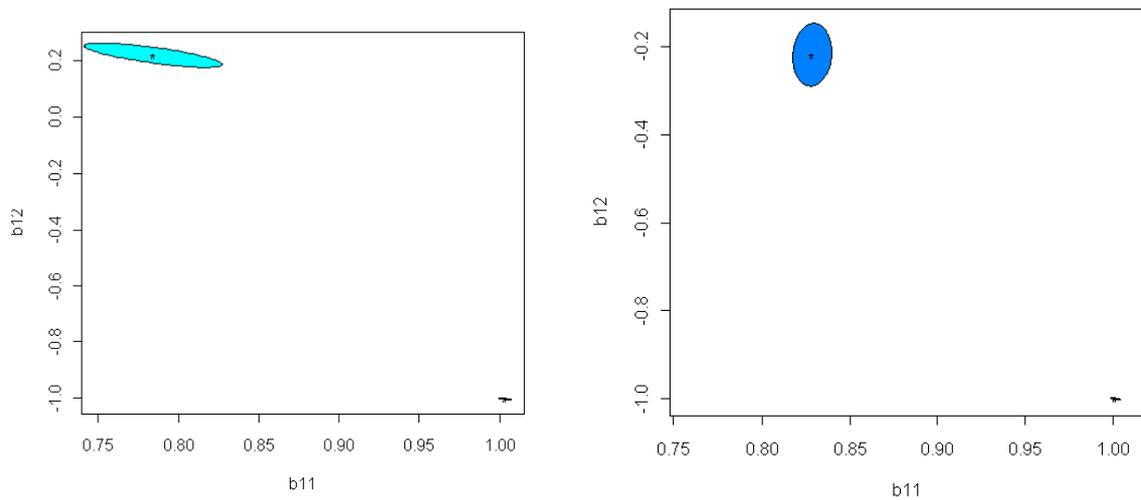


Figura 5.10 Comparación de los estimadores MCC y MP frente a un 20 % de contaminación de los píxeles de la textura: elipses de confianza del 95% para estimar el vector $\theta_{H-V} = (\beta_{11}, \beta_{12})$. Considerando el caso $\theta_{H-V} = (1, -1)$.

Si ahora se comparan las elipses de confianza obtenidas a través de las estimaciones del MCC y del MP, para estimar el vector $\theta_{H-V} = (\beta_{11}, \beta_{12})$ del modelo Autobinomial H-V, puede advertirse que el MP resulta ser más eficiente que el MCC. Esto se manifiesta en el tamaño de las elipses y en el alejamiento existente entre las construidas mediante el modelo puro comparadas con las del contaminado el 20%. Se percibe que las estimaciones logradas por el método MCC han sido afectadas en gran medida por este grado de contaminación. Resulta ser el método más sensible ante distintos grados de contaminación a patrones homogéneos de textura.

CLASIFICACIÓN SUPERVISADA DE UNA IMAGEN SATELITAL

En este capítulo primeramente se ilustrarán algunos conceptos básicos concernientes a la formación de imágenes de teledetección, principalmente desde sensores ubicados en plataformas espaciales. Posteriormente se describirán brevemente las técnicas de clasificación de imágenes digitales. Finalmente, se comparará la capacidad de los métodos de estimación *de Mínimos Cuadrados Condicional* y el *Máxima Pseudo-verosimilitud* y la de los modelos Autobinomiales en la clasificación supervisada de las diferentes texturas existentes en una imagen satelital concreta.

6.1 Teledetección espacial

La teledetección es una perfecta conjunción de ciencias y tecnologías que se han desarrollado a lo largo de los dos últimos siglos y que han dado lugar a una técnica multidisciplinar basada en el conocimiento de la Física, las Matemáticas y que se ha desarrollado, principalmente, gracias a los avances tecnológicos en telecomunicación y computación.

La Teledetección (Remote Sensing), podría definirse como una técnica que consiste en la adquisición de información acerca de las características físicas y biológicas de objetos de la superficie terrestre, mediante mediciones realizadas a distancia por sensores instalados en plataformas espaciales o aerotransportados, y su posterior análisis por medio de procesamiento digital e interpretación de imágenes. La información obtenida producto de la medición de la interacción electromagnética entre la Tierra y el sensor se resume en matrices, dando lugar a lo que se denomina imagen satelital. Este trabajo se centrará su atención en datos provenientes desde plataformas espaciales: teledetección espacial.

Un sistema de teledetección espacial está constituido básicamente, por los siguientes elementos: la fuente de energía, la cubierta terrestre, el sistema sensor, el tratamiento de la información.

- *Fuente de energía.* La primera de las componentes del sistema es la fuente generadora de energía o de radiación electromagnética detectada por el sensor, que puede ser radiada por los objetos, o reflejada por la superficie terrestre. En este último caso, la fuente de radiación puede ser el Sol, en cuyo caso se habla de *teledetección pasiva*, o el propio sensor, conocida como *teledetección activa*.
- *Cubierta terrestre.* Constituida por distintas masas de vegetación, suelo, agua y construcciones, que reciben la señal energética, y la reflejan o emiten de acuerdo a sus características.
- *Sistema sensor.* Constituido por el sensor propiamente dicho y la plataforma satelital (denominada satélite) en la que se instala. Capta la energía reflejada en las diferentes cubiertas terrestres, luego de atravesar la atmósfera, y transmite la información, como señal digital, en forma matricial, al sistema receptor en la Tierra.
- *Sistema de transmisión, recepción y procesamiento.* Encargado de recibir la información transmitida por el sensor, grabarla en un formato apropiado (analógico o digital) y distribuirla a los intérpretes para la aplicación de procedimientos de tratamiento visual o digital, obteniendo de esta forma, nuevos datos.
- *Interprete* Transforma los datos en información temática orientada a la finalidad del estudio, ya sea de forma visual o de forma digital. La interpretación visual pretende extraer información espacial a partir del análisis de la imagen y sus cualidades (brillo, color textura, forma, tamaño de los elementos que la componen), mientras que el tratamiento digital, supone la extracción de la información haciendo uso de técnicas matemáticas y estadísticas mediante algoritmos en programas informáticos.

- *Usuario final.* Encargado de analizar el documento fruto de la interpretación, extraer la información necesaria para resolver una problemática específica y de dictaminar el impacto que de él se derive.

Un ejemplo de este proceso puede verse en la siguiente figura 6.1.

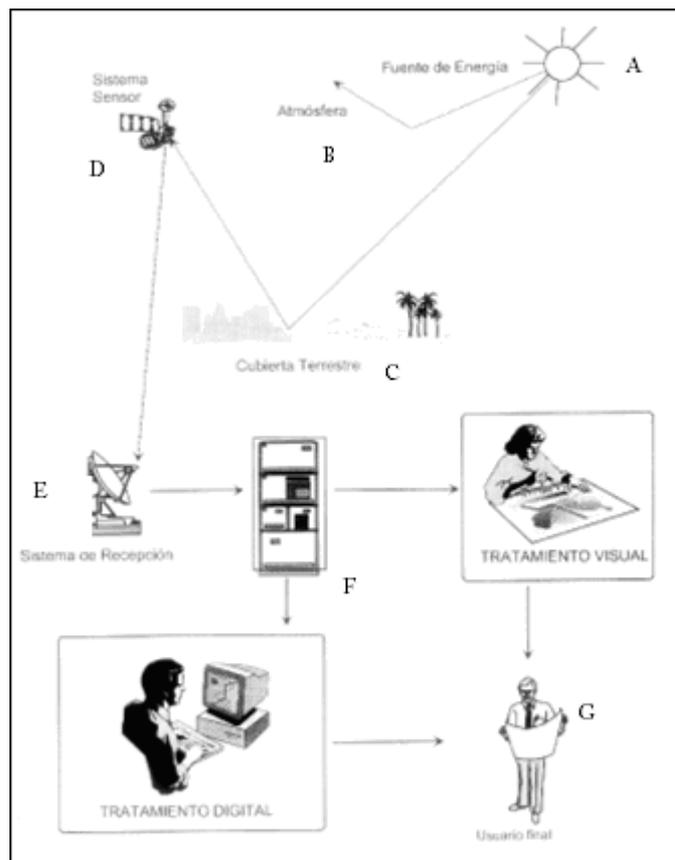


Figura 6.1 Los componentes de un sistema de Teledetección espacial:
 (A) Fuente de energía. (B) Atmósfera. (C) Cubierta terrestre.
 (D) Sensor. (E) Sistema de Transmisión, recepción y procesamiento.
 (F) Interprete. (G) Usuario.

Las características fundamentales de los datos adquiridos (la radiación, emitida o reflejada, de la superficie terrestre) en un sistema de teledetección espacial, pueden definirse a través de cuatro parámetros:

- *Resolución espacial:* que hace referencia al tamaño del objeto más pequeño que identifica el sensor. Ese tamaño corresponde a la mínima unidad de información que aparece en la imagen, que se denomina *píxel* (picture element). La resolución espacial de los sensores actuales varía desde los 4,5 km a 1,1 m. Por ejemplo la resolución del Landsat 7 ETM es de 30 m, es decir cada píxel corresponde en el terreno a una superficie de un cuadrado de 30 m x 30 m.

- *Resolución Espectral:* representa el número y el ancho de las bandas espectrales que puede discriminar o registrar el sensor. Cuanto mayor sea el número de bandas que incluye un sensor, mejor, puesto que cada banda, constituye una variable para caracterizar la superficie captada. También es importante que estas bandas sean estrechas para que aumente su poder discriminante, caso contrario, recogerán valores promediados que ocultarán elementos de diferenciación. El número de bandas y su localización en el espectro depende de los objetivos que se pretendan alcanzar.
- *Resolución Radiométrica:* hace referencia a la sensibilidad que tiene el sensor de detectar variaciones en la magnitud de medida de energía que recibe, es decir, al número de niveles digitales (ND) utilizados para expresar la intensidad de luz captada por el sensor. Este número suele expresarse en términos binarios (bits) por píxel. Por ejemplo, Landsat-TM utiliza 8 bits, lo que da 256 niveles de energía (0 - 255), que pueden ser captados. A mayor resolución radiométrica mayor número de detalles podrán captarse en la imagen y mejor será su interpretación.
- *Resolución Temporal:* La periodicidad con la que el sensor adquiere imágenes de la misma porción de la superficie terrestre, “período de revisita”.

El aumento en la resolución espacial supone una disminución en la resolución temporal y espectral.

Entre las principales aplicaciones de la teledetección pueden mencionarse:

- *Forestales:* Identificación de áreas de bosques, alteraciones forestales en parques, reservas, bosques nativos o plantados, mediciones de área, ubicación de incendios forestales etc.
- *Agrícolas:* Identificación de zonas agrícolas, medición de áreas, monitoreo del desarrollo y expansión de la frontera agrícola, etc.
- *Ambientales:* Identificación de anomalías antrópicas a lo largo de los cursos de agua, reservorios, bosques, zonas de crecimiento urbano, cartas de uso del suelo, expansión de áreas urbanas, etc.
- *Hidrológicas:* Identificación de límites costeros, represas, tajamares, ubicación de acuíferos etc.
- *Cartografía:* Dado que posibilita la obtención de pares espectroscópicos permite el análisis cartográfico y la producción de cartas imágenes con mosaicos municipales y provinciales.

- *Geología y Suelos*: Apoyo a levantamientos de suelos y mapas geológicos.

Existe también un uso intensivo de la teledetección, dentro del área militar y de la seguridad.

La teledetección, con el avance de las tecnologías en estos últimos años, ha facilitado la obtención de imágenes de la tierra desde sensores instalados en satélites, permitiendo una mayor comprensión acerca de las propiedades y comportamiento de la naturaleza, logrando prever situaciones de riesgos ambientales. Se ha convertido, en la actualidad, en la herramienta esencial que permite la adquisición de imágenes de forma casi inmediata y a cualquier escala; brindando una enorme cantidad de datos geográficos, espaciales, espectrales, etc., de gran calidad.

6.2 Imágenes digitales

Un sensor cuenta con miles de detectores diminutos encargados de recibir y medir la cantidad de radiación electromagnética reflejada por la superficie terrestre y los objetos que hay en ella, en cierta gama de longitudes de onda. Estas mediciones se denominan *espectrales*. Cada valor de reflectancia espectral se registra como un valor numérico o nivel digital (ND). Toda esta cuantificación se constituye en un arreglo matricial, que es transmitido a la Tierra, donde, a los efectos de su visualización, se traduce en una escala de grises (en la que el negro representa ausencia de energía y el blanco es el máximo de energía recibida) o grado de brillo de color, conformarse así, lo que se denomina una *imagen satelital*.

Las imágenes creadas a partir de éstas medidas de reflectancia, ofrecen una representación visual extremadamente precisa de las características y detalles de la superficie de la tierra y de los objetos implantados en ella, en términos de forma, dimensión, color y apariencia general, es decir, del *contenido espacial* de la imagen de satélite. Las imágenes satelitales también ponen de manifiesto otros rasgos y peculiaridades que son imperceptibles por el ojo humano, como por ejemplo, contenido mineral de las rocas, la humedad del suelo, la densidad del agua, la composición física de los edificios, etc., denominado *contenido espectral* de la imagen de satélite.

La diferencia entre información espacial y espectral determina el tipo de imagen de satélite: pancromática o multiespectral. Las *imágenes pancromáticas* se captan

mediante un sensor digital que mide la reflectancia de energía en una amplia parte del espectro electromagnético. En los sensores pancromáticos más modernos, esta única banda suele abarcar la parte visible y de infrarrojo cercano del espectro. Los datos pancromáticos se representan por medio de la escala de gris monocromática. Son imágenes en blanco y negro. Las imágenes multispectrales se captan mediante un sensor digital que mide la reflectancia en muchas bandas, que pueden variar desde tres a ocho. Cada píxel de una imagen multispectral, tiene un color que se crea combinando niveles de brillo que se corresponden con los valores de la reflectancia en las diferentes bandas. La imagen multispectral está constituida por tantas imágenes como bandas espectrales proporcione el sensor de una misma zona de la superficie terrestre.

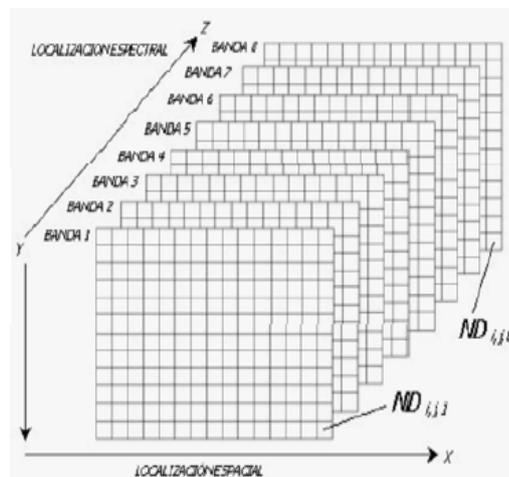


Figura 6.2 Organización de datos de ocho bandas espectrales sobre una misma escena.

Uno de los aspectos más significativos de las imágenes digitales es que pueden ser procesadas, transformadas y realzadas a través de procedimientos computacionales y de técnicas de análisis, para generar una serie de productos de información diferentes a fin de realizar diversas interpretaciones y poder resolver problemáticas específicas concretas.

A continuación se desarrollará brevemente uno de los productos más corrientes que se extrae de las imágenes obtenidas por satélite, denominado mapa de clasificación.

6.3 Clasificación de imágenes

Uno de los objetivos fundamentales de la teledetección es la elaboración de un mapa de clasificación de un área de interés. La clasificación es una técnica que se utiliza

para obtener información, interpretar y conocer el contenido de las imágenes digitales y supone un enfoque cualitativo, mediante el cual permite asignar los píxeles originales de una imagen a un conjunto de categorías o clases, basándose en los niveles digitales de los píxeles en cada una de las distintas bandas. Estas categorías, denominadas *taxones* pueden ser:

- Clase espectral: clase definida a partir de la respuesta espectral.
- Unidades geográficas: categorías definidas por el análisis del paisaje: dunas, playa de arenas, cultivos herbáceos, regadíos, zona urbana, etc.

A través de la clasificación se obtiene una nueva imagen, en la cual cada uno de los píxeles originales quedan definidos por un nivel digital o un color, que los identifica según la clase a la que se los ha incluido. Estas clases pueden describir distintos tipos de cubiertas del terreno (variable nominal o categórica), por ejemplo, una clasificación de zonas boscosas, de campo abierto, urbanas, de aguas, coberturas vegetales, o bien intervalos de una misma categoría de interés (variable ordinal), por ejemplo, un intento de señalar niveles de afectación en un incendio forestal o de humedad en una inundación, etc. Normalmente, los distintos tipos de terreno están codificados por colores.

La técnica de clasificación digital consta de tres fases:

1. *Fase de entrenamiento*: definición digital de las categorías.
2. *Fase de asignación*: agrupación de los píxeles de la imagen en una de las categorías previamente definidas.
3. *Comprobación y verificación de los resultados*.

La clasificación digital puede ser supervisada o no supervisada.

Clasificación Supervisada

El método supervisado parte de un cierto conocimiento de la zona de estudio, adquirido por experiencia previa o por trabajos de campo. Este conocimiento previo de la zona de interés, permite al interpretador delimitar sobre la imagen cierto número de áreas piloto, denominadas *campos ó áreas de entrenamiento (training fields)*, que se consideran, suficientemente representativas de las categorías que se pretenden discriminar. El término “entrenamiento” se debe a que estas áreas le sirven al algoritmo computacional para reconocer las distintas categorías. Para realizar la clasificación es necesario introducirle al programa informático una muestra de píxeles que representen adecuadamente a cada categoría. Éste, en base

a sus valores digitales calcula diversos estadísticos: media, rango, desviación estándar, matriz de varianza – covarianza, etc., que caracterizan cada clase, para luego asignar automáticamente los píxeles restantes a las distintas clases en función de sus niveles digitales. Esta asignación se establece mediante algoritmos de clasificación. Existen diversos algoritmos de clasificación de la escena, entre ellos se encuentran:

- *Clasificador de Paralelepípedos:* En este método se definen una serie de paralelepípedos (hiper-rectángulos) que constituyen las clases. Las fronteras de cada una de estas clases se determinan, teniendo en cuenta los valores máximos y mínimos de reflectividad de cada área de entrenamiento. Todos aquellos píxeles que se encuentren dentro del hipercubo son asignados a esa clase. Es un clasificador rápido y simple de implementar, pero suelen presentarse problemas que le quitan robustez; pueden existir píxeles sin clasificar o píxeles asignados a dos o más categorías.
- *Clasificador de Distancia Mínima.* Esta clasificación consiste en la determinación de las medias (centroides) de los ND de cada clase y cada píxel es asignado a la clase cuya media espectral es más cercana al ND del píxel.
- *Clasificador de Máxima probabilidad.* Se considera que los ND de cada categoría se ajustan a una distribución normal. Esto permite describir esa categoría a través de una función de probabilidad, a partir de su vector de medias y matriz de varianza-covarianza. Se calcula la probabilidad de que un determinado píxel pertenezca a una categoría; el cálculo se realiza para todas las categorías existentes y el píxel se incluirá en aquella clase a la cual tenga más probabilidad de pertenecer.
- *Clasificador de Mínima distancia de Mahalanobis.* Es un método similar al de máxima probabilidad, pero asumiendo que las covarianzas de las clases son iguales.

Otras tendencias (Solaiman, Mouchot y Maillard, 1994) resaltan la capacidad que poseen las “Redes neuronales artificiales” para la clasificación de imágenes multispectrales. El método de clasificación empleando redes neuronales fue desarrollado simultáneamente en los ámbitos del análisis estadístico y de la inteligencia artificial. Las redes neuronales constituyen una herramienta valiosa para clasificar imágenes satelitales pues no requieren hipótesis sobre la

distribución de los datos. Consiste en extraer combinaciones lineales de los atributos presentes en los objetos obteniendo una serie de características y modelizar las clases como funciones no lineales de dichas características.

Autores como Chuvieco (1996), señalan otros clasificadores como el “*Clasificador del árbol*”. Es un algoritmo de clasificación y regresión basado en la construcción de un árbol, en el que se parte de la raíz y a partir de los valores de la entrada se profundiza hacia los nodos hoja, eligiendo un camino u otro dependiendo de los valores de la entrada. Cada nodo hoja representa una clase de la clasificación.

Se ha propuesto, también, el uso de la Teoría de la Lógica Difusa (Duda, Canty y Klaus, 1999) para solventar el problema de la existencia de zonas de transición o usos mezclados en la superficie terrestre. Es decir, los *Clasificadores difusos* son empleados cuando se examinan áreas heterogéneas donde con resoluciones medias, no puede asegurarse que un píxel dado contenga una sola clase de cobertura. En este caso, el píxel no pertenece totalmente a una sola categoría sino que tiene distintos grados de pertenencia (comprendidos entre 0 y 1), lo que permite una asignación simultánea a varias categorías.

Clasificación No Supervisada

Este enfoque se utiliza cuando no se conocen "a priori", las características del terreno, o porque las características espectrales de la superficie no están claramente definidas, por lo que, la intervención humana se centra en la interpretación de los resultados.

