



**UNIVERSIDAD NACIONAL AUTÓNOMA  
DE MÉXICO**

---

---

**FACULTAD DE CIENCIAS**

**ESTADÍSTICA ESPACIAL APLICADA  
A LA PESCA EN MEXICO**

**T E S I S**

**QUE PARA OBTENER EL TÍTULO DE**

**M A T E M Á T I C A**

**P R E S E N T A**

**VELVET MALVAEZ OROZCO**



**DIRECTOR DE TESIS:  
DR. CARLOS DIAZ AVALOS**

2011



Universidad Nacional  
Autónoma de México

Dirección General de Bibliotecas de la UNAM

**Biblioteca Central**



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

# Índice general

<b>Introducción</b>	<b>III</b>
<b>1. Campos Aleatorios</b>	<b>1</b>
1.1. Campos Aleatorios de Markov . . . . .	2
1.1.1. Campos Aleatorios de Markov Gaussianos . . . . .	5
1.1.2. Autoregresiones Condicionales (CAR) . . . . .	9
1.2. Modelo Lineal Espacial . . . . .	12
<b>2. Inferencia Bayesiana</b>	<b>14</b>
2.1. Conceptos Básicos . . . . .	14
2.2. Modelos Jerárquicos Bayesianos . . . . .	16
2.3. Distribuciones Condicionales Totales . . . . .	18
2.4. Simulación Monte Carlo de Cadenas de Markov . . . . .	22
2.4.1. Algoritmo Metropolis-Hastings . . . . .	22
2.4.2. Muestreo de Gibbs . . . . .	25
2.4.3. Implementación . . . . .	26
2.4.4. Diagnósticos de Convergencia . . . . .	27
<b>3. Reconstrucción de Imágenes</b>	<b>29</b>

<i>ÍNDICE GENERAL</i>	II
3.1. El Modelo . . . . .	29
3.2. Muestreo de Gibbs en la Reconstrucción de Imágenes . . . . .	31
3.3. Resultados . . . . .	33
<b>4. La Captura</b>	<b>36</b>
4.1. El Modelo . . . . .	36
4.2. Muestreo de Gibbs en la Estimación de los Parámetros de la Captura . . .	40
4.3. Resultados . . . . .	41
<b>5. Conclusiones</b>	<b>42</b>
<b>Bibliografía</b>	<b>43</b>

# Introducción

La producción de alimentos es uno de los mayores problemas para los países en vías de desarrollo ya que no cuentan con los conocimientos tecnológicos necesarios, mano de obra especializada, inversiones e infraestructura tecnológica y también presentan problemas socioeconómicos como la falta de educación.

El problema de la escasez de alimentos se hace más grave entre estos países por el crecimiento de la población y como consecuencia al régimen alimenticio deficiente se crea el ambiente propicio para el desarrollo de enfermedades que a su vez causen inestabilidad e inseguridad.

Desde hace varios años, México ha realizado esfuerzos para resolver este problema. Uno de los programas que han recibido mayor impulso en el país es el “Programa de Desarrollo Pesquero”, ya que se considera que los recursos acuáticos pueden ser una importante fuente de alimentos, así como la posibilidad de generar empleos que permitan a la población adquirir este tipo de comida.

La República Mexicana posee 11 592.77 km de costas, de los cuales 8475.06 corresponden al litoral del Pacífico y 3 117.71 al del Golfo de México y Caribe incluyendo islas; su plataforma continental es de aproximadamente 394 603 km<sup>2</sup>, siendo mayor en el Golfo de México. Gracias al régimen establecido en 1976 de 200 millas náuticas de “zona económica exclusiva”, quedan bajo jurisdicción nacional 2 946 885 km<sup>2</sup> de región marina nacional.

Gracias al impulso que se ha dado al “Programa de Desarrollo Pesquero” la captura se incrementó notablemente; se ha obtenido una tasa anual de crecimiento importante que ubica al sector pesquero como uno de los de más rápido crecimiento en la economía nacional.

En las aguas nacionales se aprovechan 305 especies diferentes, y se ha calculado que existen alrededor de 1 200 especies posibles de ser capturadas. Las principales especies que forman la captura total mexicana son para consumo humano directo, entre éstas, el atún.

Bajo el nombre de *atunes* se incluyen diversos tipos de peces: los considerados verdaderos atunes, como el *atún aleta azul*, el *atún aleta amarilla* y la *albácora*, y otros con características similares como el *barrilete* y el *bonito del Atlántico*.

Los atunes son peces oceánicos con características morfológicas que les permiten ser buenos nadadores; su cuerpo les permite nadar largas distancias y alcanzar altas velocidades de hasta 70 km/hr. Junto con los esturiones, los atunes se encuentran entre los peces de mayor tamaño; uno de los más grandes es el atún aleta azul, que habita en el Atlántico y llega a medir 3 m de longitud y a pesar 680 kg. En los mares cálidos, donde es muy abundante, los individuos pesan de 15 a 100 kg y miden de 40 cm a 1 m de longitud, como es el caso de los bonitos y los barriletes.

El atún aleta azul y la albácora se localizan en aguas templadas, mientras que el atún aleta amarilla y el barrilete se localizan en aguas cálidas con temperaturas que van de los 17 a los 33°C. Los atunes tienen una gran sensibilidad a los cambios en temperatura, salinidad y turbidez del agua, así como a la cantidad de alimento disponible.

Los atunes se alimentan durante todas las estaciones del año excepto en el periodo de reproducción, que se lleva a cabo generalmente durante los meses de primavera y verano, y comen prácticamente todo lo que encuentran. Un ejemplar de barrilete consume 25 % de su peso de alimento.

La pesquería de atunes en México es una de las mejor establecidas; en el año 1986, México fue el 2° lugar mundial en récords de captura y en 1988 la captura anual fue de 113 607 toneladas, pescándose 113 324 en el Pacífico y sólo 283 en el Atlántico. La flota atunera nacional está integrada con 58 embarcaciones, de las cuales algunas cuentan con los sistemas más modernos para realizar la captura como sonares, navegador por satélite, localizadores de dirección, guardadores de redes, monitores de redes, indicadores de la temperatura del agua y ecosondas con video a color.

La mayor parte de la captura se realiza en la zona noroccidental del país, concentrándose ahí gran parte tanto de la flota como del personal experimentado en esta pesquería, así como la infraestructura portuaria y comercial.

Las principales especies de atunes que se capturan en México son: el aleta amarilla, que se encuentra en la costa del Pacífico oriental y en las aguas de California, el aleta azul que se encuentra en toda la costa occidental de Baja California y se concentra alrededor de las islas, y el barrilete y el bonito, que se pescan tanto en el Pacífico como en el Atlántico.

Los cardúmenes de atún se localizan visualmente ya sea por un vigía con enormes catalejos, desde avionetas que acompañan a la flota, o desde un helicóptero perteneciente

al barco. Es por esto que en este trabajo se construye una herramienta que facilite la localización de los cardúmenes de atún para así incrementar la eficiencia en su captura.

Para hacer uso de esta herramienta primero se debe valer de otra: la percepción remota, que se define simplemente como coleccionar e interpretar información acerca de un objetivo sin estar en contacto con el objeto. Las plataformas más comunes para observaciones de percepción remota son aeronaves y satélites. En este caso el objetivo es el Océano Pacífico Mexicano, ya que la mayor parte de la captura de atún se realiza en esta zona, y las imágenes son tomadas desde un satélite.

De las imágenes recolectadas por el satélite, sólo se toman los datos que nos interesan: la temperatura y la clorofila, ya que son los factores más importantes que influyen en la presencia de un cardumen de atún en una zona. Con estos datos se desea principalmente predecir el tiempo y lugar en el que ocurrirá una surgencia. Este término se utiliza para denominar los movimientos verticales ascendentes que transportan aguas de niveles subsuperficiales hasta la capa superficial del océano, así como los movimientos horizontales que las remueven de las regiones donde ocurren estos eventos. Las aguas subsuperficiales que alcanzan la capa superficial tienen mayor contenido de nutrientes y menor de oxígeno disuelto que las aguas superficiales reemplazadas. Estas aguas ricas en nutrientes que alcanzan la capa superficial en la que penetra la luz, generan un aumento acelerado del fitoplancton, que constituye la base de la cadena alimenticia en el océano; así, la biomasa de los elementos superiores de la cadena alimenticia también aumenta. Por esta razón, las zonas de surgencia son de gran importancia en la pesca.

El problema inicial que se presenta es de la ausencia de datos en las imágenes proveídas por el satélite, es decir, las imágenes no salen completas. Esta ausencia de datos puede ser producida principalmente por la presencia de una nube al momento de tomar la imagen o de alguna falla en el satélite. Es así que surge el primer objetivo: “reconstruir imágenes”, procedimiento que se verá en el capítulo 3. Ya teniendo la imagen “reconstruída”, es decir, teniendo los valores de temperatura y clorofila de cada pixel de la imagen, se procede a aproximar la captura de túnidos por pixel, contenido del capítulo 4. Todo esto se logra haciendo uso de Métodos Monte Carlo, que serán presentados en el capítulo 1.

# Capítulo 1

## Campos Aleatorios

Una series de datos espaciales pueden ser modelados mediante un modelo estocástico conocido como campo aleatorio.

**Definición 1.1.** Sea  $D$  un conjunto finito de índices (conjunto de sitios o posiciones). Para cada sitio  $s \in D$  existe un espacio  $\Lambda_s$  de estados  $x_s$ . El espacio de configuraciones  $\mathbf{x} = (x_s)_{s \in D}$  es el producto  $\Omega = \prod_{s \in D} \Lambda_s$ . Se consideran medidas de probabilidad o distribuciones  $\pi$  en  $\Omega$ ; éstas pueden ser representadas por vectores  $\pi = (\pi(\mathbf{x}))_{\mathbf{x} \in \Omega}$  tal que  $\pi(x) \geq 0$  y  $\sum_{\mathbf{x} \in \Omega} \pi(\mathbf{x}) = 1$ . Se dice que un campo  $\mathbf{X}$  definido en  $D$  con una medida de probabilidad asociada  $\pi$  tal que  $\pi(\mathbf{x}) > 0$  para toda  $\mathbf{x} \in \Omega$ , i.e., si la distribución o medida cumple la condición de positividad, se conoce como campo aleatorio.