Este tipo de clasificación es una técnica de “agrupamiento”, que no necesita del proceso de muestreo de la imagen, por cuanto consiste en una búsqueda automática de grupos (clusters) de píxeles de la imagen con un comportamiento espectral homogéneo, de acuerdo a algún criterio estadístico. De esta forma, la imagen es segmentada en clases desconocidas, que generalmente representan las clases espectrales naturales. Para definir los grupos con un comportamiento espectral similar se utilizan, en general técnicas de clasificación automática de datos. Estos algoritmos de clasificación sólo requieren de la escena a clasificar y algunos parámetros que limiten el número de clases a identificar. Entre ellos pueden mencionarse:

- *El método de las K-medias*: Consiste en elegir las medias de las clases uniformemente distribuidas en el espacio de datos y luego de forma iterativa, los píxeles son asignados a las clases más cercanas, mediante el empleo de técnicas de mínima distancia. En cada iteración se recalcula la media de la clase, con los píxeles que fueron incluidos en el cluster y se vuelven a reclasificar todos los píxeles. Este proceso se detiene una vez que el número de píxeles que cambia de cada clase sea menos que el límite de cambio de píxeles seleccionado o cuando se haya alcanzado el número máximo de iteraciones fijado. Una de las desventajas que presenta esta técnica es una alta dependencia en la elección de los centros iniciales.
- *ISODATA* (Interactive Self-Organizing Data Analysis Technique (Algorithm)) (Richards, 1995). Es un método iterativo que requiere que se determine a priori el número máximo de clusters que el algoritmo debe identificar. Consiste en calcular los promedios de las clases o agrupaciones que eventualmente están distribuidos en el espacio de decisión. En cada una de las iteraciones se recalcula los promedios y reclasifica los píxeles de la imagen con respecto a los nuevos promedios. Este proceso continúa hasta que el cambio del número de píxeles en cada cluster es mínimo respecto a un valor de tolerancia o hasta que se realiza cierto número de iteraciones predefinidas.

Con posterioridad, el usuario debe identificar a qué categoría geográfica corresponde cada una de las clases espectrales definidas, es decir, qué cluster se corresponde con vegetación, suelos, área urbana, etc.

El método supervisado pretende definir clases informacionales, mientras que el no supervisado tiende a identificar las clases espectrales presentes en la imagen, aunque, ninguno de los dos, puede considerarse completamente efectivo.

Evaluación de la clasificación

Para evaluar calidad del proceso de clasificación de una imagen en términos de los errores que ha cometido el clasificador, al confundir las muestras de una clase con las de otra, y por tanto, conocer qué clases han sido mejor y peor identificadas se construye la denominada *Matriz de Confusión*, también llamada *Matriz de Error* o *Matriz de Contingencia* de las clases. Se trata de una matriz cuadrada de orden “n”, donde “n” indica el número de categorías. (Tabla 6.1) Usualmente, las filas de esta

matriz, son ocupadas por la categorización de los píxeles deducida por el clasificador, mientras que las columnas corresponden a las clases referencia establecidas por observaciones del terreno. Cada elemento, de esta matriz, debe interpretarse como el número de píxeles que han sido adjudicados a la clase fila, perteneciendo a la clase columna. La diagonal expresa el número de puntos clasificados donde se produce un total acuerdo entre imagen y clasificación, mientras que los marginales suponen errores de asignación.

Tabla 6.1 Matriz de confusión. Efectividad de la clasificación

Datos de Referencia	
Datos de	..
C ₁	..
⋮	..
C _k	..
Total	..

donde

n_{ii} representa el número de píxeles de la imagen satelital que fueron clasificados pertenecientes a la clase i , perteneciendo a la clase i ,

n_{i+} : totales marginales de la fila i ,

n_{+i} : totales marginales de la columna i ,

n : número total de píxeles de la imagen considerada para el estudio.

A partir de una Matriz de Confusión pueden deducirse varios índices relativos a la exactitud de la clasificación. Los índices se determinan a través de las siguientes expresiones,

$$\% \text{ de Fiabilidad global} = \frac{\text{n}^\circ \text{ de píxeles correctamente clasificados}}{\text{n}^\circ \text{ total de píxeles de referencia}} * 100\% \quad (6.1)$$

El índice (6.1), representa la fiabilidad o efectividad global de la imagen clasificada: establece la relación entre el número de píxeles correctamente asignados y el total.

$$\% \text{ de error por omisión} = \frac{n_{+i} - n_{ii}}{n_{+i}} * 100 \quad (6.2)$$

El índice (6.2), representa el error de omisión: indica el número de píxeles que, perteneciendo a una determinada categoría, no fueron adjudicados a ella.

$$\% \text{ de error de comisión} = \frac{n_{i+} - n_{ii}}{n_{i+}} * 100 \quad (6.3)$$

El índice (6.3), representa el error de comisión: indica el número de píxeles que fueron incluidos a una determinada categoría, perteneciendo realmente a otra.

Desde otro punto de vista, algunos autores cuando se refieren a los errores de “omisión” y “comisión” en una clasificación lo expresan en relación inversa, como exactitud del usuario y del productor.

Resulta de gran interés analizar las relaciones entre las distintas clases. Para este análisis puede utilizarse la técnica multivariada discreta *Kappa* (κ). Este estadístico, mide el grado de ajuste debido exclusivamente a la exactitud de la clasificación, prescindiendo del causado por factores aleatorios. Es un criterio que resume la bondad de la clasificación en un único valor. Se basa en la proporción de coincidencia observada relación a la proporción de coincidencia que se esperaría por azar.

La estimación de κ se obtiene a partir de la fórmula de Hudson y Ramn (1987) (6.2). Su cálculo se basa en la matriz de confusión. Si el estadístico $\hat{\kappa}$ es igual a 1 esto indica un acuerdo pleno, mientras que un valor cercano a 0 sugiere que el acuerdo observado es puramente debido al azar. En el caso en que el acuerdo observado sea inferior al esperado por azar, el estadístico kappa es negativo.

El estimador de máxima verosimilitud de Kappa es:

$$\hat{\kappa} = \frac{p_0 - p_a}{1 - p_a} = \frac{n \sum_{i=1}^k n_{ii} - \sum_{i=1}^k n_{i+} * n_{+i}}{n^2 - \sum_{i=1}^k n_{i+} * n_{+i}}, \quad (6.4)$$

donde p_0 , representa la proporción de concordancia observada, y se calcula

$$p_0 = \sum_{i=1}^k p_{ii},$$

y p_a , la proporción de concordancia esperada por puro azar, y se determina:

$$p_a = \sum_{i=1}^k p_{+i} p_{i+}.$$

La expresión del estimador de la varianza asintótica del estadístico Kappa obtenida a partir del método Delta, es:

$$\hat{\sigma}_{\hat{\kappa}}^2 = 1/n \left[\frac{\varphi_1(1-\varphi_1)}{(1-\varphi_2)^2} + \frac{2(1-\varphi_1)(2\varphi_1\varphi_2 - \varphi_3)}{(1-\varphi_3)^3} + \frac{2(1-\varphi_1)^2(\varphi_4 - 4\varphi_2^2)}{(1-\varphi_2)^4} \right], \quad (6.5)$$

donde

$$\varphi_1 = \frac{1}{n} \sum_{i=1}^k n_{ii}$$

$$\varphi_2 = \frac{1}{n^2} \sum_{i=1}^k n_{i+} * n_{+i}$$

$$\varphi_3 = \frac{1}{n^2} \sum_{i=1}^k n_{ii} (n_{i+} + n_{+i})$$

$$\varphi_4 = \frac{1}{n^3} \sum_{i=1}^k \sum_{j=1}^k n_{ij} (n_{i+} + n_{+i}).$$

Para el testear la significación del coeficiente κ , se realiza un contraste de hipótesis (Fleiss, 1981), siendo la hipótesis nula $\kappa = 0$, y, la alternativa $\kappa \neq 0$, empleando el estadístico: $\frac{|\hat{\kappa}|}{\hat{\sigma}_{\kappa}}$. Si se verifican las condiciones de aproximación de la distribución

Binomial a la distribución Normal, este estadístico se distribuye como una normal tipificada.

A continuación se realizará una aplicación de la técnica de clasificación supervisada a una imagen satelital real, en la que se considerará que los ND de cada categoría se ajustan a una distribución Autobinomial H-V e I-D y se hará uso de los métodos de estimación CMC y MP junto a una métrica de símil aridad, como método de reconocimiento de patrones.

6.4 Aplicación: Clasificación supervisada de una imagen satelital

El objetivo de esta sección es poder discriminar en una imagen satelital, regiones similares a un área de referencia o entrenamiento, mediante la aplicación de los

métodos de estimación de los parámetros del modelo Autobinomial H-V e I-D, el PMV y el CMC, en la clasificación supervisada de datos multiespectrales Landsat 5 TM (Mapeador Temático).

Para el estudio, se utilizó una imagen satelital Landsat 5 TM 232-081, de 399 x 577 píxeles, en falso color (BANDA 3 = VERDE), en formato GeoTIFF, de fecha 17 de noviembre de 2008. La imagen corresponde al embalse Ullum, represa sobre el río San Juan a 768 msnm, localizada en el cañón de la Quebrada de Ullum, a 18 km de la ciudad capital provincial de San Juan.

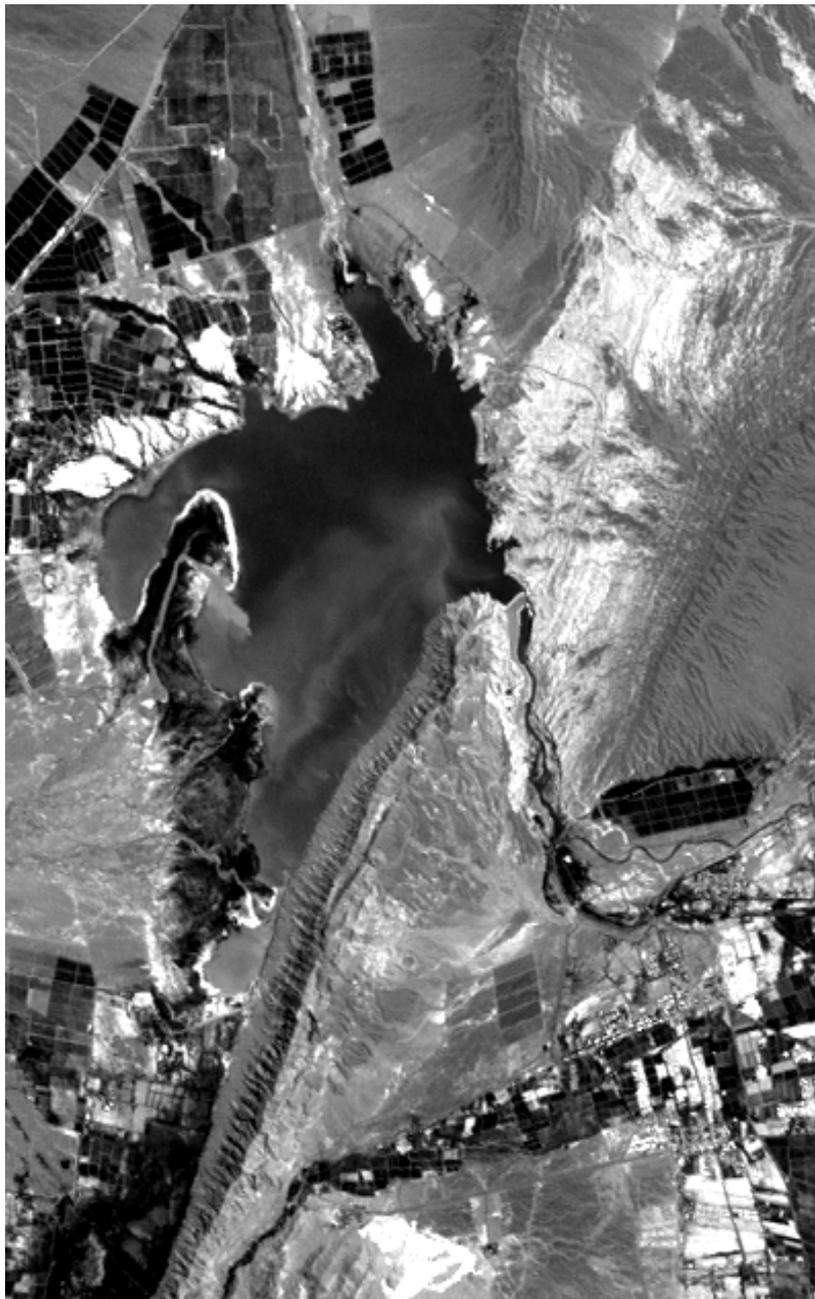


Figura 6.3 Imagen satelital elegida.

Se seleccionó una subimagen de 100 x 100 píxeles, con la que se trabajó. La misma se indica en la figura 6.4.

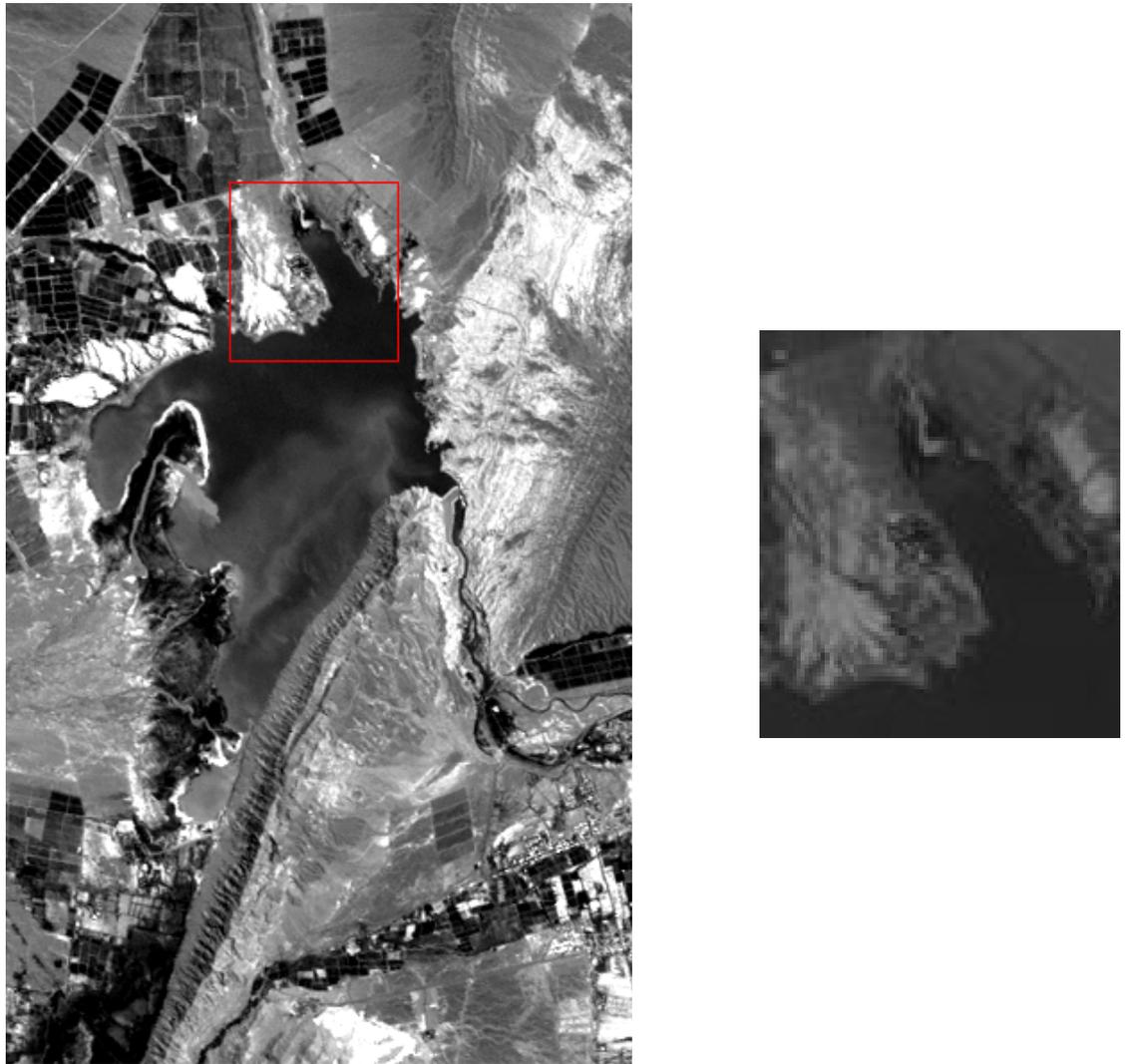


Figura 6.4 Imagen satelital y subimagen seleccionada para el estudio.

Para evaluar la capacidad de los modelos Autobinomiales en la extracción de información espacial, se eligieron en la subimagen, dos áreas de entrenamiento de igual tamaño: 9 x 9 píxeles (Ver figuras 6.5 y 6.6).

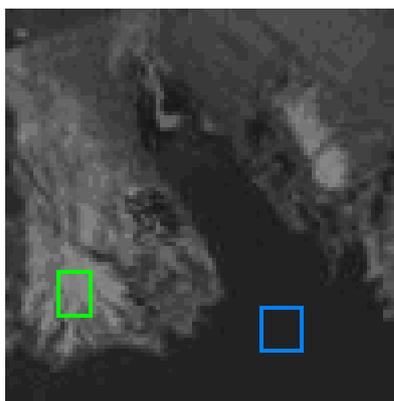


Figura 6.5 Marcadores de región: región I en verde y la región II en azul.



Figura 6.6 Regiones de referencia elegidas en la imagen. A la derecha, la región I y a la izquierda la región II.

La región I, corresponde a una porción de superficie montañosa y rocosa, desprovista de vegetación natural; mientras que la región II, muestra una porción de agua del embalse Ullum.

Como se indicó anteriormente, el objetivo es clasificar la subimagen satelital en diferentes regiones de textura mediante la estimación de los parámetros de los modelos Autobinomiales H-V e I-D a través de los estimadores PM y MCC, recorriendo la misma de izquierda a derecha por medio de ventanitas de 9 x 9 píxeles, tomando como referencia cada una de las dos regiones indicadas: I y II. Para establecer la similaridad entre las regiones, se utilizó la norma Euclídea,

$$\|\Delta\hat{\theta}\|^2 = \|\hat{\theta}^{(1)} - \hat{\theta}^{(0)}\|^2 = \sum_{ij} (b_{ij}^{(1)} - b_{ij}^{(0)})^2 \quad (6.6)$$

donde los superíndices (0) y (1), indican las estimaciones de los parámetros de la región testeada y la de referencia, respectivamente. La condición de similaridad considerada fue la establecida por Schöder et al (1997), $\|\Delta\hat{\theta}\|^2 \leq 2$.

6.4.1 Resultados

Primeramente, se determinaron las estimaciones de los parámetros de los modelos considerados para cada área de entrenamiento, obteniéndose los valores que exhibe la tabla 6.2.

Tabla 6.2 Estimaciones de los parámetros de las regiones de referencia.

Modelo	$\hat{\theta}$	Región I		Región II	
		$\hat{\theta}_{MP}$	$\hat{\theta}_{CMC}$	$\hat{\theta}_{MP}$	$\hat{\theta}_{CMC}$
H - V	b_{11}	-1.542	-1.554	-3.146	-3.137
	b_{12}	1.363	1.375	-3.868	-3.878
I - D	b_{21}	0.223	0.227	-3.012	-3.010
	b_{22}	-0.405	-0.410	-4.001	-4.003

Para cada una de las 121 ventanitas de la imagen satelital, se calcularon las estimaciones de los parámetros de los dos modelos con los dos métodos: MCC y MP. A partir de estos resultados, se evaluó el grado de acuerdo existente entre ellos, mediante el coeficiente de Lin, ρ_c , denominado coeficiente de correlación de concordancia (Concordance correlation coefficient).

Este coeficiente, desarrollado por Lin en 1989, evalúa si las mediciones sobre una misma variable cuantitativa, obtenidas a partir de dos métodos distintos, producen resultados equivalentes; valora el grado reproducibilidad de las mediciones. Su rango de valores oscila entre -1 y 1 y su valor absoluto no puede ser mayor que el coeficiente de correlación de Pearson r . El coeficiente de Lin sólo toma el valor cero si el coeficiente de correlación de Pearson es también cero.

La reproducibilidad perfecta es alcanzada cuando los puntos (el par de resultados de ambos métodos) caen sobre la línea de concordancia (línea de 45° que pasa por el origen).