Para  $A \subset D$  sea  $\Lambda_A = \prod_{s \in A} \Lambda_s$  el espacio de configuraciones  $\mathbf{x}_A = (x_s)_{s \in A}$ . Como notación se usará  $X_s$  para  $X_{\{s\}}$ ,  $\{\mathbf{X}_A = \mathbf{x}_A\}$  para  $\{\mathbf{x} \in \Omega : \mathbf{X}_A = \mathbf{x}_A\}$ . Las probabilidades condicionales

$$\pi(\mathbf{X}_A = \mathbf{x}_A | \mathbf{X}_{D \setminus A} = \mathbf{x}_{D \setminus A}),, \quad A \subset D, \quad \mathbf{x}_A \in \Lambda_A, \quad \mathbf{x}_{D \setminus A} \in \Lambda_{D \setminus A}$$

son llamadas *características locales* de  $\pi$ . Para hacer más sencilla la notación, se usará  $\mathbf{X}_{-A}$  en lugar de  $\mathbf{X}_{D \setminus A}$ . Estas probabilidades están bien definidas ya que  $\pi$  es estrictamente positiva. En el contexto de estadística bayesiana, a las características locales

$$\pi(X_s = x_s | \mathbf{X}_{-s} = \mathbf{x}_{-s}),$$

se les conoce como distribuciones condicionales totales.

Usualmente  $D$  se supone un subconjunto fijo de  $\mathbb{R}^2$ , en cuyo caso se pueden tener dos tipos diferentes de datos:



- Datos geoestadísticos
- Datos en malla

Para datos geoestadísticos,  $D$  es un subconjunto fijo de  $\mathbb{R}^d$  que contiene un rectángulo  $d$ -dimensional de volumen positivo;  $\mathbf{X}(s)$  es un vector aleatorio en el punto  $s \in D$ .

En el caso de datos en malla,  $D$  es una colección fija (regular o irregular) de una cantidad numerable de puntos de  $\mathbb{R}^d$ ;  $\mathbf{X}(s)$  es un vector aleatorio en el punto  $s \in D$ . A la colección de sitios  $D$  se le conoce como malla, que después se suplementará con información de vecindades. Matemáticamente hablando, los sitios  $s$  de  $D$  se convierten en vértices conectados por líneas (esta estructura de vecindad se verá en la siguiente sección).

Los datos que se analizan en este trabajo son en una malla, donde los sitios son los pixeles que conforman las imágenes de temperatura y clorofila tomadas vía satélite y los datos de la captura.

## 1.1. Campos Aleatorios de Markov

Ciertos procesos, como un campo magnético o una imagen, pueden verse como una malla de estados. A estos modelos se les conoce como campos aleatorios de Markov (MRF) y son una extensión de las cadenas de Markov, en los que se sustituye el índice del tiempo por un índice espacial.

Para poder definir un campo aleatorio de Markov se debe introducir la noción de vecindad.

**Definición 1.2.** *Una familia  $\partial = \{\partial\{s\} : s \in D\}$  de conjuntos se llama sistema de vecindades si  $s \notin \partial\{s\}$  y  $s \in \partial\{t\}$  si y solo si  $t \in \partial\{s\}$ . Los sitios  $t \in \partial\{s\}$  se llaman vecinos de  $s$ . También se podrá escribir  $s \sim t$  en lugar de  $t \in \partial\{s\}$ . Es claro que  $\sim$  es simétrica. Un subconjunto  $C$  de  $D$  es un clic si cualesquiera dos elementos diferentes de  $C$  son vecinos.*

En la Figura 1.1 se muestran vecindades de primero, segundo y cuarto grado para mallas regulares, mientras que en la Figura 1.2 se ven todos los posibles clics en una malla regular con un sistema de vecindades de primer orden.

Ya con el concepto de vecindad claro es posible definir un campo aleatorio de Markov.

**Definición 1.3.** *Un campo aleatorio  $X$  con medida de probabilidad asociada  $\pi$  es un campo aleatorio de Markov con respecto al sistema de vecindades  $\partial$  si para toda  $s \in D$ ,*

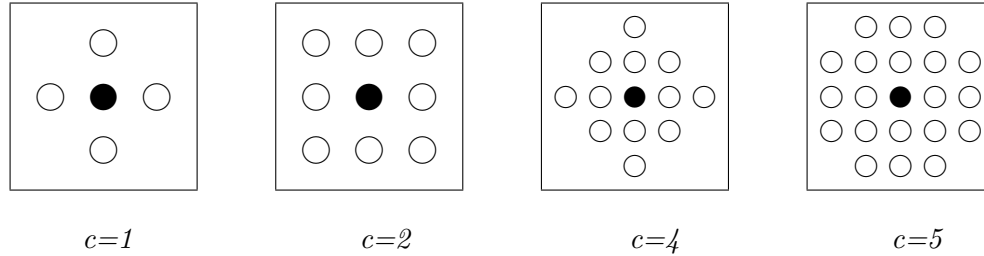


Figura 1.1: Vecindades de órdenes 1, 2, 4 y 5 en una malla regular.

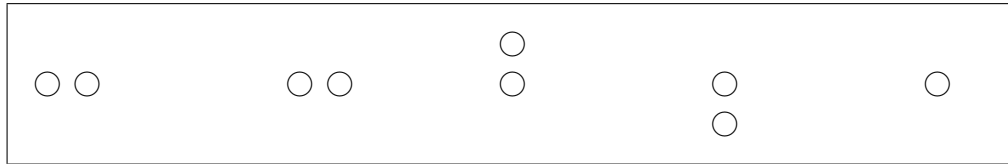


Figura 1.2: Cliques en vecindad de orden 1 en una malla regular.

$$\pi(X_s = x_s | \mathbf{X}_{-s} = \mathbf{x}_{-s}) = \pi(x_s | x_t, t \in \partial\{s\}).$$

También es posible definir un sistema de vecindades a través de una gráfica no dirigida, definida a continuación.

**Definición 1.4.** Una gráfica no dirigida  $G$  es una dupla  $G=(V,E)$ , donde  $V$  es el conjunto de nodos en la gráfica y  $E$  es el conjunto de bordes  $\{i,j\}$ , donde  $i,j \in V$  y  $i \neq j$ . En la mayoría de los casos se asume que  $V=\{1,2,\dots,n\}$ , en cuyo caso la gráfica se denomina etiquetada.

Dada  $G$ , los vecinos del nodo  $i$  son todos aquellos nodos que tienen un borde con el nodo  $i$ ,

$$\partial(i) = \{j \in V : \{i,j\} \in E\}.$$

Esta definición se puede extender a un conjunto  $A \subset V$ , donde se define a los vecinos de  $A$  como

$$\partial(A) = \bigcup_{i \in A} \partial(i) \setminus A.$$

Los vecinos de  $A$  son todos los nodos que no estan en  $A$ , pero que son adyacentes a algún nodo en  $A$ .

A continuación se da un ejemplo, mostrado por J.E. Besag, en el que la condición de positividad no se cumple.

**Ejemplo 1.1.** Considérese el modelo de infección por contacto y asóciase con cada sitio una variable  $X(\cdot)$  que es 1 si el sitio está infectado y 0 si está sano. Considérese la evolución de un sitio de sano a infectado un proceso de una sola vía y supóngase que los sitios están en una malla cuadrada de  $\mathbb{R}^2$ . Sea  $\Pr(x_s|x_{-s})$  la probabilidad condicional de observar  $x_s = x(u, v)$  en el punto  $s$  que depende sólo de los cuatro vecinos más cercanos (vecindad de primer orden) que son  $x(u-1, v)$ ,  $x(u+1, v)$ ,  $x(u, v-1)$ ,  $x(u, v+1)$ . Dada la naturaleza de infección por contacto de una vía, un evento como

$$\begin{array}{ccc} & 0 & \\ 0 & 1 & 0 \\ & 0 & \end{array} \quad (1.1)$$

debe tener probabilidad cero.

Supóngase que la infección comenzó con dos sitios consecutivos infectados, en  $(0,0)$  y  $(1,0)$ . Después de que evolucione durante un tiempo fijo, el patron de ceros y unos se toman como los datos y se analizan vía un modelo espacial. Ahora se mostrará que la condición de positividad (y por tanto también la propuesta de campo aleatorio de Markov que se dio) es inapropiada para estos datos.

Considérese la variable aleatoria  $X(u, v)$  con valores vecinos y valores extra que aparecen como sigue

$$\begin{array}{cccc} x(u-2, v+1) & 0 & x(u, v+1) & x(u+1, v+1) \\ & 0 & 1 & X(u, v) & x(u+1, v) \\ x(u-2, v-1) & 0 & x(u, v-1) & x(u+1, v-1) \end{array}$$

$\Pr\{X(u, v) = 0 | \text{valores vecinos y valores extra}\}$  debe ser cero, pues de otra manera resultaría en un evento excluído (1.2). Pero dados sólo los valores vecinos de  $X(u, v)$ ,

$$\begin{array}{ccccc}
 & & x(u, v + 1) & & \\
 & & \downarrow & & \\
 & & 1 & X(u, v) & x(u + 1, v) \\
 & & \downarrow & & \\
 & & x(u, v - 1) & & 
 \end{array}$$

$Pr\{X(u, v) | \text{valores vecinos}\}$  no será cero. Entonces, modelos de probabilidad condicional para datos de este tipo no pueden tener definida una vecindad de primer orden. La condición de positividad falla en este caso ya que cada realización individual por sitio tiene probabilidad positiva, a pesar de que la realización (1.2) tenga probabilidad cero.

### 1.1.1. Campos Aleatorios de Markov Gaussianos

Para definir un Campo Aleatorio de Markov Gaussiano (GMRF), primero se recordará la distribución normal multivariada, así como algunas propiedades, luego se definirá el concepto de independencia condicional y finalmente se definirá lo que es un GMRF.