El coeficiente de Lin se obtiene mediante la expresión:

Sean (x_{i1}, y_{i2}) $i = 1, \dots, n$ pares de mediciones independientes muestreadas de una población de distribución normal bivariada con medias μ_X y μ_Y y con matriz de varianza - covarianza

$$\begin{pmatrix} \sigma^2_X & \sigma^2_{XY} \\ \sigma^2_{XY} & \sigma^2_Y \end{pmatrix}$$

El coeficiente de correlación de concordancia poblacional de Lin es

$$\rho_C = \frac{2\sigma^2_{xy}}{\sigma^2_X + \sigma^2_Y + (\bar{x} - \bar{y})^2}$$

Este coeficiente es estimado reemplazando las medias, las varianzas y covarianzas poblacionales por sus respectivos estimadores puntuales:

$$\hat{\rho}_C = r_C = \frac{2S^2_{xy}}{S^2_X + S^2_Y + (\bar{x} - \bar{y})^2}$$

Donde:

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}, \quad \bar{y} = \frac{\sum_{i=1}^n y_i}{n}, \quad S^2_X = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}, \quad S^2_Y = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n-1}, \quad y$$

$$S^2_{XY} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n-1}$$

El error estándar de ρ_C , se obtiene mediante la siguiente fórmula:

$$S_{\hat{\rho}_C} = \sqrt{\frac{1}{n-2} \left[\left(\frac{1-r^2}{r^2} \rho_C^2 (1-\rho_C^2) \right) + \left(2\rho_C^3 (1-\rho_C) \frac{(\bar{x} - \bar{y})^2}{S_X S_Y r} \right) - \rho_C^4 \frac{(\bar{x} - \bar{y})^4}{2S_X^2 S_Y^2 r^2} \right]}$$

donde r es el estimador del coeficiente de correlación de Pearson.

El análisis de concordancia de las estimaciones efectuadas arrojó los siguientes resultados:

Tabla 6.3 Estimaciones de los coeficientes de Lin.

Modelo	Estimaciones del parámetro	Estimación de ρ_C	IC 95%
H - V	β_{11}	0.9969	(0.9956, 0.9978)
	β_{12}	0.9971	(0.9959, 0.9980)
I - D	β_{21}	0.9911	(0.9873, 0.9938)
	β_{22}	0.9980	(0.9871, 0.9986)

Puede observarse, en todos los casos, que el grado de concordancia entre los resultados emitidos por ambos métodos, fue “casi perfecto” (mayor de 0.99). Es decir, las estimaciones logradas por el método MCC son reproducidas en un alto grado por el MP.

Evaluación de los modelos H-V e I-D en la clasificación de la imagen satelital

Visto que en el capítulo anterior, se comprobó que el método de Máxima Pseudo-verosimilitud presenta una mejor performance en las estimaciones que el CMC, a efectos de comparar la capacidad en la clasificación supervisada de los modelos considerados: H-V e I-D, se tuvo en cuenta los resultados logrados a partir del mismo.

Como producto del proceso de clasificación supervisada, se obtuvieron los mapas que se observan en las figuras 6.7 y 6.8.

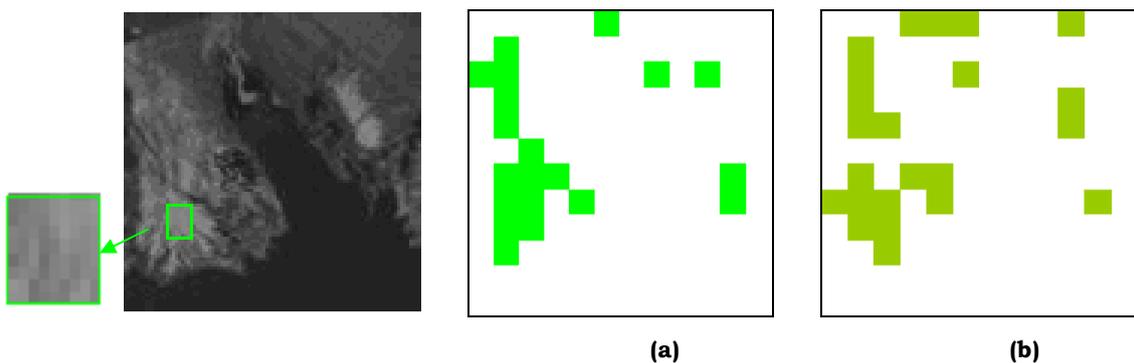


Figura 6.7 Grafico de la región I considerando: **(a)** el modelo Autobinomial H-V y **(b)** el modelo Autobinomial I-D.

Como puede observarse, la clasificación establecida por el modelo I-D resulta ser levemente más efectiva que la lograda a través del H-V. Esto puede deberse a las características que presenta el terreno en la imagen de referencia seleccionada como región I, se trata de una superficie montañosa, rocosa, desprovista de vegetación natural.

Respecto a los resultados obtenidos, teniendo como referencia la región II, el modelo I-D aparenta ser algo más discriminador que el H-V.

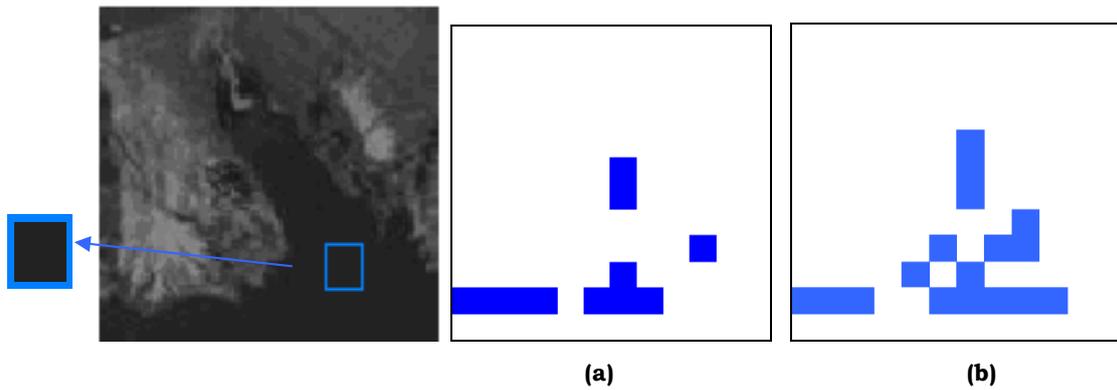


Figura 6.8 Grafico de la región II considerando: **(a)** el modelo Autobiomial H-V y **(b)** el modelo Autobiomial I-D.

Comparación de los Métodos de clasificación MCC y MP

A modo de comparar los resultados obtenidos de la clasificación de la imagen, a través de los métodos de estimación MCC y MP, se codificó cada una de las imágenes clasificadas según las regiones de referencia consideradas, mediante la siguiente función, $f(M)$:

$$f(M) = \begin{cases} 1 & \text{si } M \text{ es similar a la región } i \\ 0 & \text{en otro caso} \end{cases}, \quad i = 1, 2.$$

donde M representa cada una de las ventanitas de 9×9 píxeles de la imagen bajo estudio, y la región i es la región de referencia i , $i = 1, 2$.

Los datos alcanzados se resumieron en matrices y para cada región se calculó el valor absoluto de la diferencia entre los elementos de las matrices obtenidas por la clasificación de ambos métodos.

En la siguiente figura, se observa algunos de los resultados obtenidos al comparar la performance de los clasificadores ante una misma región de referencia:



Figura 6.9 Comparación de los clasificadores: **(a)** según región I, considerando el modelo Autobinomial H-V. **(b)** según región II, considerando el modelo Autobinomial I-D. En ambos casos, lo negro son coincidencias y los parches blancos representan las discordancias entre los dos clasificadores.

Las discrepancias, en el caso (a), sólo fueron del 1,7 % (2/121), mientras que en el caso (b) fueron del orden del 3,3% (4/121). Ambos clasificadores mostraron una fuerte concordancia al clasificar la imagen.

Con el propósito de continuar evaluando cuantitativamente el desempeño de los dos clasificadores estadísticos, MP y MCC, en la imagen satelital en estudio, se determinó el coeficiente Kappa (κ) y otros índices alternativos, en base a la matriz de confusión, para cada una de las áreas de entrenamiento consideradas. Para la elaboración de la “matriz de confusión” se tomó como referencia el clasificador MP, por presentar anteriormente, una mejor performance en las estimaciones que el MCC. La imagen quedó segmentada en dos clases, las texturas que eran similares al área de entrenamiento (indicadas con el valor 1) y las que diferían de ésta (indicadas con 0). Esta información fue comparada con el producto de la clasificación supervisada obtenida por el método MCC.

En este caso el elemento n_{ij} , de la matriz de confusión, representa el número de ventanitas de 9 x 9 de la imagen satelital que fueron clasificadas por MP pertenecientes a la clase i y por MCC a la clase j, n_{i+} : totales marginales de la fila i, n_{+i} : totales marginales de la columna i, n : número total de ventanitas de la imagen considerada para el estudio.

Las matrices de confusión obtenidas se muestran en las tablas 6.4.

Tablas 6.4 Matrices de confusión generadas.

(a) considerando a la región I como área de entrenamiento y el modelo Autobinomial H-V.

Región I		Clasificador MP		
Clasificador CMC	0	1	Total	
0	99	0	99	
1	2	20	22	
Total	101	20	121	

(c) considerando a la región II como área de entrenamiento y el modelo Autobinomial H-V.

Región II		Clasificador MP		
Clasificador CMC	0	1	Total	
0	117	1	118	
1	0	3	3	
Total	117	4	121	

(b) considerando a la región I como área de entrenamiento y el modelo Autobinomial I-D.

Región I		Clasificador MP		
Clasificador CMC	0	1	Total	
0	96	0	96	
1	4	21	25	
Total	100	21	121	

(d) considerando a la región II como área de entrenamiento y el modelo Autobinomial I-D.

Región II		Clasificador MP		
Clasificador CMC	0	1	Total	
0	109	1	110	
1	3	8	11	
Total	112	9	121	

Analizando las tablas 6.4 se deduce que la fiabilidad global es elevada en todas las matrices: 97.5%, 96.7%, 99.1%, 96.7%, respectivamente, mientras que los errores de omisión y comisión para las distintas clases son los que muestran la tabla 6.5.

Tablas 6.5 Porcentajes de errores de omisión y comisión.

Región I : H - V		% de Error	
Clasificador CMC		Omisión	Comisión
0		1.98	0
1		0	9.1

Región II : H - V		% de Error	
Clasificador		Omisión	Comisión
0		4	0
1		0	16

Región I: I - D		% de Error	
Clasificador CMC		Omisión	Comisión
0		0	0.85
1		25	0

Región II : I - D		% de Error	
Clasificador CMC		Omisió	Comisión
0		2.68	0.91
1		11	27

Los valores obtenidos de los estadísticos kappa se muestran en la tabla 6.6.

Tabla 6.6 Estimaciones del coeficiente Kappa para cada región y modelo Autobinomial.

Modelo	H - V		I - D	
Clasificador A	$\hat{\kappa}$	IC_{95%}	$\hat{\kappa}$	IC_{95%}
Región I	0.942	(0.863, 1.022)	0.893	(0.790, 0.996)
Región II	0.853	(0.566, 1.140)	0.782	(0.572, 0.992)

Estos valores revelan la existencia de un nivel de concordancia entre los clasificadores “muy bueno”, salvo el último de concordancia “sustancial” (según escala establecida por Landis y Koch, 1977). Todos ellos resultaron estadísticamente significativos ($P < 0.0001$), reafirmando así, la concordancia significativa entre las clasificaciones comparadas. Por lo tanto podría concluirse que serían clasificadores intercambiables.

CONCLUSIONES Y FUTURAS LÍNEAS DE INVESTIGACIÓN

En este último Capítulo se pretende exponer, de forma breve, las principales conclusiones obtenidas del trabajo desarrollado, y discutir las líneas de investigación que quedan abiertas.

7.1 Conclusiones

En el presente trabajo, se han definido dos modelos Gibbs-Autobinomiales: H-V (Horizontal -Vertical) y I-D (Izquierda-Derecha), que permiten generar texturas sintéticas y constituyen una herramienta fundamental para la extracción de información espacial de las imágenes de teledetección. Ellos modelan las dependencias estocásticas entre los píxeles definidos en ciertos sistemas de vecinos. Esta dependencia está caracterizada, en forma general, mediante una función de energía, y ponderada a través de su vector de parámetros θ . La captura de la información obtenida de una imagen se hace efectiva con la estimación del vector θ . El proceso de estimación se basó en el método de Máxima Pseudo-verosimilitud y el de Mínimos Cuadrados Condicional.

En la exposición de los contenidos de este trabajo, el acento se ha puesto en la comparación de estos dos métodos de estimación del vector θ : la evaluación de sus comportamientos ante distintos grados de contaminación a diferentes patrones homogéneos de textura y en la clasificación supervisada de las diferentes texturas existentes en una imagen satelital concreta. Algunas conclusiones relevantes obtenidas se enuncian a continuación:

- Al comparar los métodos de estimación en cuatro situaciones diferentes para el par de parámetros de cada uno de los dos modelos considerados, respecto a los

resultados numéricos obtenidos, no se vislumbraron diferencias significativas. Si bien, las magnitudes de los sesgos, en algunas de las situaciones consideradas, fueron relativamente altas para ambos métodos, las estimaciones obtenidas a partir de MCC presentaron mayor sesgo y dispersión que las logradas por MP. Esto indicaría que el estimador MP resulta ser levemente más eficiente que el método MCC.

- Analizando los resultados alcanzados en la comparación de los dos métodos de estimación frente a distintos grados de contaminación de diferentes patrones homogéneos de textura, puede advertirse la progresiva distorsión en las estimaciones a medida que aumenta el porcentaje de píxeles contaminados en la imagen. Revelando una alteración substancial en las magnitudes proporcionadas por ambos al contaminar al 20%. Si bien ambos manifestaron incrementos en los sesgos de las estimaciones a medida que aumentaba el grado de contaminación, las obtenidas por el método MP mantienen un sesgo y un desvío estándar menor que las alcanzadas con el MCC. El método MCC resultó ser más sensible que el MP frente a los distintos grados de contaminación considerados.
- Respecto a la capacidad de los modelos Autobinomiales H-V e I-D y la de los métodos de estimación MCC y el MP y, en la clasificación supervisada de las diferentes texturas presentes en una imagen satelital concreta:
 - El grado de concordancia (coeficiente de Lin) entre los resultados emitidos por ambos métodos al estimar los parámetros de los dos modelos para cada una de las 121 ventanitas consideradas en la imagen satelital, fue “casi perfecto”. Las estimaciones logradas por MCC fueron reproducidas en un alto grado por el MP.
 - La clasificación establecida por el modelo I-D resultó ser levemente más efectiva que la lograda a través del H-V, en las dos regiones (I y II) de la imagen consideradas como de referencia. En el primer caso, esto pudo deberse a las características que presentaba el terreno de la imagen de referencia seleccionada como región I, puesto que se trataba de una superficie montañosa, rocosa, desprovista de vegetación natural, es decir, presentaba texturas con orientaciones / o \. En el segundo caso, puede que se trate de una región con una mala calidad de resolución lo que produjo tal efecto.

- A modo de comparar los resultados obtenidos de la clasificación de la imagen, a través de los métodos de estimación MCC y MP y de evaluar sus desempeños, se codificó (1 - 0) cada una de las imágenes clasificadas según las regiones de referencia consideradas. Los datos alcanzados se resumieron en matrices y para cada región se calculó el valor absoluto de la diferencia entre los elementos de las matrices obtenidas por la clasificación de ambos métodos. Ambos métodos mostraron una fuerte correspondencia en los resultados obtenidos, las mayores discrepancias fueron del orden del 3,3% (4/121). Los valores de los coeficientes Kappa revelaron la existencia de un nivel de concordancia “muy bueno” (Landis y Koch, 1977) entre los clasificadores. Todos ellos resultaron estadísticamente significativos ($P < 0.0001$). Por lo tanto, podría concluirse que serían clasificadores intercambiables.
- Para evaluar la calidad con que habían quedado establecidas las categorías deducidas después del proceso de clasificación, a partir de MCC y MP, para cada una de las áreas de entrenamiento consideradas, se determinaron los índices basados en las “matrices de confusión”, tomando como referencia el clasificador MP,” por presentar anteriormente, una mejor performance en las estimaciones que el CMC. La fiabilidad global obtenida, resultó elevada en todas ellas.

En el análisis presentado, en general, los métodos MCC y MP emitieron resultados equivalentes. Si bien el estimador MP ofrece un desempeño levemente mejor que el MCC, es pertinente reconocer que éste último, requiere de algoritmos más sencillos para su determinación, a diferencia del MP que debe valerse de métodos de optimización global que dificultan su cálculo, ya sea por el tiempo que insumen en encontrar el valor máximo de la función de Pseudo-verosimilitud, como en no encontrar el punto máximo si éste difiere demasiado del punto inicial. Estas particularidades hacen del MCC, un estimador aconsejable, como para tener una estimación inicial del vector de parámetro de estudio.

7. 2 Aportes esperados de la tesis

El trabajo de tesis propuesto pretende presentar una reseña sobre procesos markovianos y distribuciones de Gibbs, ampliamente utilizados en la caracterización de imágenes, y se focaliza en el modelo Autobinomial, fundamentalmente en análisis de determinados aspectos de los estimadores clásicos de los parámetros de este modelo: Máxima Pseudo- verosimilitud y el estimador de Mínimos Cuadrados Condicional.

La contribución del mismo, se centra en el análisis inédito sobre la performance de los estimadores clásicos de los parámetros de modelos Autobinomiales ante desvíos de las suposiciones que los sustentan.

7. 3 Líneas Futuras de Trabajo

A partir de esta tesis han quedado abiertas diferentes líneas para la investigación futura, que se describen a continuación.

- Como objetivo inmediato, se pretende estudiar el Modelo Generalizado de textura, modelo construido sobre la base de un Campo Aleatorio de Markov con la introducción de un nuevo sistema de vecinos no estándar, denominado sistema de *vecindad funcional*, que permite modelar y sintetizar tanto micro como macrotexturas en una misma escala de resolución y resulta válido para innumerables campos de aplicación.
- También resulta fundamental el poder perfeccionar los algoritmos de estimación con el propósito de mejorar tanto la precisión como la velocidad de ejecución.
- Dada la diversidad de aplicaciones e intereses en la utilización de las imágenes satelitales y ante la existencia de demandas crecientes a cerca de la extracción de información espacial asociada al contenido de una imagen en las prácticas concretas de investigación y de asistencia de carácter profesional, especialmente en las áreas de Geografía, Geología, Agronomía, otra línea de investigación a bordar en un futuro, es la fusión de imágenes de satélite, como respuesta a la frecuente necesidad de resumir en una sola imagen, datos de alta resolución espectral y espacial.

En el ámbito de la teledetección, existe un amplio rango de fenómenos a estudiar que requieren disponer de imágenes de satélite que combinen una alta resolución espacial y espectral, como por ejemplo, la creación y actualización de mapas temáticos, la detección de cambios o la monitorización y gestión de recursos naturales. Las limitaciones tecnológicas de los sensores en el espacio, no permiten aprovechar toda la riqueza de información de las imágenes.

Una manera de disponer de imágenes de alta resolución tanto espacial como espectral, es el empleo de técnicas de *fusión de imágenes*. La *fusión imágenes* puede entenderse como la combinación sinérgica de información proporcionada por varios sensores o por el mismo sensor en diferentes escenarios (espaciales, espectrales y temporales) (Lillo-Saavedra et al., 2006). Permite ampliar la dimensionalidad del espacio observacional, incrementar la cobertura espacial y temporal, reducir las ambigüedades, suplir carencias de las observaciones consideradas individualmente. La fusión de imágenes Multiespectrales y Pancromáticas consiste en combinar de una forma coherente, la información espectral de la imagen Multiespectral, con la espacial de la imagen Pancromática, con el propósito que la imagen fusionada conserve las características más valiosas de cada imagen individual, proporcionando así una imagen más completa y rica en información, es decir, una imagen con una calidad global (espacial-espectral) de la información visual adquirida más alta.

Con el fin de profundizar en esta temática, y lograr cubrir la necesidad de aplicaciones concretas, se proponen las siguientes tareas de investigación:

- Realizar una revisión, clasificación y selección continua de bibliografía sobre la fusión de imágenes de satélite.
- Revisar el software disponible e implementar las modificaciones necesarias que permitan adecuarlo a distintas aplicaciones.
- Evaluar las técnicas existentes de fusión de Imágenes: transformada de Hermite multiescala y multidireccional, análisis de componentes principales, transformada wavelet, transformaciones piramidales, método de Brovey, filtros steerable, entre otros.
- Trabajar en el desarrollo de metodologías novedosas de fusión de imágenes.
- Evaluar métodos que permite valorar la calidad de las imágenes fusionadas.