La función de densidad de un vector aleatorio normal  $\mathbf{X} = (X_1, \dots, X_n)^T$ ,  $n < \infty$ , con media  $\mu$  (vector de  $n \times 1$ ) y matriz de covarianza  $\Sigma$  (matrix definida positiva y simétrica), es

$$\pi(\mathbf{X} = \mathbf{x}) = (2\pi)^{-n/2} |\Sigma|^{-1/2} \exp\left(-\frac{1}{2}(\mathbf{x} - \mu)^T \Sigma^{-1}(\mathbf{x} - \mu)\right), \quad \mathbf{x} \in \mathbb{R}^n.$$

Aquí,

$$\begin{aligned}
 \mu_i &= E(X_i) \\
 \Sigma_{ij} &= \text{Cov}(X_i, X_j) \\
 \Sigma_{ii} &= \text{Var}(X_i) > 0 \\
 \text{Corr}(X_i, X_j) &= \Sigma_{ij} / (\Sigma_{ii} \Sigma_{jj})^{1/2}.
 \end{aligned}$$

Esto se escribe como  $\mathbf{X} \sim N(\mu, \Sigma)$ .

**Definición 1.5.** Sean  $X$ ,  $Y$ ,  $Z$  variables aleatorias. Se dice que  $X$  y  $Y$  son **condicionalmente independientes** dado  $Z$ , si y sólo si  $\pi(X, Y | Z) = \pi(X | Z)\pi(Y | Z)$ . La notación es  $X \perp Y | Z$ .

**Teorema 1.1.** Sean  $X$ ,  $Y$ , y  $Z$  tal que  $\pi(Z) > 0$

$$X \perp Y | Z \Leftrightarrow \pi(X, Y, Z) = f(X, Z)g(Y, Z),$$

para dos funciones  $f$  y  $g$ .

**Demostración.** La prueba de la ida es trivial. Para la vuelta sabemos que

$$\pi(X, Y, Z) = \pi(X|Y, Z)\pi(Y, Z)$$

y así,

$$f(X, Z)g(Y, Z) = \pi(X|Y, Z)\pi(Y, Z).$$

Como el segundo término de la derecha depende sólo de  $Y$  y  $Z$ , entonces  $g(Y, Z) = \pi(Y, Z)$ . Así,  $f(X, Z) = \pi(X|Y, Z)$  y claramente  $\pi(X|Y, Z)$  no depende de  $Y$ , por lo que  $f(X, Z) = \pi(X|Z)$ . Por lo tanto,

$$\pi(X, Y|Z) = \frac{\pi(X, Y, Z)}{\pi(Z)} = \frac{f(X, Z)g(Y, Z)}{\pi(Z)} = \frac{\pi(X|Z)\pi(Y, Z)}{\pi(Z)} = \pi(X|Z)\pi(Y|Z)$$

lo que concluye la prueba.  $\square$

El resultado anterior puede extenderse fácilmente al caso multivariado.

Para calcular la densidad condicional de  $\mathbf{X}_A$  dado  $\mathbf{X}_{-A}$ , se usará lo siguiente

$$\pi(\mathbf{X}_A | \mathbf{X}_{-A}) = \frac{\pi(\mathbf{X}_A, \mathbf{X}_{-A})}{\pi(\mathbf{X}_{-A})} \propto \pi(\mathbf{X}). \quad (1.2)$$

Esto es cierto dado que el denominador no depende de  $\mathbf{X}_A$ .

Sea  $\mathbf{X} = (X_1, \dots, X_n)^T$  tal que  $\mathbf{X} \sim N(\mu, \Sigma)$ . Defínase la gráfica  $G = (V, E)$ , donde  $V = \{1, \dots, n\}$  y  $E$  es tal que  $\{i, j\} \notin E$  si y sólo si  $X_i \perp X_j | \mathbf{X}_{-ij}$ , donde  $\mathbf{X}_{-ij} = \mathbf{X}_{-\{i, j\}}$ . Entonces se dice que  $\mathbf{X}$  es un GMRF con respecto a la gráfica  $G$ .

El siguiente teorema muestra la relación entre los parámetros de la distribución normal y la gráfica  $G$ . Se usará, en lugar de la matriz de covarianza  $\Sigma$ , la matriz de precisión  $\mathbf{Q} = \Sigma^{-1}$ , así

$$\pi(\mathbf{x}) = (2\pi)^{-n/2} |\mathbf{Q}|^{1/2} \exp\left(-\frac{1}{2}(\mathbf{x} - \mu)^T \mathbf{Q}(\mathbf{x} - \mu)\right). \quad (1.3)$$

**Teorema 1.2.** Sea  $\mathbf{X}$  vector aleatorio con distribución normal con media  $\mu$  y matriz de precisión  $\mathbf{Q} > 0$ . Entonces para  $i \neq j$ ,

$$X_i \perp X_j | \mathbf{X}_{-ij} \Leftrightarrow Q_{ij} = 0.$$

**Demostración.** Se parte  $\mathbf{X}$  como  $(X_i, X_j, \mathbf{X}_{-ij})$ . Se fija  $i \neq j$  y se asume  $\mu = 0$  sin pérdida de la generalidad. De (2.2) se tiene

$$\begin{aligned} \pi(x_i, x_j, \mathbf{x}_{-ij}) &\propto \exp\left(-\frac{1}{2} \sum_{k,l} x_k Q_{kl} x_l\right) \\ &\propto \exp\left(-\frac{1}{2} x_i x_j (Q_{ij} + Q_{ji}) - \frac{1}{2} \sum_{\{k,l\} \neq \{i,j\}} x_k Q_{kl} x_l\right). \end{aligned}$$