BIBLIOGRAFÍA

- [1] **Acuna, C.** (1992). "*Texture modelling using Gibbs distributions*", GVGIP: Graphical models and image processing, Vol. 54, n ° 3pp. 210 – 222.
- [2] **Acuna, C.** (1994). "An Application of Bivariate Pseudolikelihood to Texture Synthesis", IEEE Trans Vol.4, No. 3, pp. 431 - 435.
- [3] **Albert, J. and Chib, S.** (1990). "*Bayesian Analysis of Binary and polychotomous Response Data* ", Technical Report, Bowling Green State University, Dept. of Mathematics and Statistics.
- [4] **Berkson, J.** (1955). "*Maximum likelihood and Minimum χ^2 of the Logistic Function*", Journal of the American Statistical Association, vol.50, pp.130-162.
- [5] **Besag, J. E.** (1972). "*Nearest - Neighbour Systems and the Auto-Logistic Model for Binary Data*", Journal of the Royal Statistical Society, Vol. B34, No. 1, pp. 75 – 83.
- [6] **Besag, J. E.** (1974). "*Spatial Interaction and the Statistical Analysis of Lattice Systems*", Journal of Royal Statistic Society, Vol.36, pp. 192 - 236.
- [7] **Besag, J. E.** (1975). "*Statistical Analysis of no-lattice Systems*", The Statistician, Vol.24, pp.192 – 236.
- [8] **Besag, J. E. and Moran, P. A.** (1975). "*On the Estimation of Spatial Interaction in Gaussian Lattice Processes*", Biometrika, Vol. 62, No. 3, pp. 555 – 562.
- [9] **Besag, J. E.** (1977). "*Efficiency of Pseudo-likelihood Estimation for Simple Gaussian Fields*", Biometrika, Vol. 64, No. 3, pp. 616 – 618.
- [10] **Besag, J. E. and Moran, P. A.** (1977). "Efficiency of Pseudo- likelihood estimation for simple Gaussian Fields", Biometrika, Vol.64, pp.616-618.
- [11] **Chellappa, R. and Chatterjee, S.** (1985). "*Classification of textures using Markov random fields*", IEEE transactions on Acoustics, Speech, and Signal Processing, vol.ASSP-33, n°4, pp.959-963.
- [12] **Chellapa R. and Jain, A.** (1993). "*Markov random fields: Theory and Application*", Academic Press, Boston San Diego.
- [13] **Chen, C. C. and Dubes, R. C.** (1989). "*Experiments in fitting discrete Markov random fields to texture*",IEEE Computer vision and Pattern Recognition, pp. 298 – 303.

- [14] **Cohen, F.S. and Cooper, D. B.** (1987). “*Simple parallel hierarchical relaxation algorithms for segmentation noncausal Markovian random models*”, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 5, pp. 195 – 219.
- [15] **Comets F.** (1992). “*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.
- [16] **Comets, F. and Gidas, B.** (1992). “*Parameter Estimation for Gibbs Distributions from Partially Observed Data*”, The Annals of Applied Probability, Vol. 2, No. 1, pp. 142 -170.
- [17] **Cross, G. R. and Jain, R. K.** (1983). “*Markov random field texture models*”, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 5, No. 1, pp. 25 – 39.
- [18] **Dass, S. C. and Jain, A.** (2001). “*Markov face models*”, Computer vision 2001.ICCV 2001.Proceedings. Eight IEEE International Conference on Volume 2, pp. 680-687, Vol. 2, Issue.
- [19] **Datcu, M. Seidel, K. and Walessa, M.** (1981). “*Spatial information retrieval from remote sensing images: Part A. Information theoretical perspective*”, IEEE Tr. on Geoscience and Remote Sensing.
- [20] **Datcu, M., Daschiel, H., Pelizzari, A., Quartulli, M., Galoppo, A., Colapicchioni, A., Pastori, M., Seidel, K., Marchetti, P.G. and D’Elia, S.** (2003). “*Information mining in remote sensing image archives –Part A: System Concepts*”, IEEE Trans on Geoscience and Remote Sensing, pp. 2923 – 2936.
- [21] **Dellaportas, P. and Smith, A.F.M.** (1990). “*Bayesian inference for Generalized Linear Models*”, Technical Report, University of Nottingham, Dept. of Mathematics.
- [22] **Derin, H. and Eliot, H.** (1984). “*Modelling and segmentation of noisy and texture images using Gibbs Random Fields*”, IEEE Trans Pattern Anal: machine Intell. Vol. 9, pp.39 – 55.
- [23] **Diebolt, J. and Robert, C.** (1990). “*Estimation of Finite Mixture Distributions Through Bayesian Sampling*”, Technical Report, Université Paris VI,vL. S.T.A.
- [24] **Dubes, R. C. and Jain, R. K.** (1989). “*Random Field models in image analysis*”. Journal of Applied Statistics. 16 (2), pp. 131 – 164.
- [25] **Eddy, W. F. and Schervish, M. J.** (1990). “*The Asymptotic Distribution of Runninig Time Quicksort*”, Technical Report, Carnegie Mellon University, Dept. of Statistics.
- [26] **Fernández, M. M.** (2001). “*Contribuciones al análisis automático y semiautomático de ecografía fetal tridimensional mediante Campos Aleatorios de Markov y contornos activos. Ayudas al diagnóstico precoz de malformaciones*”, Tesis Doctoral. Departamento de teoría de la señal y comunicaciones e ingeniería telemática, Escuela Técnica Superior de Ingenieros de Telecomunicación. Universidad de Valladolid.

- [27] **Frigessi A., Stefano, P. Hwang, C. R. and Sheu, S. J.** (1993). “*Convergence Rates of the Gibbs Sampler, the Metropolis Algorithm, and Other Single-Site Updating Dynamics*”, Journal of the Royal Statistical Society, Vol. B55, No. 1, pp. 205 – 219.
- [28] **Gelfand, A. E., Hill A. F. M. and Lee, T. M.** (1992). “*Bayesian Analysis of Constrained Parameter and truncated Data Problems Using Gibbs Sampling*”, Journal of the American Statistical Association, vol.87, pp.523-532.
- [29] **Gelfand, S. B. and Mitter, S. K.** (1993). “*Metropolis - Type Annealing Algorithms for Global Optimization in R, SIAM Journal of Control and Optimization*”, Vol. 31, No. 1, pp. 111 – 131.
- [30] **Gelfand, A. E. and Smith, A.F.M.** (1990). “*Sampling –Based Approaches to calculating Marginal Densities*”, Journal of the American Statistical Association, vol.85, pp.398-409.
- [31] **Geman, D.** (1991). “*Random Fields and inverse problems in imaging*”, in Lecture Notes in Mathematics, Vol.1427, pp.113-193, Springer-Verlag.
- [32] **Geman, S. and Graffigne, C.** (1986). “*Markov random Field image models and their applications to computer vision,*” Proceedings of the International Congress of Mathematicians, pp. 1496 – 1517.
- [33] **Geman S. and Geman, D.** (1984). “*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.
- [34] **Geman, S. and Huang, C.** (1986). “*Diffusions for Global Optimization, SIAM Journal of Control and Optimization*”, Vol. 24, No. 5, pp. 1031 – 1043.
- [35] **Georgii, H. O.** (1988). “*Gibbs Measure and Phase transition*”, de Gruyter, Berlin; New York.
- [36] **Gilks, W. R. and Wild, P.** (1992). “*Adaptive Rejection Sampling for Gibbs Sampling*”, Journal of the Royal Statistical Society, Ser. C, vol.41, pp.337-348.
- [37] **Gimel’fard, G. L.** (1999). “*Image Textures and Gibbs Random Fields*”. University of Auckland, Auckland, New Zealand. Kluwer Academic Publishers: Dordrecht e. a.
- [38] **Gimel’ farb, G. and Farag, A. A.** (2005). “*Texture analysis by accurate identification of simple Markovian Models*”, Journal Cybernetics and Systems analysis. Springer New York, Vol. 41, n°1, pp. 1- 156.
- [39] **Guyon, X. and Künsch, H. R.** (1996). “*Asymptotic comparison of estimators in the Ising model*”, in Stochastic Models, Statistical Methods and Algorithms in Image Analysis, P. Barone, A.Frigessi and M. Piccioni Eds., Berlin, Vol. 74 of Lecture Notes in Statistics, pp.177-198, Springer.
- [40] **Grimmett, G. R.**(1973). “A Theorem about random fields. Bull. London Math. Soc. 5, pp. 81-84
- [41] **Hammersley, J. M. and Handscomb, D. C.** (1964). “*Monte Carlo Methods*”, Methuen, London.

- [42] **Hammersley, J. M. and Clifford, P.** (1971). "Markov fields on finite graphs and lattices". In Unpublished.
- [43] **Hassner, M. and Sklanky, J.** (1980). "The use of Markov random field as models of texture", *Comput. Graphics Image Processing*, Vol. 12, pp. 357 – 370.
- [44] **Hastings, W. K.** (1970). "Monte Carlo Sampling Methods Using Markov Chains and Their Applications", *Biometrika*, Vol.57, pp.97-109.
- [45] **Hebar, M., Gleich, D. and Cucej, Z.** (2009). "Autobinomial Model for SAR Image Despeckling and Information Extraction". *Geoscience and Remote Sensing*, IEEE Transactions on. Vol. 47, No. 8, August, pp. 2818-2835.
- [46] **Ising, E.** (1925). Beitrag zur Theorie der ferromagnetismus. *Z. Phys.*, 31, 253 – 258.
- [47] **Howard, D.** (1995). "Markov random fields and image processing". PhD thesis, Flinders University, Adelaide.
- [48] **Jackson, M. C.** (2003) "A comparison of estimation methods for spatial data analysis with discrete data on lattice", Department of Mathematics, University of Maryland.
- [49] **Jain, A. K.** (1989) "Fundamentals of Digital Image Processing", Prentice Hall.
- [50] **Johansson, J. O.** (2001). "Parameter - estimation in the auto - binomial model using coding and pseudo-likelihood method approached with simulated annealing and numerical optimization", *Pattern Recognition Letters*, pp. 1233 - 1246.
- [51] **Jensen, J. L. Moller, J.** (1991). "Pseudolikelihood for Exponential Family Models of Spatial Point Processes", *The Annals of Applied Probability*, Vol. 1, No. 3, pp. 445 – 461.
- [52] **Kashyap, R., Chellappa, R. and Khotanzad, A.** (1982). "Texture classification using features derived from random field models", *Pattern Recognition Lett.* Vol.1, pp. 43 – 50.
- [53] **Kato, Z., Zerubia, J. and Berthod, M.** (1995). "Unsupervised parallel image classification using a hierarchical Markovian model", *IEEE 5th international Conference on Computer (B. Werner, ed.)*, (Los Alamitos), pp.169-174, IEEE Computer Society Press.
- [54] **Khotanzad, A. and Kashyap, R. L.** (1987). "Feature selection for texture recognition based on image synthesis", *IEEE Transactions on Systems, Man and Cybernetics*, Vol. 17, pp. 1087-1095.
- [55] **Kindermann, R. and Snell, J. L.** (1980). "Markov Random Fields and their applications". American Mathematical Society.
- [56] **Kirkpatrick, S., Gelatt, C. D. and Vecchi, M. P.** (1983). "Optimization by Simulated Annealing", *Science*, Vol. 220, pp. 671-680.

- [57] **Lakshmanan, S. and Derin, H.** (1989). “*Simultaneous Parameter Estimation and Segmentation of Gibbs Random Fields Using Simulated Annealing*”, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 11, No. 8, pp. 799-813.
- [58] **Lange, N., Carlin, B. P. and Gelfand, A. E.** (1990). “*Hierarchical Bayes Models for the progression of HIV Infection Using longitudinal CD4+Counts*”, Technical Report, Carnegie Mellon University, Dept. of Statistics.
- [59] **Lele, S. R. and Ord, J. K.** (1986). “*Conditional least squares estimation for spatial processes: some asymptotics results*”, Tech. Rep. 65, Dept. Stat., The Pennsylvania State Univ.
- [60] **Li, S. Z.** (1995). “*Markov Random Fields Modelling Computer Vision*”, Springer, First Edition.
- [61] **Li, S. Z.** (2001). “*Markov Random Field Modelling in Image Analysis*”, Springer.
- [62] **Lippok, A. and Reulke, R.** (2006). “*Discrimination between urban area and vegetation in highresolution images using markov random fields (MRF)*”. 1st EARSeL Workshop of the SIG Urban Remote Sensing Humboldt-Universität zu Berlin.
- [63] **Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. and Teller, E.** (1953). “*Equations of State Calculations by Fast Computing Machines*”, Journal of Chemical Physics, Vol. 21, pp. 1087-1091.
- [64] **Newell, G. F. and Montroll, E. W.** (1953). “*On the Theory of the Ising Model of Ferromagnetism*”, The Review of Modern Physics, Vol. 25, No. 2, pp. 353 – 389.
- [65] **Onsager, L.** (1944). “*Crystal Statistics. I. A Two-Dimensional Model with an Order Disorder Transition*”, Phys. Rev. 65, 117-149.
- [66] **Otsu, N.** (1979) “*A Threshold Selection Method from Gray-Level Histograms*”, IEEE Transactions on System, Man, and Cybernetics, Vol. 9, No. 1, pp. 62 - 66.
- [67] **Ozyildiz, E. Krahnstöver, N. and Sharma, R.** (2001). “*Adaptive texture and colour segmentation for tracking moving objects*”, Department of computer science and engineering, Pennsylvania State University. Preprint submitted to Elsevier preprint.
- [68] **Paget, R. and Longstaff, D.** (1997). “*Extracting the Cliques from a Neighbourhood System*”, IEE Proceedings - Vision, Image, and Signal Processing, Vol. 144, No. 3, pp. 168 - 170.
- [69] **Paget, R. and Longstaff, D.** (1995). “*Texture synthesis via a nonparametric Markov random field,*” in Proceedings of DICTA-95, Digital Image Computing: Techniques and Applications (A. Maeder and B. Lovell, eds.), Vol. 1, (Brisbane, Australia), pp. 547- 552, Australian Pattern Recognition Society. <http://www.vision.ee.ethz.ch/~rpaget>.
- [70] **Pajares Martinsanz, G. P y De la Cruz García, J. M.** (2008). *Visión por computador. Imágenes Digitales y Aplicaciones*. Alfaomega Grupo Editor, S. A. de C.V., México.

- [71] **Peierls, R. E.** (1936). "On Ising's Model of Ferromagnetism", Proc. Camb. Phil. Soc. 32, 477- 481.
- [72] **Pujol, M., Rizo, R. Arques, P. , Compañ, P., Escolano, F. and Molina, R.** (2000). "Aplicación de los modelos de campos aleatorios de Markov en visión artificial", Departamento de la Computación e Inteligencia Artificial. Departamento de Tecnología Informática y Computación. Universidad de Alicante. Revista Electrónica de Visión por Computador. N°4.
- [73] **Richards, J. A.** (1993). "Remote Sensing Digital Image Analysis: An Introduction". Berlin: Springer-Verlag.
- [74] **Rignot, E. and Kwok, R.** (1990). "Extraction of textural features in SAR images: Statistical model and sensitivity," in Proceedings International Geoscience and Remote Sensing Symposium, (Washington, DC), pp. 1979 - 1982.
- [75] **Rota, G. C.** (1964) "On the foundations of combinatorial theory", Zeitschrift Fur Wahrscheinlichkeitstheorie Und Verwandte Gebietee, vol. 2, pp. 340-368.
- [76] **Seong-Wook Joo,** (2003). "An Overview of Markov Random Field and Application to Texture Segmentation".
- [77] **Seymour, L.** (1993). "Parameter estimation and model selection in image analysis using Gibbs-Markov random fields". PhD thesis, The University of North Carolina, Chapel Hill.
- [78] **Schöder, M., Rehrauer, H., Seidel, K. and Dactu, M.** (1998). "Spatial Information from remote sensing images: Part B. Gibbs Markov Random Fields", IEEE Transactions on information Theory, 35(6), pp.1446 -1455.
- [79] **Schöder, M., Seidel, K., and Dactu, M.** (1997). "Gibbs Random Field Models for image Content Characterization", In a Scientific Vision for Sustainable Development, IGARSS' 97.T. I. Stein. Vol.1, pp.258 - 260.
- [80] **Schöder, M., Walessa, M.,Rehrauer, H., Seidel, K. and Dactu, M.** (2000). "Gibbs Random Field Models: a toolbox for Spatial Information Extraction", Computer & Geoscience. Vol.26, No.4, pp.423 - 432.
- [81] **Schultz, T. D., Mattis, D.C. and Lieb, E. H.** (1964). Rev.Mod. Phys. 36,856.
- [82] **Spitzer, F.** (1971). "Markov random fields and Gibbs ensembles", American Mathematical Monthly, Vol. 78, pp. 142 - 154.
- [83] **Tuceryan, M. and Jain, A. K.** (1993). "Texture analysis," in *Handbook of Pattern Recognition and Computer Vision* (C. H. Chen, L. F. Pau, and P. S. P. Wang, eds.), pp. 235 - 276, Singapore: World Scientific.
- [84] **Verdinelli, I. and Wasserman, L.** (1990). "Bayesian Analysis of Outliers problems using Gibbs Sampler", Technical Report 469, Carnegie Mellon University, Dept. of Statistics.
- [85] **Wechsler, H.** (1980). "Texture analysis - a survey," Signal Processing, Vol. 2, pp. 271 282.

- [86] **Winkler, G.** (1995). *“Image analysis, Random Fields and Dinamic Monte Carlo Methods”*, Springer, New York.
- [87] **Yang, C. N.** (1952). Phys.Rev.85, 809.
- [88] **Younes, L.** (1988). *“Estimation and Annealing for Gibbsian Fields”*, Annales de l’ Institut Henri Poincare - Probabilites et Statistiques, Vol. 24, No. 2, pp. 269 – 294.
- [89] **Zeger, S. and Riuzaul Karim, M.** (1991). *“Generalized Linear Models with Random Effects: A Gibbs sampling Approach”*, Journal of the American Statistical Association, vol.86, pp.79-86.
- [90] **Zhang, J.** (1992). *“The Mean Field Theory in EM Procedures for Markov Random Fields”*, IEEE Transactions on Signal Processing, Vol. 40, No. 10, pp. 2570 - 2583.

APENDICES

NOCIONES MATEMÁTICAS DE LAS CADENAS DE MARKOV Y DE METODOS DE OPTIMIZACIÓN

En este apéndice se presenta la teoría esencial requerida para el desarrollo de los métodos Monte Carlo basados en Cadenas de Markov. Se incluyen teoremas vitales para probar la convergencia de los métodos Monte Carlo. Los resultados han sido extraídos de [21]. No se han incluido las demostraciones que pueden ser consultadas en dicha referencia. Finalmente, se presenta el Método de Optimización de Newton.

A.1 Cadenas de Markov

A.1.1 Conceptos básicos

Definición A.1.1 Se denomina *probabilidad de transición* o *kernel de Markov* a la familia de distribuciones de probabilidad $(\mathbf{P}(\mathbf{x}, \cdot))_{\mathbf{x} \in \Omega}$.

Una probabilidad de transición \mathbf{P} puede representarse también, mediante una matriz, denotada por \mathbf{P} , donde $\mathbf{P}(\mathbf{x}, \mathbf{y})$ es el elemento de la fila \mathbf{x} y de la columna \mathbf{y} de una matriz cuadrada de probabilidades de orden $|\Omega|$ con vectores de probabilidad por filas.

Si \mathbf{v} es una distribución de probabilidad en Ω entonces $\mathbf{v}(\mathbf{x})\mathbf{P}(\mathbf{x}, \mathbf{y})$ es la probabilidad de extraer de forma aleatoria \mathbf{x} de \mathbf{v} y luego extraer de forma aleatoria \mathbf{y} de $\mathbf{P}(\mathbf{x}, \cdot)$. La distribución de probabilidad \mathbf{v} puede representarse mediante una matriz de orden $1 \times |\Omega|$, y denotarse \mathbf{v} . La probabilidad de arribar a la configuración \mathbf{y} partiendo de cualquier configuración \mathbf{x} es:

$$\mathbf{v}\mathbf{P}(\mathbf{y}) = \sum_{\mathbf{x}} \mathbf{v}(\mathbf{x})\mathbf{P}(\mathbf{x}, \mathbf{y}).$$

$\mathbf{v}\mathbf{P}(\mathbf{y})$ es una nueva distribución de probabilidad sobre Ω ($\sum_{\mathbf{y} \in \Omega} \mathbf{v}\mathbf{P}(\mathbf{y}) = 1$). Si se parte de una configuración \mathbf{x} , y se aplica un kernel de Markov \mathbf{P} y luego otro \mathbf{Q} , se obtendrá la configuración \mathbf{y} con probabilidad

$$\mathbf{P}\mathbf{Q}(\mathbf{y}) = \sum_{\mathbf{z}} \mathbf{P}(\mathbf{x}, \mathbf{z})\mathbf{Q}(\mathbf{z}, \mathbf{y}).$$

La composición de probabilidades de transición \mathbf{P} y \mathbf{Q} , $\mathbf{P}\mathbf{Q}$, es un también un kernel de Markov. Tanto $\mathbf{v}\mathbf{P}$ como $\mathbf{P}\mathbf{Q}$ representan un producto de matrices. Dada una distribución \mathbf{v} y kernels \mathbf{P}_i , $i = 1, 2, \dots, n$ puede definirse recursivamente $\mathbf{v}\mathbf{P}_1\mathbf{P}_2\dots\mathbf{P}_n = (\mathbf{v}\mathbf{P}_1\mathbf{P}_2\dots\mathbf{P}_{n-1})\mathbf{P}_n$. Todas las propiedades de producto de matrices pueden aplicarse a la composición de kernels.

Definición A.1.2 Una cadena de Markov en el espacio finito Ω queda determinada por una distribución inicial \mathbf{v} y kernels de Markov $\mathbf{P}_1, \mathbf{P}_2, \dots$ en Ω . La cadena se dice que es *homogénea* si $\mathbf{P}_i = \mathbf{P} \forall i$; en caso contrario, es *no homogénea*.

Dada una cadena de Markov, la probabilidad de que en los instantes de tiempo 0, 1, . . . , n los estados sean x_0, x_1, \dots, x_n es

$$\mathbf{v}(x_0)\mathbf{P}_1(x_0, x_1)\mathbf{P}_2(x_1, x_2)\dots\mathbf{P}_n(x_{n-1}, x_n).$$

Esto define una distribución de probabilidad $\mathbf{P}^{(n)}$ en el espacio $\Omega^{\{0,1,\dots,n\}}$ de tales secuencias de longitud $n + 1$. Estas distribuciones son consistentes, $\mathbf{P}^{(n+1)}$ induce $\mathbf{P}^{(n)}$ por

$$\mathbf{P}^{(n)}((x_0, \dots, x_n)) = \sum_{x_{n+1}} \mathbf{P}^{(n+1)}((x_0, x_n, x_{n+1})).$$

Definición A.1.3 Se denomina *camino* de una cadena de Markov a una secuencia infinita de estados (x_0, x_1, \dots) . El conjunto de todos los *caminos* se lo denotará por $\Omega^{\mathbb{N}_0}$, es un conjunto numerable con infinitos elementos. Debido a la consistencia, pueden definirse probabilidades sobre los elementos de $\Omega^{\mathbb{N}_0}$ siempre que dependan de un número finito de índices temporales. Sea $A \subset \Omega^{\mathbb{N}_0}$ tal que $A = B \times \Omega^{\{n+1, \dots\}}$, con $B \subset \Omega^{\{0, \dots, n\}}$, luego se define la probabilidad de A como

$$\mathbf{P}(A) = \mathbf{P}^{(n)}(B)$$

Esta probabilidad se denomina *probabilidad de A* (respecto de la cadena dada).

Las cadenas de Markov pueden introducirse también a través de secuencias de variables aleatorias ξ_i que verifican la propiedad de Markov

$$\mathbf{P}(\xi_n = x_n / \xi_0 = x_0, \dots, \xi_{n-1} = x_{n-1}) = \mathbf{P}(\xi_n = x_n / \xi_{n-1} = x_{n-1}),$$

$\forall n \geq 1$ y $x_0, \dots, x_{n-1} \in \Omega$.

Para obtener cadenas de Markov como en el sentido anterior, la distribución inicial \mathbf{v} esta dada por $\mathbf{v}(x) = \mathbf{P}(\xi_0 = x)$ y la probabilidad de transición \mathbf{P} por

$$\mathbf{P}_n(x, y) = \mathbf{P}(\xi_n = y / \xi_{n-1} = x).$$

Recíprocamente, dada una cadena de Markov $(x_i)_{i \geq 0}$ para una distribución inicial y unos kernels dados, las variables aleatorias ξ_i pueden definirse como la proyección del espacio $\Omega^{\{0, \dots\}}$ en sus coordenadas Ω , es decir, mediante las aplicaciones

$$\xi_n : \Omega^{\{0, \dots\}} \rightarrow \Omega$$

$$(x_i)_{i \geq 0} \rightarrow x_n$$

La distribución de probabilidad en el instante n , \mathbf{v}_n , definida

$$\begin{aligned}
\mathbf{v}_n(\mathbf{x}) &= \mathbf{P}(\boldsymbol{\xi}_n = \mathbf{x}) = \sum_{x_0, \dots, x_{n-1}} \mathbf{P}((x_0, \dots, x_{n-1}, \mathbf{x})) \\
&= \sum_{x_0, \dots, x_{n-1}} \mathbf{v}(x_0) \mathbf{P}_1(x_0, x_1) \mathbf{P}_2(x_1, x_2) \dots \mathbf{P}_n(x_{n-1}, \mathbf{x}) \\
&= \mathbf{v} \mathbf{P}_1 \mathbf{P}_2 \dots \mathbf{P}_n(\mathbf{x}),
\end{aligned}$$

se denomina *distribución de probabilidad marginal unidimensional* de la cadena de Markov en el instante n.

Si $m < n$ la probabilidad

$$\mathbf{v}_{mn}(x, y) = \mathbf{P}(\boldsymbol{\xi}_m = x, \boldsymbol{\xi}_n = y) = \mathbf{v} \mathbf{P}_1 \mathbf{P}_2 \dots \mathbf{P}_m(x) \mathbf{P}_{m+2} \dots \mathbf{P}_n(x, y),$$

se denomina *distribución de probabilidad marginal bidimensional* de la cadena de Markov en los instantes m y n. Las distribuciones de probabilidad marginal para más dimensiones $\mathbf{P}(\boldsymbol{\xi}_{n_1} = x_1, \boldsymbol{\xi}_{n_2} = x_2, \dots, \boldsymbol{\xi}_{n_k} = x_k)$ se definen de igual manera.

Se suele denominar a \mathbf{P} *ley* de la cadena de Markov $(\xi_n)_{n \geq 0}$.

Sean f las funciones definidas sobre $\Omega^{\mathbf{N}_0}$ que dependen de un número finito de índices temporales, es decir, $\exists k \geq 0$ tal que para toda secuencia $(x_n)_{n \geq 0}$, $f((x_n)_{n \geq 0}) = f(x_0, x_1, \dots, x_k)$, luego la esperanza matemática, $\mathbf{E}(f)$, con respecto a la ley \mathbf{P} se define como

$$\mathbf{E}(f) = \sum_{x_0, \dots, x_k} f(x_0, x_1, \dots, x_k) \mathbf{P}((x_0, x_1, \dots, x_k)).$$

Debido a que se está interesado en el comportamiento en el límite de las cadenas de Markov, a continuación se presentan dos diferentes tipos de convergencia que se utilizarán:

Definición A.1.4 Sean ξ y $\xi_0, \xi_1, \xi_2, \dots$ variables aleatorias, se dice que $(\xi_n)_{n \geq 0}$ converge a ξ :

1. De forma *cuadrática* o en L^2 si $\lim_{n \rightarrow \infty} \mathbf{E}(\xi_n - \xi)^2 = 0$.

2. En *probabilidad* si $\forall \varepsilon > 0$ se verifica que $\lim_{n \rightarrow \infty} P(|\xi_n - \xi| > \varepsilon) = 0$.

Proposición A.1.1 Sean ξ y ξ_0, ξ_1, \dots variables aleatorias, se verifica que:

1. Si $(\xi_n)_{n \geq 0}$ converge en L^2 a ξ , entonces $(\xi_n)_{n \geq 0}$ converge en *probabilidad* a ξ .
2. Si $(\xi_n)_{n \geq 0}$ converge en *probabilidad* a ξ y $(\xi_n)_{n \geq 0}$ está acotada, entonces lo hace en L^2 . En este caso ambas convergencias son equivalentes.

Finalmente, puede probarse que una cadena de Markov con distribución de probabilidad inicial y kernels de Markov estrictamente positivos induce un CAM (finito) de forma natural. En cada intervalo temporal $\mathbf{I} = \{0, \dots, n\}$ se define el sistema de vecinos como $\partial(k) = \{k-1, k+1\} \cap \mathbf{I}$. Para $k \in \mathbf{I}/\{0\}$

$$P(\xi_k / \xi_i, 0 \leq i \leq n, i \neq k) = \frac{P(\xi_i = x_i, 0 \leq i \leq n)}{P(\xi_i, 0 \leq i \leq n, i \neq k)} = P(\xi_k = x_k / \xi_{k-1} = x_{k-1}, \xi_{k+1} = x_{k+1})$$

Y además

$$P(\xi_0 = x_0 / \xi_i = x_i, 0 \leq i \leq n) = P(\xi_0 = x_0 / \xi_1 = x_1)$$

$$P(\xi_n = x_n / \xi_i = x_i, 0 \leq i \leq n-1) = P(\xi_n = x_n / \xi_{n-1} = x_{n-1}).$$

Esta es la propiedad de Markov espacial definida en la sección 2.2 para los campos aleatorios.

A.1.2 Coeficiente de Contracción

El método de *contracción* se utiliza para demostrar los teoremas de límite de cadenas de Markov homogéneas y no homogéneas. Las demostraciones presentadas son válidas para espacios de configuraciones Ω finitos.

Definición A.1.4 Para las distribuciones μ y ν en Ω , la norma de la variación total de la diferencia $\mu - \nu$ está dada por

$$\|\mu - \nu\| = \sum_x |\mu(x) - \nu(x)|.$$

Es la norma L^1 de la diferencia.

Lema A.1.1 Sean μ y ν distribuciones de probabilidad en Ω , entonces ¹¹ ¹²

¹¹ Si a y b son dos números reales, el mínimo de a y b se denota $a \wedge b$.

$$\begin{aligned}
\|\boldsymbol{\mu} - \boldsymbol{\nu}\| &= 2 \sum_{\mathbf{x} \in \Omega} (\boldsymbol{\mu}(\mathbf{x}) - \boldsymbol{\nu}(\mathbf{x}))^+ \\
&= 2 \left(1 - \sum_{\mathbf{x} \in \Omega} (\boldsymbol{\mu}(\mathbf{x}) \wedge \boldsymbol{\nu}(\mathbf{x})) \right) \\
&= \max \left\{ \left| \sum_{\mathbf{x} \in \Omega} h(\mathbf{x}) \boldsymbol{\mu}(\mathbf{x}) - \boldsymbol{\nu}(\mathbf{x}) \right| : |h| \leq 1 \right\}.
\end{aligned}$$

Mediante la desigualdad triangular se tiene que $\|\boldsymbol{\mu} - \boldsymbol{\nu}\| \leq 2$. La igualdad de esta última identidad se alcanza si las distribuciones de probabilidad $\boldsymbol{\mu}$ y $\boldsymbol{\nu}$ son ortogonales¹³, y $\|\boldsymbol{\mu} - \boldsymbol{\nu}\| = 0 \leftrightarrow \boldsymbol{\mu} = \boldsymbol{\nu}$.

Definición A.1.5 El *coeficiente de contracción* de un kernel de Markov \mathbf{P} se define

$$c(\mathbf{P}) = \frac{1}{2} \max_{\mathbf{x}, \mathbf{y}} \|\mathbf{P}(\mathbf{x}, \cdot) - \mathbf{P}(\mathbf{y}, \cdot)\|.$$

$c(\mathbf{P}) \leq 1$ y la igualdad se verifica si y sólo si al menos dos de las distribuciones $\mathbf{P}(\mathbf{x}, \cdot)$ son ortogonales. $c(\mathbf{P}) = 0$ si y sólo si todas las distribuciones $\mathbf{P}(\mathbf{x}, \cdot)$ son iguales. Por consiguiente, $c(\mathbf{P})$ es una medida de la ortogonalidad de las distribuciones $\mathbf{P}(\mathbf{x}, \cdot)$.

Lema A.1.2 Sean $\boldsymbol{\mu}$ y $\boldsymbol{\nu}$ dos distribuciones de probabilidad, y \mathbf{P} y \mathbf{Q} dos kernels de Markov definidos en Ω , entonces

$$\|\boldsymbol{\mu}\mathbf{P} - \boldsymbol{\nu}\mathbf{P}\| \leq c(\mathbf{P}) \|\boldsymbol{\mu} - \boldsymbol{\nu}\|,$$

$$c(\mathbf{P}\mathbf{Q}) \leq c(\mathbf{P})c(\mathbf{Q}).$$

En particular

$$\|\boldsymbol{\mu}\mathbf{P} - \boldsymbol{\nu}\mathbf{P}\| \leq \|\boldsymbol{\mu} - \boldsymbol{\nu}\|,$$

$$\|\boldsymbol{\mu}\mathbf{P} - \boldsymbol{\nu}\mathbf{P}\| \leq 2c(\mathbf{P}).$$

¹² Para un vector $\boldsymbol{\varphi} = (\varphi(\mathbf{x}))_{\mathbf{x} \in \Omega}$ se define la parte positiva de $\boldsymbol{\varphi}$ $\varphi^+ = \begin{cases} \varphi(\mathbf{x}) & \text{si } \varphi(\mathbf{x}) > 0 \\ 0 & \text{en otro caso} \end{cases}$, y la parte negativa $\varphi^- = (-\varphi)^+$.

¹³ Se denomina *soporte de una distribución de probabilidad* $\boldsymbol{\nu}$ al conjunto $\{\mathbf{x} \in \Omega / \boldsymbol{\nu}(\mathbf{x}) > 0\}$. Dos distribuciones de probabilidad $\boldsymbol{\mu}$ y $\boldsymbol{\nu}$ se dicen *ortogonales* si poseen soporte disjunto.

Una consecuencia inmediata del Lema A.1.2 es la pérdida asintótica de memoria o también llamada *ergodicidad débil* de las cadenas de Markov:

Sean $(\mathbf{P}_n)_{n \geq 1}$ kernels de Markov, μ y ν dos distribuciones iniciales de probabilidad, entonces $c(\mathbf{P}_1 \dots \mathbf{P}_n) \rightarrow 0$ implica que $\|\mu \mathbf{P}_1 \dots \mathbf{P}_n - \nu \mathbf{P}_1 \dots \mathbf{P}_n\| \rightarrow 0$.

La convergencia de las cadenas de Markov es rápida si el coeficiente de contracción es pequeño.

Lema A.1.3 Para todo kernel de Markov \mathbf{Q} en un espacio finito Ω ,

$$c(\mathbf{Q}) \leq 1 - \frac{1}{|\Omega|} \min \{ \mathbf{Q}(x,y) : x,y \in \Omega \} \leq 1 - \min \{ \mathbf{Q}(x,y) : x,y \in \Omega \}.$$

En particular, si $\mathbf{Q}(x,y) > 0 \forall x,y \in \Omega$, entonces $c(\mathbf{Q}) < 1$.

A.1.3. Convergencia de cadenas de Markov Homogéneas

Según la **Definición A.1.2** una cadena de Markov se dice *homogénea* si todas las probabilidades de transición son iguales.

Definición A.1.6 Sea \mathbf{P} kernel de Markov, se dice que \mathbf{P} es *primitivo* si tiene una potencia \mathbf{P}^r estrictamente positiva. Una cadena de Markov homogénea con kernel \mathbf{P} primitivo puede alcanzar en algún instante cualquier estado con probabilidad positiva desde cualquier estado. Esta propiedad suele denominarse *irreducibilidad*. Desde un punto de vista probabilístico, cuando la matriz de las probabilidades de transición de una cadena de Markov homogénea es primitiva se dice que es *irreducible* y *aperiódica*.

Lema A.1.4 Para cada kernel de Markov \mathbf{P} en un espacio finito de configuraciones la sucesión $(c(\mathbf{P}^n))_{n \geq 0}$ decrece. Si \mathbf{P} es primitivo $(c(\mathbf{P}^n))_{n \geq 0}$ decrece a 0.

Definición A.1.7 Sea μ una distribución de probabilidad en Ω y \mathbf{P} un kernel de Markov en Ω , se dice que μ es *invariante* o *estacionaria* para \mathbf{P} si y sólo si $\mu \mathbf{P} = \mu$.

Si μ es estacionaria, entonces $\mu P^n = \mu \forall n \geq 0$. Este tipo de distribuciones de probabilidad son candidatas para la distribución en el límite de las cadenas de Markov homogéneas.

Teorema A.1.1 Un kernel de Markov P primitivo en un espacio finito Ω tiene una única distribución invariante μ y $\lim_{n \rightarrow \infty} \nu P^n = \mu$ uniformemente en todas las distribuciones ν .

La existencia y unicidad de la distribución invariante μ forma parte del **teorema A.1.2** conocido como Teorema de Perron-Frobenius.

Teorema A.1.2 Sea P un kernel de Markov primitivo en un espacio finito Ω , entonces se verifican las siguientes condiciones:

1. $\lambda = 1$ es un autovalor de P de multiplicidad uno.
2. El autoespacio por la derecha asociado a $\lambda = 1$ está generado por la distribución uniforme y el autoespacio por la izquierda está generado por una distribución μ . Esta distribución μ es la única distribución invariante de P y es siempre estrictamente positiva.
3. $|\gamma| < \lambda$, para todo autovalor γ , $\gamma \neq \lambda$.

Las cadenas homogéneas con kernel de Markov primitivo satisfacen el **Teorema A.1.3** conocido como la Ley de los Grandes Números para cadenas de Markov homogéneas.

Teorema A.1.3 Sea P un kernel de Markov primitivo en un espacio finito Ω con distribución invariante μ , entonces para cada distribución inicial ν y para cada función f en Ω , se verifica que¹⁴

$$\frac{1}{n} \sum_{i=1}^n f(\mathfrak{S}_i) \rightarrow E_{\mu}(f), \text{ cuando } n \rightarrow \infty.$$

¹⁴ $E_{\mu}(f)$ representa la esperanza matemática de la función f en un espacio Ω con respecto a la distribución invariante μ , y se determina $E_{\mu}(f) = \sum_{x \in \Omega} f(x)\mu(x)$.

en $L^2(\mathbf{P}_v, \mathbf{P}^{15})$. $(\xi_i)_{i \geq 0}$ es la sucesión de variables aleatorias correspondiente a la cadena de Markov para cada distribución inicial v y cada kernel \mathbf{P} . Además, $\forall \varepsilon > 0$, se tiene que

$$\mathbf{P} \left(\left| \frac{1}{n} \sum_{i=1}^n f(\xi_i) - \mathbf{E}_{\mathbf{P}}(f) \right| > \varepsilon \right) \leq \frac{13 \|f\|^2}{(1 - c(\mathbf{P})) n \varepsilon^2}.$$

Donde $\|f\| = \sum_{x \in \Omega} |f(x)|$.

Si $(\xi_i)_{i \geq 0}$ es la sucesión de variables aleatorias independientes idénticamente distribuidas, el kernel de Markov $\mathbf{P}(x, y)$ no depende de x , y por consiguiente, las filas de la matriz son iguales y $c(\mathbf{P}) = 0$.

A.2 Algoritmo Gibbs Sampler

Para obtener una realización del CAG en Ω

$$\Pi(x) = \frac{e^{-\mathbf{H}(x)}}{\sum_{y \in \Omega} e^{-\mathbf{H}(y)}}.$$

Se requiere simular una cadena de Markov con distribución límite Π . Para ello es necesario hallar un kernel de Markov \mathbf{P} estrictamente positivo o al menos primitivo, para que Π sea una distribución invariante. Esto puede lograrse a partir de las características locales del campo Π .

Definición A.2.1 Sea Π un CAG en Ω con función de energía \mathbf{H} en Ω , para cada $\mathbf{N} \subset \mathbf{S}$ se define el kernel de Markov en Ω , denotado $\Pi_{\mathbf{N}}$, como

$$\Pi_{\mathbf{N}}(\mathbf{x}, \mathbf{y}) = \begin{cases} \frac{e^{-\mathbf{H}(\mathbf{y}_{\mathbf{N}}^{\mathbf{x}_{\mathbf{S}/\mathbf{N}}})}}{Z_{\mathbf{N}}} & \text{si } \mathbf{y}_{\mathbf{S}/\mathbf{N}} = \mathbf{x}_{\mathbf{S}/\mathbf{N}} \\ 0 & \text{en el resto} \end{cases}, \quad (\text{A.1})$$

¹⁵ $\mathbf{P}_{v, \mathbf{P}}$ representa la ley de probabilidades que genera la cadena de Markov para la distribución inicial v y el kernel \mathbf{P} .

donde la función de partición viene dada por

$$Z_{\mathbf{N}} = \sum_{z_{\mathbf{N}} \in \Omega_{\mathbf{N}}} e^{-\mathbf{H}(z_{\mathbf{N}}^{\mathbf{x}} \mathbf{s}/\mathbf{N})}. \quad (\text{A.2})$$

Estos kernels representan la probabilidad de transición de la configuración $\mathbf{x} \in \Omega$ a la configuración $\mathbf{y} \in \Omega$ y se denominan *características locales del campo* Π .

Se está interesado en construir una cadena de Markov homogénea para la cual la distribución que se desea muestrear, Π , sea invariante (estacionaria). A menudo, se emplean cadenas de Markov homogéneas *reversibles en tiempo* que satisfacen la condición restrictiva del *balance detallado* (en inglés, detailed balanced).

Definición A.2.2 Sea μ una distribución de probabilidad y \mathbf{P} un kernel de Markov cualesquiera en Ω . Se dice que el kernel \mathbf{P} es *reversible* con respecto a la distribución μ si verifican la ecuación de balance detallado

$$\mu(\mathbf{x}) \mathbf{P}(\mathbf{x}, \mathbf{y}) = \mu(\mathbf{y}) \mathbf{P}(\mathbf{y}, \mathbf{x}), \quad \forall \mathbf{x}, \mathbf{y} \in \Omega.$$

Esto significa que la cadena de Markov homogénea dada por μ y \mathbf{P} es reversible en tiempo.