El segundo término no depende de  $x_i x_j$  mientras que el primero depende de  $x_i x_j$  si y sólo si  $Q_{ij} \neq 0$ . Entonces

$$\pi(x_i, x_j, \mathbf{x}_{-ij}) = f(x_i, \mathbf{x}_{-ij}) g(x_j, \mathbf{x}_{-ij}),$$

para funciones positivas  $f$  y  $g$ , si y sólo si  $Q_{ij} = 0$ . Se sigue de la versión multivariada del Teorema 1 aplicada a  $\pi(x_i, x_j, \mathbf{x}_{-ij})$  que  $X_i \perp X_j \mid \mathbf{X}_{-ij}$  si y sólo si  $Q_{ij} = 0$ .

□

**Definición 1.6.** (GMRF) Un vector aleatorio  $\mathbf{X} = (X_1, \dots, X_n) \in \mathbb{R}^n$  se llama GMRF con respecto a una gráfica  $G = (V, E)$  con media  $\mu$  y matriz de precisión  $\mathbf{Q} = \Sigma^{-1}$ ,  $\mathbf{Q} > 0$ , si y sólo si su función de densidad es de la forma

$$\pi(\mathbf{x}) = (2\pi)^{-n/2} |\mathbf{Q}|^{1/2} \exp\left(-\frac{1}{2} (\mathbf{x} - \mu)^T \mathbf{Q} (\mathbf{x} - \mu)\right)$$

y

$$Q_{ij} \neq 0 \Leftrightarrow \{i, j\} \in E, \quad \forall i \neq j.$$

Con el siguiente resultado se podrá ver que los elementos de la matriz de precisión  $\mathbf{Q}$  tiene interpretaciones condicionales bonitas.

**Teorema 1.3.** Sea  $\mathbf{X}$  un GMRF con respecto a  $G = (V, E)$  con media  $\mu$  y matriz de precisión  $\mathbf{Q} > 0$ , entonces

$$E(X_i \mid \mathbf{X}_{-i}) = \mu_i - \frac{1}{Q_{ii}} \sum_{j:j \sim i} Q_{ij} (X_j - \mu_j) \quad (1.4)$$

$$\text{Prec}(X_i \mid \mathbf{X}_{-i}) = Q_{ii} \quad (1.5)$$

$$\text{Corr}(X_i, X_j \mid \mathbf{X}_{-ij}) = -\frac{Q_{ij}}{\sqrt{Q_{ii} Q_{jj}}}, \quad i \neq j. \quad (1.6)$$

Los elementos de la diagonal de  $\mathbf{Q}$  son las precisiones condicionales de  $X_i$  dado  $\mathbf{X}_{-i}$ , mientras que los elementos fuera de la diagonal, pueden proporcionar información acerca de la correlación condicional entre  $X_i$  y  $X_j$ , dado  $\mathbf{X}_{-ij}$ . Estos resultados deben ser comparados con los elementos de la matriz de covarianza  $\mathbf{\Sigma} = (\Sigma_{ij})$ .

**Demostración.** [Teorema 1.3] Primero recuérdese que una variable aleatoria normal univariada  $X_i$  con media  $\gamma$  y precisión  $\kappa$  tiene densidad proporcional a

$$\exp\left(-\frac{1}{2}\kappa X_i^2 + \kappa X_i \gamma\right). \quad (1.7)$$

Asúmase por el momento que  $\mu = \mathbf{0}$  y aplíquese (1.2) a (1.3), entonces se obtiene

$$\begin{aligned} \pi(X_i | \mathbf{X}_{-i}) &\propto \exp\left(-\frac{1}{2}\mathbf{X}^T \mathbf{Q} \mathbf{X}\right) \\ &\propto \exp\left(-\frac{1}{2}X_i^2 Q_{ii} - X_i \sum_{j:j \sim i} Q_{ij} X_j\right) \end{aligned} \quad (1.8)$$

Comparando (1.7) y (1.8) se puede observar que  $\pi(X_i | \mathbf{X}_{-i})$  es normal. Comparando los coeficientes para el término cuadrático se obtiene (2.4). Comparando los coeficientes para el término lineal, se tiene

$$E(X_i | \mathbf{X}_{-i}) = -\frac{1}{Q_{ii}} \sum_{j:j \sim i} Q_{ij} X_j.$$

Si  $\mathbf{X}$  tiene media  $\mu$ , entonces  $\mathbf{X} - \mu$  tiene media cero, así reemplazando  $X_i$  y  $X_j$  por  $X_i - \mu_i$  y  $X_j - \mu_j$ , respectivamente, da (1.4). Para mostrar (1.6), se procede de forma similar y se considera

$$\pi(X_i, X_j | \mathbf{X}_{-ij}) \propto \exp\left(-\frac{1}{2}(X_i, X_j) \begin{pmatrix} Q_{ii} & Q_{ij} \\ Q_{ji} & Q_{jj} \end{pmatrix} \begin{pmatrix} X_i \\ X_j \end{pmatrix} + \text{terminos lineales}\right). \quad (1.9)$$