Lema A.2.1 Para todo $\mathbf{N} \subset \mathbf{S}$, sean $\Pi_{\mathbf{N}}$ características locales del CAG Π definido en Ω . Los kernels de Markov $\Pi_{\mathbf{N}}$ son reversibles con respecto a Π . En otras palabras, $\Pi_{\mathbf{N}}$ y Π , satisfacen la condición del balance detallado,

$$\Pi(\mathbf{x}) \Pi_{\mathbf{N}}(\mathbf{x}, \mathbf{y}) = \Pi(\mathbf{y}) \Pi_{\mathbf{N}}(\mathbf{y}, \mathbf{x}), \quad \forall \mathbf{x}, \mathbf{y} \in \Omega.$$

Teorema A.2.1 Sea μ una distribución de probabilidad y \mathbf{P} un kernel de Markov en Ω . Si el kernel \mathbf{P} es *reversible* con respecto a la distribución μ , entonces μ es una distribución invariante para \mathbf{P} .

$$\mu = \mu \mathbf{P}.$$

En efecto,

$$\sum_{\mathbf{y}} \mu(\mathbf{y}) \mathbf{P}(\mathbf{y}, \mathbf{x}) = \sum_{\mathbf{y}} \mu(\mathbf{x}) \mathbf{P}(\mathbf{x}, \mathbf{y}) = \mu(\mathbf{x}) \sum_{\mathbf{y}} \mathbf{P}(\mathbf{x}, \mathbf{y}) = \mu(\mathbf{x}).$$

En particular, si se considera el lema A.2.1, un CAG Π es invariante para sus características locales $\Pi_{\mathbf{N}}$, para todo $\mathbf{N} \subset \mathbf{S}$, luego,

$$\Pi \Pi_{\mathbf{N}} = \Pi.$$

Definición A.2.3 Una distribución *de exploración o de propuesta* es una probabilidad cualquiera $\mathbf{G}(\mathbf{s})$ definida sobre el conjunto \mathbf{S} .

Generalmente $\mathbf{G}(\mathbf{s})$ es una distribución uniforme. Se define un kernel de Markov $\hat{\Pi}$

$$\hat{\Pi}(\mathbf{x}, \mathbf{y}) = \begin{cases} \mathbf{G}(\mathbf{s})\Pi_{\{\mathbf{s}\}}(\mathbf{x}, \mathbf{y}) & \text{si } \mathbf{s}/\{\mathbf{s}\} = \mathbf{x}/\{\mathbf{s}\} \\ 0 & \text{en el resto} \end{cases} = \sum_{\mathbf{s} \in \mathbf{S}} \mathbf{G}(\mathbf{s})\Pi_{\{\mathbf{s}\}}(\mathbf{x}, \mathbf{y}), \text{ para algún } \mathbf{s} \in \mathbf{S} \quad (\text{A.3})$$

donde $\Pi_{\{\mathbf{s}\}}$, representan las características locales del CAG Π , con $\mathbf{s} \in \mathbf{S}$.

Luego el kernel de Markov para cada barrido se define como,

$$\tilde{\mathbf{P}} = \hat{\Pi}^{|\mathbf{S}|}. \quad (\text{A.4})$$

La cadena de Markov homogénea con matriz de probabilidades de transición, $\tilde{\mathbf{P}}$, definida por las expresiones (A.3) y (A.4), inducen el algoritmo B.2 para el esquema de visita aleatorio dado por la distribución de exploración $\mathbf{G}(\mathbf{s})$.

Lema A.2.2 Si Π es una CAG en Ω y $\tilde{\mathbf{P}}$ un kernel de Markov en Ω , para cada barrido de un esquema de visita aleatorio con distribución \mathbf{G} en \mathbf{S} , definido por las expresiones (A.3) y (A.4), entonces si \mathbf{G} es estrictamente positivo, $\tilde{\mathbf{P}}$ también lo es.

Teorema A.2.2 Si Π es una CAG en Ω y $\tilde{\mathbf{P}}$ un kernel de Markov en Ω , para cada barrido de un esquema de visita aleatorio con distribución \mathbf{G} en \mathbf{S} , definido por las expresiones (A.3) y (A.4), entonces si \mathbf{G} es estrictamente positivo para cada $\mathbf{x} \in \Omega$,

$$\lim_{n \rightarrow \infty} \mathbf{v}\tilde{\mathbf{P}}^n(\mathbf{x}) = \Pi(\mathbf{x}),$$

en L^1 uniformemente en todas las distribuciones iniciales \mathbf{v} definidas en Ω .

Definición A.2.4 Sea \mathbf{H} la función de energía definida en Ω y sea $\mathbf{s} \in \mathbf{S}$ una posición cualesquiera, la oscilación local $\delta_{\mathbf{s}}$ de \mathbf{H} en la posición \mathbf{s} se define,

$$\delta_{\mathbf{s}} = \sup_{\mathbf{x}, \mathbf{y} \in \Omega} \left\{ |\mathbf{H}(\mathbf{x}) - \mathbf{H}(\mathbf{y})| / : \mathbf{x}_{\mathbf{S} - \{\mathbf{s}\}} = \mathbf{y}_{\mathbf{S} - \{\mathbf{s}\}} \right\}. \quad (\text{A.5})$$

La oscilación local máxima Δ de \mathbf{H} es

$$\Delta = \max_{\mathbf{s} \in \mathbf{S}} \{\delta_{\mathbf{s}}\}. \quad (\text{A.6})$$

Lema A.2.3 Para el kernel de Markov $\tilde{\mathbf{P}}$ en Ω , definido por las expresiones (A.3) y (A.4), considerando un esquema de visita aleatorio con distribución de exploración \mathbf{G} en \mathbf{S} , se verifica que

$$c(\tilde{\mathbf{P}}) \leq 1 - \eta e^{-|\mathbf{S}|\Delta},$$

donde Δ es la oscilación local máxima y

$$\eta = \min_{\mathbf{s} \in \mathbf{S}} \left\{ \mathbf{G}(\mathbf{s})^{|\mathbf{S}|} \right\}. \quad (\text{A.7})$$

Entonces si \mathbf{H} no es una función constante y \mathbf{G} es estrictamente positivo, $c(\tilde{\mathbf{P}}) < 1$.

Demostración. Sea $\mathbf{x} \in \Omega$, y $z_{\mathbf{s}}$ un minimizador local en \mathbf{s} tal que

$$\mathbf{H}(z_{\mathbf{s}} \mathbf{x}_{\mathbf{S} - \{\mathbf{s}\}}) = m_{\mathbf{s}} = \min \left\{ \mathbf{H}(v_{\mathbf{s}} \mathbf{x}_{\mathbf{S} - \{\mathbf{s}\}}) : v_{\mathbf{s}} \in \Lambda_{\mathbf{s}} \right\}.$$

Las características locales $\Pi_{\{\mathbf{s}\}}$ se determinan a partir de (A.1) y (A.2)

$$\Pi_{\{\mathbf{s}\}}(\mathbf{x}, \mathbf{y}) = \begin{cases} \frac{e^{-\mathbf{H}(y_{\mathbf{s}} \mathbf{x}_{\mathbf{S} - \{\mathbf{s}\}})}}{\sum_{v_{\mathbf{s}} \in \Lambda_{\mathbf{s}}} e^{-\mathbf{H}(v_{\mathbf{s}} \mathbf{x}_{\mathbf{S} - \{\mathbf{s}\}})}} & \text{si } y_{\mathbf{S} - \{\mathbf{s}\}} = \mathbf{x}_{\mathbf{S} - \{\mathbf{s}\}} \\ 0 & \text{en el resto} \end{cases}$$

Luego se tiene que

$$\Pi_{\{s\}}(\mathbf{x}, \mathbf{y}_s, \mathbf{x}_{\mathbf{S}/\{s\}}) = \frac{e^{-\mathbf{H}(\mathbf{y}_s, \mathbf{x}_{\mathbf{S}/\{s\}}) - m_s}}{\sum_{\mathbf{v}_s \in \Lambda_s} e^{-\mathbf{H}(\mathbf{v}_s, \mathbf{x}_{\mathbf{S}/\{s\}}) - m_s}}.$$

Puede probarse que $\mathbf{H}(\mathbf{y}_s, \mathbf{x}_{\mathbf{S}/\{s\}}) - m_s \leq \delta_s$, siendo δ_s la oscilación local de \mathbf{H} en la posición s , dada por la expresión (A.5) y que $\mathbf{H}(\mathbf{y}_s, \mathbf{x}_{\mathbf{S}/\{s\}}) - m_s \geq 0$. Por lo tanto puede escribirse

$$\Pi_{\{s\}}(\mathbf{x}, \mathbf{y}_s, \mathbf{x}_{\mathbf{S}/\{s\}}) \geq \frac{e^{-\delta_s}}{|\Lambda_s|}.$$

Para el kernel de Markov $\hat{\Pi}$ definido por la expresión (A.3)

$$\hat{\Pi}_{\{s\}}(\mathbf{x}, \mathbf{y}_s, \mathbf{x}_{\mathbf{S}/\{s\}}) = \mathbf{G}(s) \Pi_{\{s\}}(\mathbf{x}, \mathbf{y}_s, \mathbf{x}_{\mathbf{S}/\{s\}}) \geq \mathbf{G}(s) \frac{e^{-\delta_s}}{|\Lambda_s|}.$$

Para el kernel de Markov $\tilde{\mathbf{P}}$ para cada barrido definido en (A.4), se tiene que,

$$\min_{\mathbf{x}, \mathbf{y} \in \Omega} \hat{\mathbf{P}}(\mathbf{x}, \mathbf{y}) \geq \prod_{s=1}^{|\mathbf{S}|} \min_{\mathbf{s} \in \mathbf{S}} \left\{ \mathbf{G}(s) \frac{e^{-\delta_s}}{|\Lambda_s|} \right\} = \min_{\mathbf{s} \in \mathbf{S}} \left\{ \mathbf{G}(s)^{|\mathbf{S}|} \right\} \frac{e^{-\Delta|\mathbf{S}|}}{|\Omega|} = \eta \frac{e^{-\Delta|\mathbf{S}|}}{|\Omega|}$$

donde Δ es la oscilación local máxima definida en (A.6).

Mediante el lema A.1.3. con $\tilde{\mathbf{P}}$ se tiene finalmente que

$$c(\tilde{\mathbf{P}}) \leq 1 - |\Omega| \min_{\mathbf{x}, \mathbf{y} \in \Omega} \tilde{\mathbf{P}}(\mathbf{x}, \mathbf{y}) \leq 1 - \eta e^{(-|\mathbf{S}|\Delta)}.$$

Si \mathbf{G} es estrictamente positivo, entonces $\eta > 0$, luego

$$\min_{\mathbf{x}, \mathbf{y} \in \Omega} \tilde{\mathbf{P}}(\mathbf{x}, \mathbf{y}) > 0,$$

y el kernel es primitivo, irreducible y aperiódico, por lo tanto $c(\tilde{\mathbf{P}}) < 1$. ♦

El siguiente teorema es la ley débil de los grandes números del algoritmo GS.

Teorema A.2.3 Sea Π es una CAG en Ω , con función de energía \mathbf{H} en Ω con oscilación local máxima Δ . Sea $(\xi_i)_{i \geq 0}$ la sucesión de variables aleatorias inducidas

por la cadena de Markov homogénea con distribución inicial \mathbf{v} arbitraria en Ω , y el kernel primitivo $\tilde{\mathbf{P}}$ en Ω , en el caso de un esquema de visita aleatorio con distribución de exploración \mathbf{G} en \mathbf{S} . Entonces para cada función f definida en Ω , se verifica que

$$\frac{1}{n} \sum_{i=1}^{n-1} f(\mathbf{S}_i) \rightarrow E_{\Pi}(f(\mathbf{X})), \text{ cuando } n \rightarrow \infty.$$

en $L^2(P_{\mathbf{v}, \tilde{\mathbf{P}}})$. Además, en probabilidad $P_{\mathbf{v}, \tilde{\mathbf{P}}}$, $\forall \varepsilon > 0$, se tiene que

$$\mathbf{P} \left(\left| \frac{1}{n} \sum_{i=1}^{n-1} f(\mathbf{S}_i) - E_{\Pi}(f(\mathbf{X})) \right| > \varepsilon \right) \leq c \frac{e^{|\mathbf{S}|\Delta}}{n\varepsilon^2}.$$

donde $c = \frac{13\|f\|^2}{\eta}$ y η definido por la ecuación (A.7).

Demostración. Por el teorema A.2.1 se concluye que la distribución Π es invariante con respecto a la cadena de Markov de kernel $\tilde{\mathbf{P}}$. Puesto que $\tilde{\mathbf{P}}$ es primitivo, si se aplica el teorema A.1.3 para $\tilde{\mathbf{P}}$ con distribución invariante Π , se obtiene

$$\mathbf{P} \left(\left| \frac{1}{n} \sum_{i=1}^{n-1} f(\mathbf{S}_i) - E_{\Pi}(f(\mathbf{X})) \right| > \varepsilon \right) \leq \frac{13\|f\|^2}{(1-c(\tilde{\mathbf{P}}))n\varepsilon^2}.$$

A partir del lema A.2.3 se tiene que

$$1 - c(\tilde{\mathbf{P}}) \geq \eta e^{(-|\mathbf{S}|\Delta)},$$

Por lo que

$$\frac{13\|f\|^2}{(1-c(\tilde{\mathbf{P}}))n\varepsilon^2} \leq \frac{13\|f\|^2}{\eta n\varepsilon^2} e^{(-|\mathbf{S}|\Delta)} = c \frac{e^{(-|\mathbf{S}|\Delta)}}{\eta n\varepsilon^2}. \blacklozenge$$

A.3 Método de Maximización

A.3.1 Preliminares

Definición A.3.1.1 Un subconjunto \mathbf{B} de \mathfrak{R}^n se dice que es *convexo* si para todo $\mathbf{x}, \mathbf{y} \in \mathbf{B}$, el segmento de línea $[\mathbf{x}, \mathbf{y}] = \{ \lambda \mathbf{x} + (1-\lambda) \mathbf{y} : 0 \leq \lambda \leq 1 \}$ esta contenido en \mathbf{B} .

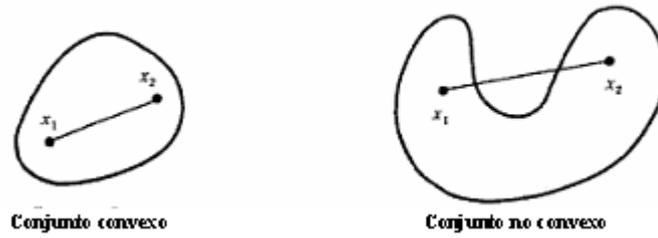


Figura A.3.1 Conjunto convexo y no convexo.

Definición A.3.1.2 Sean $\mathbf{x} = (x_1, \dots, x_n), \mathbf{y} = (y_1, \dots, y_n) \in \mathfrak{R}^n$, se define el *producto escalar Euclídeo* como

$$\langle \mathbf{x}, \mathbf{y} \rangle = \sum_i x_i y_i.$$

Definición A.3.1.3 Sea $\mathbf{x} = (x_1, \dots, x_n) \in \mathfrak{R}^n$, se define la *norma Euclídea* como

$$\|\mathbf{x}\| = \sqrt{\sum_i x_i^2} = \langle \mathbf{x}, \mathbf{x} \rangle.$$

Definición A.3.1.4 Una función a valores reales f , definida en \mathfrak{R}^n se dice *continua en* $\mathbf{x}_0 \in \mathfrak{R}^n$ si para cualquier $\varepsilon > 0$, existe un $\delta(\varepsilon, \mathbf{x}_0) > 0$ tal que $\|\mathbf{x} - \mathbf{x}_0\| < \delta$ entonces $|f(\mathbf{x}) - f(\mathbf{x}_0)| < \varepsilon$.

Definición A.3.1.5 Una función a valores reales f , definida en \mathfrak{R}^n se dice *uniformemente continua* en $\mathbf{D} \subset \mathfrak{R}^n$, si para cualquier $\varepsilon > 0$, existe un $\delta > 0$ tal que $\forall \mathbf{x}, \mathbf{y} \in \mathbf{D}$ con $\|\mathbf{x} - \mathbf{y}\| < \delta$ entonces $|f(\mathbf{x}) - f(\mathbf{y})| < \varepsilon$.

Definición A.3.1.6 Si una función real f , definida en \mathfrak{R}^n es diferenciable, el *gradiente* de f en el punto \mathbf{x} se define como:

$$\nabla f(\mathbf{x}) = \left(\frac{\partial}{\partial x_1} f(\mathbf{x}), \dots, \frac{\partial}{\partial x_n} f(\mathbf{x}) \right),$$

donde $\frac{\partial}{\partial x_i} f(\mathbf{x})$ representa la derivada parcial de la función f respecto de x_i en el punto \mathbf{x} .

Si además f , es dos veces diferenciable en \mathfrak{R}^n , la matriz *Hessiana* en el punto \mathbf{x} esta dada por

$$\nabla^2 f(\mathbf{x}) = \left(\frac{\partial^2}{\partial x_i \partial x_j} f(\mathbf{x}) \right)_{1 \leq i, j \leq n},$$

donde $\frac{\partial}{\partial x_i \partial x_j} f(\mathbf{x})$ representa la derivada parcial segunda de la función f respecto de x_i y a x_j en el punto \mathbf{x} .

Definición A.3.1.7 Una función a valores reales f definida en \mathfrak{R}^n , se dice *diferenciable continua* en $\mathbf{x} \in \mathfrak{R}^n$ si $\frac{\partial}{\partial x_i} f(\mathbf{x})$ existe y es continua, $i = 1, 2, \dots, n$. Una f que es *diferenciable continua* para todo punto del conjunto abierto $\mathbf{D} \subset \mathfrak{R}^n$, se denota $f \in C^1(\mathbf{D})$. Una función f a valores reales f definida en \mathfrak{R}^n , se denomina *dos veces diferenciable continua* en $\mathbf{x} \in \mathfrak{R}^n$ si $\frac{\partial^2}{\partial x_i \partial x_j} f(\mathbf{x})$ existe y es continua, $i, j = 1, 2, \dots, n$. Si f es dos veces diferenciable continua para todo punto del conjunto abierto $\mathbf{D} \subset \mathfrak{R}^n$, decimos que $f \in C^2(\mathbf{D})$.

Teorema A.3.1.1 Si una función a valores reales f definida en \mathfrak{R}^n es dos veces diferenciable, entonces la matriz Hesiana $\nabla^2 f(\mathbf{x})$ es simétrica, es decir,

$$\frac{\partial^2}{\partial x_i \partial x_j} f(\mathbf{x}) = \frac{\partial^2}{\partial x_j \partial x_i} f(\mathbf{x}).$$

Definición A.3.1.8 Una función real f definida en un subconjunto convexo no vacío \mathbf{D} de \mathfrak{R}^n se dice que f es convexa en \mathbf{D} si

$$f(\lambda \mathbf{x} + (1 - \lambda) \mathbf{y}) \leq \lambda f(\mathbf{x}) + (1 - \lambda) f(\mathbf{y}) \quad \forall \mathbf{x}, \mathbf{y} \in \mathbf{D} \text{ y } 0 \leq \lambda \leq 1.$$

Si la desigualdad es estricta, f se dice *estrictamente convexa*. Si existe una constante $c > 0$ tal que para cada $\mathbf{x}, \mathbf{y} \in \mathbf{D}$

$$f(\lambda \mathbf{x} + (1 - \lambda) \mathbf{y}) \leq \lambda f(\mathbf{x}) + (1 - \lambda) f(\mathbf{y}) - \frac{1}{2} c \lambda (1 - \lambda) \|\mathbf{x} - \mathbf{y}\|^2, \quad 0 \leq \lambda \leq 1,$$

entonces f se dice *uniformemente convexa* en \mathbf{D} . Si $-f$ es una función convexa (estrictamente convexa, uniformemente convexa), entonces f es una función cóncava (estrictamente cóncava, uniformemente cóncava).

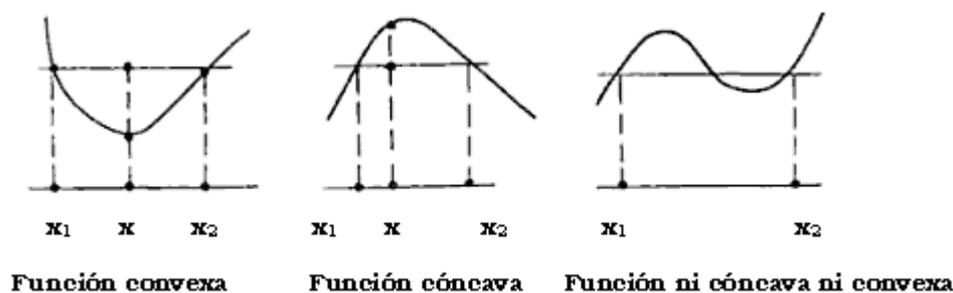


Figura A.3.2 Función convexa y función cóncava.

Definición A.3.1.9 Una matriz \mathbf{A} se dice *semi-definida negativa* si $z \mathbf{A} z^T \leq 0, \forall z \in \mathfrak{R}^n$ (donde z es un vector fila y z^T su traspuesto). Se dice que \mathbf{A} es *definida negativa* si la desigualdad es estricta. \mathbf{A} es *(semi-) definida positiva* si $-\mathbf{A}$ es *(semi-)definida negativa*.

Teorema A.3.1.2 Sea \mathbf{D} un subconjunto convexo de \mathfrak{R}^n y sea $f: \mathbf{D} \rightarrow \mathfrak{R}$ una función tal que $f \in C^2(\mathbf{D})$. Entonces

- (a) f es convexa si y sólo si su matriz Hessiana semidefinida positiva en cada punto de \mathbf{D} .
- (b) f es estrictamente convexa si su matriz Hessiana es definida positiva en cada punto de \mathbf{D} .
- (c) f es uniformemente convexa si y sólo si su matriz Hessiana es uniformemente definida positiva en cada punto de \mathbf{D} , es decir, existe una constante $m > 0$ tal que

$$m \|u\|^2 \leq u^T \nabla^2 f(x) u, \forall x \in \mathbf{D}, u \in \mathfrak{R}^n.$$

Teorema A.3.1.3 Sea \mathbf{D} un subconjunto no vacío convexo de \mathfrak{R}^n . f una función convexa definida en \mathbf{D} , $\alpha \in \mathfrak{R}$. Entonces el conjunto de nivel $L_\alpha = \{ x \in \mathbf{D} / f(x) \leq \alpha \}$ es un conjunto convexo.

Definición A.3.1.10 Sea f una función definida en un subconjunto convexo \mathbf{D} de \mathfrak{R}^n , se dice que x_0 es un máximo local o relativo de f si y sólo si existe una bola cerrada $\mathbf{B}(x_0, r)$ con $r > 0$ tal que

$$f(x) \leq f(x_0) \forall x \in \mathbf{B}(x_0, r).$$

Definición A.3.1.11 Se dice que x_0 es un mínimo local o relativo de f si y sólo si existe una bola cerrada $\mathbf{B}(x_0, r)$ con $r > 0$ tal que

$$f(\mathbf{x}_0) \leq f(\mathbf{x}), \forall \mathbf{x} \in \mathbf{B}(\mathbf{x}_0, r).$$

Teorema A.3.1.4 Sea f una función tal que $f \in C^2(\mathbf{D})$, donde \mathbf{D} un subconjunto no vacío convexo de \mathfrak{R}^n . Si se supone que existe un $m > 0$ tal que

$$\mathbf{u}^T \nabla^2 f(\mathbf{x}) \mathbf{u} \geq m \|\mathbf{u}\|^2, \forall \mathbf{x} \in L(\mathbf{x}_0), \mathbf{u} \in \mathfrak{R}^n$$

Entonces el conjunto de nivel $L(\mathbf{x}_0) = \{ \mathbf{x} \in \mathbf{D} / f(\mathbf{x}) \leq f(\mathbf{x}_0) \}$ es un conjunto convexo cerrado y acotado.

Lema A.3.1.2 Sea f es una función a valores reales dos veces diferenciable con derivadas segundas continuas definida en un subconjunto convexo abierto Θ de \mathfrak{R}^n . Si ∇f se anula en \mathbf{x}_0 y la matriz Hessiana $\nabla^2 f$ es *definida negativa* en \mathbf{x}_0 , entonces \mathbf{x}_0 es un máximo local de f .

A.3.2 Método de Newton

Es un método de optimización iterativo basado en aproximar la función objetivo mediante la serie de Taylor de segundo orden.

Sea $\varphi(\boldsymbol{\theta}) = -\mathbf{L}(\mathbf{x}; \boldsymbol{\theta}) = -\sum_{s \in \mathbf{S}} [\boldsymbol{\eta}_s x_s - M \ln(1 + e^{\boldsymbol{\eta}_s})]$ la función a minimizar, con $\boldsymbol{\theta} =$

(β_{i1}, β_{i2}) , $i = 1, 2$, según sea el modelo Autobinomial considerado. El propósito es determinar para qué valores del vector $\boldsymbol{\theta} = (\beta_{i1}, \beta_{i2}) \in \mathfrak{R}^2$, $i = 1, 2$, $\varphi(\boldsymbol{\theta})$ es mínima.

En $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, $\varphi(\boldsymbol{\theta})$ puede aproximarse utilizando el desarrollo cuadrático de Taylor,

$$\varphi(\boldsymbol{\theta}) \cong h(\boldsymbol{\theta}) = \varphi(\boldsymbol{\theta}_0) + \nabla \varphi(\boldsymbol{\theta}_0)^T (\boldsymbol{\theta} - \boldsymbol{\theta}_0) + \frac{1}{2} (\boldsymbol{\theta} - \boldsymbol{\theta}_0)^T H(\boldsymbol{\theta}_0) (\boldsymbol{\theta} - \boldsymbol{\theta}_0), \quad (\text{A.8})$$

donde $\nabla \varphi(\boldsymbol{\theta}_0)$ es el gradiente de $\varphi(\boldsymbol{\theta})$ y $H(\boldsymbol{\theta}_0)$ es el hessiano de $\varphi(\boldsymbol{\theta})$.

$h(\boldsymbol{\theta})$ es una función cuadrática, se minimiza resolviendo $\nabla h(\boldsymbol{\theta}) = 0$. dado que el gradiente de $h(\boldsymbol{\theta})$ es,

$$\nabla h(\boldsymbol{\theta}) = \nabla \varphi(\boldsymbol{\theta}_0)^T (\boldsymbol{\theta} - \boldsymbol{\theta}_0) + H(\boldsymbol{\theta}_0) (\boldsymbol{\theta} - \boldsymbol{\theta}_0),$$

debe resolverse la ecuación,

$$\nabla \varphi(\theta_0)^T + H(\theta_0) (\theta - \theta_0) = 0,$$

Si la matriz $H(\theta_0)$, es inversible, se tiene que

$$\begin{aligned} (\theta - \theta_0) &= -H(\theta_0)^{-1} \nabla \varphi(\theta_0)^T \\ \theta &= \theta_0 - H(\theta_0)^{-1} \nabla \varphi(\theta_0)^T \end{aligned} \quad (\text{A.9})$$

$d(\theta) = -H(\theta)^{-1} \nabla \varphi(\theta)^T$ representa, la dirección de búsqueda, se denomina *dirección de Newton*. La dirección de búsqueda se determina minimizando $h(\theta + d)$. La expresión (A.9) se emplea como ecuación de recurrencia para, dado un punto inicial θ_0 , generar una sucesión de puntos que van a converger bajo ciertas condiciones al óptimo de $\varphi(\theta)$. El paso \mathbf{k} del proceso iterativo, se formula:

$$\begin{aligned} d_k(\theta) &= -H(\theta_k)^{-1} \nabla \varphi(\theta_k)^T \\ \theta_{k+1} &= \theta_k + \alpha_k d_k \end{aligned} \quad (\text{A.10})$$

donde α_k es un escalar positivo que representa la magnitud del paso.

La selección del punto inicial que garantice la convergencia del método no es un asunto trivial. Dicho punto debe estar cerca del óptimo para que esto suceda.

Las ecuaciones (A.10) corresponden al siguiente algoritmo:

Algoritmo A.3 *Método de Newton con búsqueda lineal.*

1. Seleccionar el valor inicial $\theta_0 \in \mathfrak{R}^n$, un $\varepsilon > 0$.
2. Se fija $\mathbf{k} = 0$.
3. Calcular $\nabla \varphi(\theta_k)$.
4. Hacer lo siguiente:
 - a. Si $\|\nabla \varphi(\theta_k)\| \leq \varepsilon$, parar y el resultado es θ_k . Si no ir a **b.**
 - b. Resolver la ecuación: $H(\theta_k) d_k = -\nabla \varphi(\theta_k)^T$, respecto de d_k .
 - c. Encontrar el α_k tal que $\varphi(\theta_k + \alpha_k d_k) = \min_{\alpha > 0} \varphi(\theta_k + \alpha d_k)$.
 - d. Se fija $\theta_{k+1} = \theta_k + \alpha_k d_k$.
 - e. Se fija \mathbf{k} igual a $\mathbf{k}+1$.

f. Ir al paso 4..

Antes de probar la convergencia del algoritmo veamos el siguiente resultado.

Teorema A.3.2.1 Sea $\nabla\varphi(\theta)$ una función uniformemente continua en el conjunto de nivel $L(\theta_0) = \{\theta : \varphi(\theta) \leq \varphi(\theta_0)\}$. Sea β_k el ángulo entre $-\nabla\varphi(\theta_k)$ y la dirección d_k generado por el algoritmo **A.3.** acotado uniformemente por $\pi/2$, es decir,

$$\beta_k \leq \pi/2 - \mu, \text{ para algún } \mu > 0.$$

Entonces $\nabla\varphi(\theta_k) = 0$ para algún k o $\varphi(\theta) \rightarrow -\infty$ o $\nabla\varphi(\theta) \rightarrow 0$.

Teorema A.3.2.2 Sea $\varphi : \mathfrak{R}^n \rightarrow \mathfrak{R}$ una función diferenciable con derivadas segundas continuas en un conjunto abierto conexo $\mathbf{D} \subset \mathfrak{R}^n$. Para cada θ_0 existe una constante $m > 0$ tal que $\varphi(\theta)$ satisface:

$$u^T \nabla^2 \varphi(\theta) u \geq m \|u\|^2, \forall u \in \mathfrak{R}^n, \theta \in L(\theta_0), \quad (\text{A.11})$$

donde $L(\theta_0)$ el correspondiente conjunto de nivel. La sucesión $\{\theta_k\}$ generada por el algoritmo **A.3.** satisface:

- a. Si $\{\theta_k\}$ es una sucesión finita, $\nabla\varphi(\theta_k) = 0$ para algún k .
- b. Si $\{\theta_k\}$ es una sucesión infinita, $\{\theta_k\}$ converge a un único mínimo θ^* de φ .

Demostración. Por (A.11), φ es una función estrictamente convexa en \mathfrak{R}^n , por lo tanto, el punto estacionario, es el único mínimo global (absoluto).

Luego el conjunto de nivel $L(\theta_0)$ es conjunto convexo cerrado y acotado. Como $\{\varphi(\theta_k)\}$ es monótona decreciente $\{\theta_k\} \subset L(\theta_0)$ y $\{\theta_k\}$ es acotada. Por lo tanto existe un punto $\theta^* \in L(\theta_0)$, tal que $\theta_k \rightarrow \theta^*$, más aún $\varphi(\theta_k) \rightarrow \varphi(\theta^*)$. Al ser $\varphi \in C^2(\mathbf{D})$, por el teorema A.3.2.1. se tiene que $\nabla\varphi(\theta_k) \rightarrow \nabla\varphi(\theta^*) = 0$. Finalmente, como el punto estacionario es único, entonces $\{\theta_k\}$ converge a un único mínimo θ^* de φ . ♦

ALGORITMOS DE LOS MÉTODOS MONTE CARLO

B.1 Algoritmo de Metropolis

Versión básica del algoritmo de Metrópolis cuando se utiliza el esquema de visita secuencial.

Algoritmo

1. Muestrear al azar una configuración \mathbf{x}^0 de una distribución arbitraria ν en Ω .
2. Se fija $\mathbf{n} = 1$.
3. Hacer lo siguiente:
 - a. Se fija la configuración \mathbf{y}^0 igual a la configuración $\mathbf{x}^{\mathbf{n}-1}$.
 - b. Se fija la posición $\mathbf{s} = 1$.
 - c. Hacer lo siguiente:
 - i. Muestrear al azar un valor $\lambda_{\mathbf{s}} \in \Lambda = \{ 0, 1, \dots, 255 \}$ de una distribución arbitraria.
 - ii. Determinar $e^{-[\mathbf{H}(\lambda_{\mathbf{s}}, \partial_{\mathbf{s}}; \boldsymbol{\theta}) - \mathbf{H}(\mathbf{x}_{\mathbf{s}}, \partial_{\mathbf{s}}; \boldsymbol{\theta})]}$.
 - iii. Sea $\alpha = \min\{ 1, e^{-[\mathbf{H}(\lambda_{\mathbf{s}}, \partial_{\mathbf{s}}; \boldsymbol{\theta}) - \mathbf{H}(\mathbf{x}_{\mathbf{s}}, \partial_{\mathbf{s}}; \boldsymbol{\theta})]} \}$.
 - iv. Se selecciona al azar un valor \mathbf{u} de la distribución $\mathbf{U}(0, 1)$.
 - v. Si $\mathbf{u} < \alpha$ entonces fijar $\mathbf{y}_{\mathbf{s}} = \lambda_{\mathbf{s}}$, sino $\mathbf{y}_{\mathbf{s}} = \mathbf{x}_{\mathbf{s}}$.
 - d. Se fija $\mathbf{s} = \mathbf{s} + 1$.
 - e. Si $\mathbf{s} \leq |\mathbf{S}|$ ir a c.
 - f. Fijar la configuración $\mathbf{x}^{\mathbf{n}}$ igual a la configuración $\mathbf{y}^{|\mathbf{S}|}$.
4. Se fija \mathbf{n} igual a $\mathbf{n} + 1$.
5. Ir al paso 3.

Todos los algoritmos que se presentan a continuación fueron desarrollados mediante el software matemático **MATLAB** (abreviatura de *MATrix LABoratory*).

B.2. Algoritmo de Gibbs Sampler

Versión básica del algoritmo Gibbs Sampler que permite obtener una realización de un Campo aleatorio de Gibbs, a partir de sus características locales, cuando se utiliza el esquema de visita aleatorio. Este algoritmo se implementó a partir de los resultados del apartado A.2

B.2.1. Algoritmo

1. Muestrear al azar una configuración \mathbf{x}^0 de una distribución arbitraria ν en Ω .
2. Fijar $n = 1$.
3. Hacer lo siguiente:
 - a. Fijar la configuración \mathbf{y}^0 igual a la configuración \mathbf{x}^{n-1} .
 - b. Fijar $t = 1$.
 - c. Muestrear al azar un sitio \mathbf{s} a través de la distribución Uniforme \mathbf{G} en \mathbf{S} (conjunto de sitios).
 - d. Hacer lo siguiente:
 - i. Determinar la característica local $\Pi_{\{\mathbf{s}\}}(\mathbf{y}^{t-1}, X_{\mathbf{s}} \mathbf{y}_{\mathbf{S}-\{\mathbf{s}\}}^{t-1})$.
 - ii. Cambiar la posición \mathbf{s} de la configuración anterior \mathbf{y}^{t-1} muestreando al azar $X_{\mathbf{s}}$ de $\Pi_{\{\mathbf{s}\}}(\mathbf{y}^{t-1}, X_{\mathbf{s}} \mathbf{y}_{\mathbf{S}-\{\mathbf{s}\}}^{t-1})$, obteniendo una nueva configuración \mathbf{y}^t .
 - e. Se fija $t = t + 1$.
 - f. Si $t \leq |\mathbf{S}|$ ir a c.
 - g. Fijar la configuración \mathbf{x}^n igual a la configuración $\mathbf{y}^{|\mathbf{S}|}$.
4. Se fija n igual a $n + 1$.

5. Ir al paso 3.

El paso 3. del algoritmo va actualizando la configuración inicial \mathbf{x}^0 al visitar aleatoriamente los sitios \mathbf{s} de la grilla \mathbf{S} , regidas por una distribución de probabilidad de *exploración* o de *propuesta* \mathbf{G} en \mathbf{S} . Cuando visita el sitio \mathbf{s} , en la configuración \mathbf{y} , quedan fijos los valores \mathbf{x}_j , con $j \neq \mathbf{s}$ y cambia solamente el valor de \mathbf{x}_s con probabilidad $\Pi_{\{\mathbf{s}\}}(\mathbf{y})$ definida a través de características locales. Esta $\Pi_{\{\mathbf{s}\}}$ es la probabilidad de transición para una cadena de Markov $\mathbf{X}(n)$.

El procedimiento de evaluación de $\Pi_{\{\mathbf{s}\}}$ resulta ser en general, bastante sencillo gracias a la naturaleza de las funciones potenciales del modelo Autobinomial.

B.2.2 Algoritmo Gibbs sampler H-V

Versión del algoritmo Gibbs Sampler H-V que permite obtener una realización de un Campo aleatorio de Gibbs - Autobinomial H-V, a partir de sus características locales, cuando se utiliza el esquema de visita aleatorio.

Algoritmo

1. Fijar el tamaño de la imagen.
2. Muestrear de forma aleatoria una configuración inicial \mathbf{x}^0 de la distribución uniforme en $\{0, \dots, 255\}$.
3. Fijar los valores de los parámetros del modelo Autobinomial H-V, de β_{11} y de β_{12} , para generar la realización del campo.
4. Fijar $n = 1$ (inicialización del número de barridos).
5. Hacer lo siguiente:
 - a. Fijar la configuración \mathbf{y}^0 igual a la \mathbf{x}^{n-1} .
 - b. Fijar $r = 1$.
 - c. Hacer lo siguiente:
 - c.1. Muestrear al azar un sitio \mathbf{s} a visitar a través de la distribución uniforme \mathbf{G} en \mathbf{S} .
 - c.2. Calcular la probabilidad de éxito $w = 1/(1 + \exp(-\text{exponente}))$, donde

$$\text{exponente} = \beta_{11} \frac{(y_t + y_{t^*})}{M} + \beta_{12} \frac{(y_u + y_{u^*})}{M}.$$

(donde y_t , y_{t^*} , y_u , y_{u^*} , representan los niveles de gris de los vecinos horizontales y verticales de \mathbf{y}_s respectivamente, característica local).

c.3. Cambiar la posición \mathbf{s} de la configuración \mathbf{y}^{r-1} muestreando de forma aleatoria \mathbf{X}_s de la distribución Binomial de parámetros \mathbf{M} y \mathbf{w} . De esta manera, se obtiene una nueva configuración \mathbf{y}^r .

d) Fijar \mathbf{r} igual a $\mathbf{r}+1$.

f) Si $\mathbf{r} \leq |\mathbf{S}|$ ir a **c.**

g) Fijar la configuración \mathbf{x}^n igual a la $\mathbf{y}^{|\mathbf{s}|}$.

6. Fijar \mathbf{n} igual a $\mathbf{n}+1$.

7. Ir al paso **5**.

B.2.3 Algoritmo Gibbs sampler I-D

Versión del algoritmo Gibbs Sampler I-D que permite obtener una realización de un Campo aleatorio de Gibbs-Autobinomial I-D, a partir de sus características locales, cuando se utiliza el esquema de visita aleatorio.

Algoritmo

1. Fijar el tamaño de la imagen.

2. Muestrear de forma aleatoria una configuración inicial \mathbf{x}^0 de la distribución uniforme en $\{0, \dots, 255\}$.

3. Fijar los valores de los parámetros del modelo Autobinomial I-D, de β_{21} y de β_{22} , para generar la realización del campo.

4. Fijar $\mathbf{n} = 1$ (inicialización del número de barridos).

5. Hacer lo siguiente:

a. Fijar la configuración \mathbf{y}^0 igual a la $\mathbf{x}^{\mathbf{n}-1}$.

b. Fijar $r = 1$.

c. Hacer lo siguiente:

c.1. Muestrear al azar un sitio s a visitar a través de la distribución uniforme G en S .

c.2. Calcular la probabilidad de éxito $w = 1/(1+\exp(-\text{exponente}))$, donde

$$\text{exponente} = \beta_{21} \frac{(y_v + y_{v^*})}{M} + \beta_{22} \frac{(y_z + y_{z^*})}{M}$$

(donde $y_v, y_{v^*}, y_z, y_{z^*}$, representan los niveles de gris de los vecinos de y_s de la diagonal izquierda y derecha respectivamente, característica local).

c.3. Cambiar la posición s de la configuración y^{r-1} muestreando de forma aleatoria X_s de la distribución Binomial de parámetros M y w . De esta manera, se obtiene una nueva configuración y^r .

d. Fijar r igual $r+1$.

f. Si $r \leq |S|$ ir a **c.1.**

g. Fijar la configuración x^n igual a la $y^{|S|}$.

6. Fijar n igual a $n+1$.

7. Ir al paso **5.**

B.3 Algoritmo de Contaminación

Versión básica del algoritmo de contaminación aleatoria del k % de los píxeles de una imagen de $m \times m$.

Algoritmo

1. Leer los elementos de la matriz X (imagen) a contaminar.

2. Leer el orden de la matriz X .

3. Construir una matriz nula \mathbf{K} del mismo orden que \mathbf{X} . (La matriz \mathbf{K} controla que no haya elementos contaminados por lo menos 2 veces).
4. Fijar el porcentaje de píxeles de la imagen \mathbf{X} a contaminar (\mathbf{t} cantidad total de píxeles contaminados).
5. Fijar $\mathbf{c} = 1$ (inicialización del número de píxeles a ser contaminados en la imagen).
6. Hacer lo siguiente:
 - a. Muestrear en forma aleatoria un sitio \mathbf{s} a ser contaminado a través de la distribución Uniforme \mathbf{G} en \mathbf{S} .
 - b. Hacer lo siguiente:
 - i. Si el elemento del sitio \mathbf{s} de la matriz \mathbf{K} es igual a cero, asignar al elemento de \mathbf{s} de la matriz \mathbf{X} el valor M y al elemento \mathbf{s} de la matriz \mathbf{K} el valor 1.
 - ii. Si el elemento de \mathbf{s} de la matriz \mathbf{K} es igual 1, fijar \mathbf{c} igual a $\mathbf{c}+1$.
 - c. Si $\mathbf{c} \leq \mathbf{t}$ ir a 6..