Se compara esta densidad con la densidad de una normal bivariada  $(X_i, X_j)^T$  con matriz de covarianza  $\mathbf{\Sigma} = (\Sigma_{ij})$ , cuya densidad es proporcional a

$$\exp\left(-\frac{1}{2}(X_i, X_j) \begin{pmatrix} \Sigma_{ii} & \Sigma_{ij} \\ \Sigma_{ji} & \Sigma_{jj} \end{pmatrix}^{-1} \begin{pmatrix} X_i \\ X_j \end{pmatrix} + \text{terminos lineales}\right). \quad (1.10)$$

Comparando (1.9) y (1.10), se obtiene

$$\begin{pmatrix} Q_{ii} & Q_{ij} \\ Q_{ji} & Q_{jj} \end{pmatrix}^{-1} = \begin{pmatrix} \Sigma_{ii} & \Sigma_{ij} \\ \Sigma_{ji} & \Sigma_{jj} \end{pmatrix},$$

lo que implica que

$$\begin{aligned} \Sigma_{ii} &= Q_{jj}/\Delta \\ \Sigma_{jj} &= Q_{ii}/\Delta \\ \Sigma_{ij} &= -Q_{ij}/\Delta \end{aligned}$$

donde  $\Delta = Q_{ii}Q_{jj} - Q_{ij}^2$ .

Usando estas expresiones y la definición de correlación condicional se tiene que

$$\begin{aligned} \text{Corr}(X_i, X_j \mid \mathbf{X}_{-ij}) &= -\frac{Q_{ij}/\Delta}{\sqrt{(Q_{jj}/\Delta)(Q_{ii}/\Delta)}} \\ &= -\frac{Q_{ij}}{\sqrt{Q_{ii}Q_{jj}}}. \end{aligned}$$

□

### 1.1.2. Autoregresiones Condicionales (CAR)

Una alternativa para especificar un GMRF por medio de su media y matriz de precisión, es especificarlo implícitamente a través de sus condicionales totales  $\pi(X_i \mid \mathbf{X}_{-i})$ . El pionero de este enfoque fue Besag (1974,1975) y estos modelos son conocidos con el nombre de *autoregresiones condicionales* y por su abreviatura en inglés CAR. Ahora se mostrará qué condiciones deben cumplir las condicionales totales para que correspondan a un GMRF válido.

Supóngase que las condicionales totales son normales con

$$E(X_i \mid \mathbf{X}_{-i}) = \mu_i - \sum_{j:j \sim i} \beta_{ij}(X_j - \mu_j) \quad (1.11)$$

$$\text{Prec}(X_i \mid \mathbf{X}_{-i}) = \kappa_i > 0, \quad (1.12)$$

para  $i = 1, \dots, n$ , para algunos  $\{\beta_{ij}, i \neq j\}$  y vectores  $\mu$  y  $\kappa$ . Claramente  $\sim$  está definida implícitamente por los términos distintos de cero de  $\{\beta_{ij}\}$ . Estas condicionales totales deben ser consistentes de tal manera que exista una densidad conjunta  $\pi(\mathbf{X})$  que dará lugar a estas condicionales totales. Dado que  $\sim$  es simétrica, se sigue que si  $\beta_{ij} \neq 0$  entonces



$\beta_{ji} \neq 0$ . Comparando término a término con (2.3) y (2.4), se puede ver que si se escogen las entradas de la matriz de precisión  $\mathbf{Q}$  como

$$Q_{ii} = \kappa_i \text{ y } Q_{ij} = \kappa_i \beta_{ij}$$

y se requiere también que  $\mathbf{Q}$  sea simétrica, i.e.,

$$\kappa_i \beta_{ij} = \kappa_j \beta_{ji},$$

entonces se tiene un candidato para la densidad conjunta dando las condicionales totales. El siguiente resultado muestra que este candidato es único.

**Teorema 1.4.** *Dadas las  $n$  condicionales totales con media y precisión condicional definidas en (1.11) y (1.12), entonces  $\mathbf{X}$  es un GMRP con respecto a una gráfica  $G=(V,E)$  con media  $\mu$  y matriz de precisión  $\mathbf{Q} = (Q_{ij})$ , donde*

$$Q_{ij} = \begin{cases} \kappa_i \beta_{ij} & i \neq j \\ \kappa_i & i = j \end{cases}$$

dadas  $\kappa_i \beta_{ij} = \kappa_j \beta_{ji}$ ,  $i \neq j$ ,  $\mathbf{Q} > 0$

Para probarlo se necesita primero el siguiente lema.