B.4 Algoritmo del método de estimación Mínimos de Cuadrados Condicional

B.4.1 Algoritmo de estimación de los parámetros del modelo de H-V

Versión básica del algoritmo de estimación de los parámetros del modelo de H-V por medio de método de Mínimos Cuadrados Condicional.

Algoritmo

1. Leer la textura generada \mathbf{X} (imagen).
2. Leer el tamaño de \mathbf{X} .
3. Fijar $\mathbf{t} = 1$.
4. Fijar $\mathbf{s} = 1$.
5. Hacer lo siguiente:
 - a. Calcular el elemento $(\mathbf{t}, 1)$ y el $(\mathbf{t}, 2)$ de la matriz \mathbf{G} :

$$\mathbf{g}_{t1} = \frac{(x_t + x_{t^*})}{M} \quad \text{y} \quad \mathbf{g}_{t2} = \frac{(x_u + x_{u^*})}{M}.$$

b. Calcular el elemento $(t,1)$ de la matriz **D**:

$$d_{t1} = -\log\left(\frac{M}{x_s} - 1\right), \quad \text{si } x_s > 0 \text{ y } x_s < M.$$

Caso contrario, establecer algún criterio adecuado al estudio.

c. Fijar **s** igual a **s+1**.

d. Fijar **t** igual a **t+1**.

e. Si $s \leq |S|$ ir a **5.**

6. Determinar la matriz traspuesta de **G**.

7. Calcular la matriz producto **P** entre la matriz traspuesta de **G** y **G**.

8. Hallar el determinante de la matriz **P** para chequear si es inversible.

9. En caso de ser inversible, determinar la matriz inversa de **P**.

10. Calcular la matriz **H**, producto de la matriz traspuesta de **G** y la matriz **D**.

11. Determinar el producto entre la inversa de **P** y la matriz **H**. De esta forma se obtiene la estimación de mínimos cuadrados condicional del vector de parámetros del modelo Autobinomial H-V.

B.4.2 Algoritmo de estimación de los parámetros del modelo de I-D

Versión básica del algoritmo de estimación de los parámetros del modelo de I-D por medio de método de Mínimos Cuadrados Condicional.

Algoritmo

1. Leer la textura generada **X** (imagen).

2. Leer el tamaño de **X**.

3. Fijar **t** = 1.

4. Fijar **s** = 1.

5. Hacer lo siguiente:

a. Calcular el elemento $(t,1)$ y el $(t,2)$ de la matriz **G**:

$$\mathbf{gt1} = \frac{(x_v + x_{v^*})}{M} \quad \text{y} \quad \mathbf{gt2} = \frac{(x_z + x_{z^*})}{M}.$$

b. Calcular el elemento $(\mathbf{t}, 1)$ de la matriz \mathbf{D} :

$$d_{t1} = -\log\left(\frac{M}{x_s} - 1\right), \text{ si } \mathbf{x}_s > 0 \text{ y } \mathbf{x}_s < M.$$

Caso contrario, establecer algún criterio adecuado al estudio.

c. Fijar \mathbf{s} igual a $\mathbf{s}+1$.

d. Fijar \mathbf{t} igual a $\mathbf{t}+1$.

e. Si $\mathbf{s} \leq |\mathbf{S}|$ ir a **5.**

6. Determinar la matriz traspuesta de \mathbf{G} .

7. Calcular la matriz producto, \mathbf{P} , entre la matriz traspuesta de \mathbf{G} y \mathbf{G} .

8. Hallar el determinante de la matriz \mathbf{P} para verificar si es inversible.

9. En caso de ser inversible, determinar la matriz inversa de \mathbf{P} .

10. Calcular la matriz \mathbf{H} , producto de la matriz traspuesta de \mathbf{G} y la matriz \mathbf{D} .

11. Determinar el producto entre la inversa de \mathbf{P} y la matriz \mathbf{H} . De esta forma se obtiene la estimación de mínimos cuadrados condicional del vector de parámetros del modelo Autobinomial H-V.

TABLAS DE RESULTADOS

A continuación se exhiben las estimaciones del vector de parámetros de los Modelos Autobinomiales H-V e I-D Puros y Contaminados, en las cuatro situaciones consideradas. Estas estimaciones, logradas mediante los métodos de Mínimos Cuadrados Condicional y Máximo Pseudo-verosimilitud, se efectuaron a partir de vectores de valores medios, obtenidos de 100 muestras aleatorias (texturas) de cada una de las situaciones contempladas. Se incluyen, además, sus respectivas matrices de varianza-covarianza.

C.1 Modelo Autobinomial H-V

C.1.1 Estimador MCC

Tabla C.1 Resultados de las estimaciones del vector de parámetros $\theta_{H-V} = (\beta_{11}, \beta_{12})$, del Modelo Autobinomial H-V Puro y Contaminado, en las cuatro situaciones consideradas.

Situación	Modelo	Modelos Contaminados		
	Puro	1%	10%	20%
$\theta_{H-V} = (0, 1)$	$\hat{\theta}_{H-V} = (0.0006, 1.0041)$	$\hat{\theta}_{H-V} = (0.0079, 1.0041)$	$\hat{\theta}_{H-V} = (0.0412, 1.0939)$	$\hat{\theta}_{H-V} = (0.2213, 0.9488)$
	$\hat{\Sigma} = \begin{pmatrix} 0.0008 & -0.0044 \\ -0.0044 & 0.0048 \end{pmatrix}$	$\hat{\Sigma} = \begin{pmatrix} 0.0021 & -0.0017 \\ -0.0017 & 0.0021 \end{pmatrix}$	$\hat{\Sigma} = \begin{pmatrix} 0.0027 & -0.0039 \\ -0.0039 & 0.0201 \end{pmatrix}$	$\hat{\Sigma} = \begin{pmatrix} 0.0192 & -0.0146 \\ -0.0146 & 0.0964 \end{pmatrix}$
$\theta_{H-V} = (1, 1)$	$\hat{\theta}_{H-V} = (1.0055, 1.0059)$	$\hat{\theta}_{H-V} = (0.9980, 1.0138)$	$\hat{\theta}_{H-V} = (1.0600, 1.0355)$	$\hat{\theta}_{H-V} = (1.1158, 1.1440)$
	$\hat{\Sigma} = \begin{pmatrix} 0.0006 & -0.0003 \\ -0.0003 & 0.0015 \end{pmatrix}$	$\hat{\Sigma} = \begin{pmatrix} 0.0052 & -0.0050 \\ -0.0050 & 0.0053 \end{pmatrix}$	$\hat{\Sigma} = \begin{pmatrix} 0.0077 & -0.0181 \\ -0.0181 & 0.0068 \end{pmatrix}$	$\hat{\Sigma} = \begin{pmatrix} 0.0263 & -0.0165 \\ -0.0165 & 0.0337 \end{pmatrix}$
$\theta_{H-V} = (1, -1)$	$\hat{\theta}_{H-V} = (1.0029, -1.0028)$	$\hat{\theta}_{H-V} = (0.9925, -0.9917)$	$\hat{\theta}_{H-V} = (0.9197, -0.9431)$	$\hat{\theta}_{H-V} = (0.7841, 0.2179)$
	$\hat{\Sigma} = \begin{pmatrix} 0.0002 & -0.0001 \\ -0.0001 & 0.0001 \end{pmatrix}$	$\hat{\Sigma} = \begin{pmatrix} 0.00025 & -0.00004 \\ -0.00004 & 0.00022 \end{pmatrix}$	$\hat{\Sigma} = \begin{pmatrix} 0.0007 & -0.00007 \\ -0.00007 & 0.00067 \end{pmatrix}$	$\hat{\Sigma} = \begin{pmatrix} 0.0295 & -0.0224 \\ -0.0224 & 0.0296 \end{pmatrix}$
$\theta_{H-V} = (-1, 1)$	$\hat{\theta}_{H-V} = (-1.0039, 1.0027)$	$\hat{\theta}_{H-V} = (-0.9944, 0.9866)$	$\hat{\theta}_{H-V} = (-0.9491, 0.9494)$	$\hat{\theta}_{H-V} = (-0.2289, 0.7597)$
	$\hat{\Sigma} = \begin{pmatrix} 0.0001 & -0.0001 \\ -0.0001 & 0.0002 \end{pmatrix}$	$\hat{\Sigma} = \begin{pmatrix} 0.0002 & -0.0001 \\ -0.0001 & 0.0014 \end{pmatrix}$	$\hat{\Sigma} = \begin{pmatrix} 0.00064 & -0.00010 \\ -0.00010 & 0.00062 \end{pmatrix}$	$\hat{\Sigma} = \begin{pmatrix} 0.0115 & 0.0043 \\ 0.0043 & 0.0238 \end{pmatrix}$

C.1.2 Estimador MP

Tabla C.2 Resultados de las estimaciones del vector de parámetros $\theta_{H-V} = (\beta_{11}, \beta_{12})$, del Modelo Autobinomial H-V Puro y Contaminado, en las cuatro situaciones consideradas.

Situación	Modelo	Modelos Contaminados		
	Puro	1%	10%	20%
$\theta_{H-V} = (0, 1)$	$\hat{\theta}_{H-V} = (0.0006, 1.0023)$	$\hat{\theta}_{H-V} = (0.0042, 1.0029)$	$\hat{\theta}_{H-V} = (0.0382, 1.0782)$	$\hat{\theta}_{H-V} = (0.2061, 0.9651)$
	$\hat{\Sigma} = \begin{pmatrix} 0.0004 & -0.00003 \\ 0.00003 & 0.0003 \end{pmatrix}$	$\hat{\Sigma} = \begin{pmatrix} 0.0020 & -0.0008 \\ 0.0008 & 0.0011 \end{pmatrix}$	$\hat{\Sigma} = \begin{pmatrix} 0.0018 & -0.0022 \\ 0.0022 & 0.0131 \end{pmatrix}$	$\hat{\Sigma} = \begin{pmatrix} 0.0158 & -0.0032 \\ 0.0032 & 0.0284 \end{pmatrix}$
$\theta_{H-V} = (1, 1)$	$\hat{\theta}_{H-V} = (1.0034, 1.0020)$	$\hat{\theta}_{H-V} = (1.0071, 1.0081)$	$\hat{\theta}_{H-V} = (1.0187, 1.0104)$	$\hat{\theta}_{H-V} = (1.0754, 1.1211)$
	$\hat{\Sigma} = \begin{pmatrix} 0.0002 & -0.000002 \\ 0.000002 & 0.0001 \end{pmatrix}$	$\hat{\Sigma} = \begin{pmatrix} 0.0006 & -0.0002 \\ 0.0002 & 0.0003 \end{pmatrix}$	$\hat{\Sigma} = \begin{pmatrix} 0.0077 & -0.0031 \\ 0.0031 & 0.0069 \end{pmatrix}$	$\hat{\Sigma} = \begin{pmatrix} 0.0249 & -0.0165 \\ 0.0165 & 0.0181 \end{pmatrix}$
$\theta_{H-V} = (1, -1)$	$\hat{\theta}_{H-V} = (1.0015, -1.0019)$	$\hat{\theta}_{H-V} = (0.9969, 0.9947)$	$\hat{\theta}_{H-V} = (0.9713, -0.9615)$	$\hat{\theta}_{H-V} = (0.8287, -0.2191)$
	$\hat{\Sigma} = \begin{pmatrix} 0.00011 & -0.00006 \\ 0.00006 & 0.00008 \end{pmatrix}$	$\hat{\Sigma} = \begin{pmatrix} 0.00067 & -0.00067 \\ 0.00067 & 0.00094 \end{pmatrix}$	$\hat{\Sigma} = \begin{pmatrix} 0.0014 & -0.0002 \\ 0.0002 & 0.0030 \end{pmatrix}$	$\hat{\Sigma} = \begin{pmatrix} 0.0019 & 0.0010 \\ 0.0010 & 0.0809 \end{pmatrix}$
$\theta_{H-V} = (-1, 1)$	$\hat{\theta}_{H-V} = (-1.0026, 1.0019)$	$\hat{\theta}_{H-V} = (0.9965, 0.9941)$	$\hat{\theta}_{H-V} = (-0.9849, 0.9852)$	$\hat{\theta}_{H-V} = (-0.3006, 0.8244)$
	$\hat{\Sigma} = \begin{pmatrix} 0.00011 & -0.00011 \\ 0.00011 & 0.00014 \end{pmatrix}$	$\hat{\Sigma} = \begin{pmatrix} 0.00081 & -0.00001 \\ 0.00001 & 0.00015 \end{pmatrix}$	$\hat{\Sigma} = \begin{pmatrix} 0.0039 & 0.0001 \\ 0.0001 & 0.0026 \end{pmatrix}$	$\hat{\Sigma} = \begin{pmatrix} 0.0030 & 0.0004 \\ 0.0004 & 0.0025 \end{pmatrix}$

C.2 Modelo Autobinomial I-D

C.2.1 Estimador MCC

Tabla C.3 Resultados de las estimaciones del vector de parámetros $\theta_{I-D} = (\beta_{21}, \beta_{22})$, del Modelo Autobinomial I-D Puro y Contaminado, en las cuatro situaciones consideradas.

Situación	Modelo	Modelos Contaminados		
	Puro	1%	10%	20%
$\theta_{I-D} = (0, 1)$	$\hat{\theta}_{I-D} = (-0.0017, 1.0080)$	$\hat{\theta}_{I-D} = (0.0055, 0.9994)$	$\hat{\theta}_{I-D} = (0.0428, 1.0426)$	$\hat{\theta}_{I-D} = (0.2069, 1.1020)$
	$\hat{\Sigma} = \begin{pmatrix} 0.0010 & -0.0010 \\ -0.0010 & 0.0010 \end{pmatrix}$	$\hat{\Sigma} = \begin{pmatrix} 0.0005 & -0.0002 \\ -0.0002 & 0.0015 \end{pmatrix}$	$\hat{\Sigma} = \begin{pmatrix} 0.0071 & -0.0012 \\ -0.0012 & 0.0048 \end{pmatrix}$	$\hat{\Sigma} = \begin{pmatrix} 0.0113 & -0.0002 \\ -0.0002 & 0.0298 \end{pmatrix}$
$\theta_{I-D} = (1, 1)$	$\hat{\theta}_{I-D} = (1.0036, 1.0012)$	$\hat{\theta}_{I-D} = (1.0088, 1.0044)$	$\hat{\theta}_{I-D} = (1.0957, 1.0399)$	$\hat{\theta}_{I-D} = (1.1231, 1.1141)$
	$\hat{\Sigma} = \begin{pmatrix} 0.00002 & -0.00002 \\ -0.00002 & 0.00032 \end{pmatrix}$	$\hat{\Sigma} = \begin{pmatrix} 0.0003 & -0.00001 \\ 0.00001 & 0.0002 \end{pmatrix}$	$\hat{\Sigma} = \begin{pmatrix} 0.0024 & -0.0023 \\ -0.0023 & 0.0033 \end{pmatrix}$	$\hat{\Sigma} = \begin{pmatrix} 0.0491 & -0.0492 \\ -0.0492 & 0.0493 \end{pmatrix}$
$\theta_{I-D} = (1, -1)$	$\hat{\theta}_{I-D} = (1.0031, -1.0030)$	$\hat{\theta}_{I-D} = (1.0059, -0.9939)$	$\hat{\theta}_{I-D} = (0.9669, -1.0648)$	$\hat{\theta}_{I-D} = (0.7098, 0.1491)$
	$\hat{\Sigma} = \begin{pmatrix} 0.0002 & -0.0002 \\ -0.0002 & 0.0002 \end{pmatrix}$	$\hat{\Sigma} = \begin{pmatrix} 0.0004 & -0.00001 \\ -0.00001 & 0.0005 \end{pmatrix}$	$\hat{\Sigma} = \begin{pmatrix} 0.0039 & 0.0004 \\ 0.0004 & 0.0091 \end{pmatrix}$	$\hat{\Sigma} = \begin{pmatrix} 0.0046 & -0.0040 \\ -0.0040 & 0.0343 \end{pmatrix}$
$\theta_{I-D} = (-1, 1)$	$\hat{\theta}_{I-D} = (-1.0021, 1.0023)$	$\hat{\theta}_{I-D} = (0.9935, -0.9921)$	$\hat{\theta}_{I-D} = (-1.0881, 1.0360)$	$\hat{\theta}_{I-D} = (-0.1994, 0.8542)$
	$\hat{\Sigma} = \begin{pmatrix} 0.0001 & -0.0001 \\ -0.0001 & 0.00001 \end{pmatrix}$	$\hat{\Sigma} = \begin{pmatrix} 0.0006 & -0.00002 \\ -0.00002 & 0.0002 \end{pmatrix}$	$\hat{\Sigma} = \begin{pmatrix} 0.0146 & 0.0008 \\ 0.0008 & 0.0043 \end{pmatrix}$	$\hat{\Sigma} = \begin{pmatrix} 0.0135 & 0.0096 \\ 0.0096 & 0.0263 \end{pmatrix}$

C.2.2 Estimador MP

Tabla C.4 Resultados de las estimaciones del vector de parámetros $\theta_{I-D} = (\beta_{21}, \beta_{22})$, del Modelo Autobinomial I-D Puro y Contaminado, en las cuatro situaciones consideradas.

Situación	Modelo	Modelos Contaminados		
	Puro	1%	10%	20%
$\theta_{I-D} = (0, 1)$	$\hat{\theta}_{I-D} = (-0.0013, 1.0077)$	$\hat{\theta}_{I-D} = (0.0031, 1.0015)$	$\hat{\theta}_{I-D} = (0.0363, 1.0393)$	$\hat{\theta}_{I-D} = (0.1558, 1.1000)$
	$\hat{\Sigma} = \begin{pmatrix} 0.00074 & -0.00054 \\ -0.00054 & 0.0005 \end{pmatrix}$	$\hat{\Sigma} = \begin{pmatrix} 0.00041 & -0.00010 \\ -0.00010 & 0.00018 \end{pmatrix}$	$\hat{\Sigma} = \begin{pmatrix} 0.0029 & -0.002 \\ -0.002 & 0.0077 \end{pmatrix}$	$\hat{\Sigma} = \begin{pmatrix} 0.0099 & -0.0023 \\ -0.0023 & 0.0391 \end{pmatrix}$
$\theta_{I-D} = (1, 1)$	$\hat{\theta}_{I-D} = (1.0033, 1.0011)$	$\hat{\theta}_{I-D} = (1.0074, 1.0041)$	$\hat{\theta}_{I-D} = (1.0277, 1.0262)$	$\hat{\theta}_{I-D} = (1.1040, 1.1021)$
	$\hat{\Sigma} = \begin{pmatrix} 0.00002 & -0.000006 \\ -0.000006 & 0.00007 \end{pmatrix}$	$\hat{\Sigma} = \begin{pmatrix} 0.0003 & -0.00008 \\ -0.00008 & 0.0002 \end{pmatrix}$	$\hat{\Sigma} = \begin{pmatrix} 0.0102 & -0.0066 \\ -0.0066 & 0.0218 \end{pmatrix}$	$\hat{\Sigma} = \begin{pmatrix} 0.0265 & -0.0109 \\ -0.0109 & 0.0220 \end{pmatrix}$
$\theta_{I-D} = (1, -1)$	$\hat{\theta}_{I-D} = (1.0023, -1.0022)$	$\hat{\theta}_{I-D} = (1.0059, -1.0012)$	$\hat{\theta}_{I-D} = (0.9858, -1.0519)$	$\hat{\theta}_{I-D} = (0.8123, -0.2323)$
	$\hat{\Sigma} = \begin{pmatrix} 0.00007 & -0.00004 \\ -0.00004 & 0.00010 \end{pmatrix}$	$\hat{\Sigma} = \begin{pmatrix} 0.0003 & 0.00001 \\ 0.00001 & 0.0004 \end{pmatrix}$	$\hat{\Sigma} = \begin{pmatrix} 0.0027 & 0.0001 \\ 0.0001 & 0.0067 \end{pmatrix}$	$\hat{\Sigma} = \begin{pmatrix} 0.0013 & 0.0003 \\ 0.0003 & 0.013 \end{pmatrix}$
$\theta_{I-D} = (-1, 1)$	$\hat{\theta}_{I-D} = (-1.0020, 1.0022)$	$\hat{\theta}_{I-D} = (-1.0032, 1.0027)$	$\hat{\theta}_{I-D} = (-1.0708, 1.0219)$	$\hat{\theta}_{I-D} = (-0.3359, 0.8559)$
	$\hat{\Sigma} = \begin{pmatrix} 0.00007 & -0.00006 \\ -0.00006 & 0.00006 \end{pmatrix}$	$\hat{\Sigma} = \begin{pmatrix} 0.0003 & -0.00004 \\ -0.00004 & 0.0003 \end{pmatrix}$	$\hat{\Sigma} = \begin{pmatrix} 0.0128 & -0.0006 \\ -0.0006 & 0.0042 \end{pmatrix}$	$\hat{\Sigma} = \begin{pmatrix} 0.0089 & 0.0032 \\ 0.0032 & 0.0196 \end{pmatrix}$