**Lema 1.1. (Lema de Brooke)** *Sea  $\pi(\mathbf{x})$  la densidad para  $\mathbf{x} \in \mathbb{R}^n$  y defínase  $\Omega = \{\mathbf{x} \in \mathbb{R}^n : \pi(\mathbf{x}) > 0\}$ . Sean  $\mathbf{x}, \mathbf{y} \in \Omega$ , entonces*

$$\frac{\pi(\mathbf{x})}{\pi(\mathbf{y})} = \prod_{i=1}^n \frac{\pi(x_i | x_1, \dots, x_{i-1}, y_{i+1}, \dots, y_n)}{\pi(y_i | x_1, \dots, x_{i-1}, y_{i+1}, \dots, y_n)} \quad (1.13)$$

$$\frac{\pi(\mathbf{x})}{\pi(\mathbf{y})} = \prod_{i=1}^n \frac{\pi(x_i | y_1, \dots, y_{i-1}, x_{i+1}, \dots, x_n)}{\pi(y_i | y_1, \dots, y_{i-1}, x_{i+1}, \dots, x_n)} \quad (1.14)$$

**Demstración..** Se empieza con la identidad

$$\frac{\pi(x_n | x_1, \dots, x_{n-1}) \pi(x_1, \dots, x_{n-1})}{\pi(y_n | x_1, \dots, x_{n-1}) \pi(x_1, \dots, x_{n-1})} = \frac{\pi(x_1, \dots, x_{n-1}, x_n)}{\pi(x_1, \dots, x_{n-1}, y_n)},$$

de donde se sigue que

$$\pi(x_1, \dots, x_n) = \frac{\pi(x_n | x_1, \dots, x_{n-1})}{\pi(y_n | x_1, \dots, x_{n-1})} \pi(x_1, \dots, x_{n-1}, y_n).$$

Si se expresa el último término de forma similar se obtiene

$$\begin{aligned}\pi(x_1, \dots, x_n) &= \frac{\pi(x_n|x_1, \dots, x_{n-1})}{\pi(y_n|x_1, \dots, x_{n-1})} \\ &\times \frac{\pi(x_{n-1}|x_1, \dots, x_{n-2}, y_n)}{\pi(y_{n-1}|x_1, \dots, x_{n-2}, y_n)} \\ &\times \pi(x_1, \dots, x_{n-2}, y_{n-1}, y_n).\end{aligned}$$

Repitiendo este proceso se llega a (1.13). La alternativa (1.14) se prueba de forma similar empezando con

$$\pi(x_1, \dots, x_n) = \frac{x_1|x_2, \dots, x_n}{\pi(y_1|x_2, \dots, x_n)}\pi(y_1, x_2, \dots, x_n).$$

□

**Demostración.** [Teorema 1.4] Asíumase  $\mu = \mathbf{0}$  y fíjese  $\mathbf{y} = \mathbf{0}$ . Entonces (1.13) se simplifica a

$$\log \frac{\pi(\mathbf{x})}{\pi(\mathbf{0})} = -\frac{1}{2} \sum_{i=1}^n \kappa_i x_i^2 - \sum_{i=2}^n \sum_{j=1}^{i-1} \kappa_i \beta_{ij} x_i x_j \quad (1.15)$$

Usando (1.14) se obtiene

$$\log \frac{\pi(\mathbf{x})}{\pi(\mathbf{0})} = -\frac{1}{2} \sum_{i=1}^n \kappa_i x_i^2 - \sum_{i=1}^{n-1} \sum_{j=i+1}^n \kappa_i \beta_{ij} x_i x_j. \quad (1.16)$$

Como (1.15) y (1.16) deben ser iguales, se sigue que  $\kappa_i \beta_{ij} = \kappa_j \beta_{ji}$  para  $i \neq j$ . La densidad de  $\mathbf{x}$  se puede entonces expresar como

$$\log(\mathbf{x}) = \text{const} - \frac{1}{2} \sum_{i=1}^n \kappa_i x_i^2 - \frac{1}{2} \sum_{i \neq j} \kappa_i \beta_{ij} x_i x_j,$$

de donde se puede ver que  $\mathbf{X}$  se distribuye normal multivariada con media cero y matriz de precisión  $\mathbf{Q} > 0$ , donde  $Q_{ij} = \kappa_i \beta_{ij}$  para  $i \neq j$  y  $Q_{ii} = \kappa_i$ .

□

## 1.2. Modelo Lineal Espacial

Un modelo estadístico es una representación matemática de un mecanismo de generación de datos. Es una abstracción de un proceso físico, biológico, químico, etc., que genera información enfatizando aquellos aspectos del proceso importantes para el análisis e ignorando o dando menos peso a los aspectos inconsecuentes. Los modelos estadísticos más generales son una descomposición de una variable aleatoria de respuesta en una estructura matemática que describe la media y una estructura estocástica aditiva que describe la variación o covariación entre las respuestas. Esta simple descomposición se expresa frecuentemente como

$$\text{Datos} = \text{Estructura} + \text{Error}$$

Como ya se vio al inicio de este capítulo, los datos espaciales pueden verse como una realización de un proceso estocástico llamado campo aleatorio.

$$Z(s), \quad s \in D \subset \mathbb{R}^d$$

donde  $s$  es un sitio en  $D$ ; la dimensión del espacio,  $d$ , es generalmente 1, 2 ó 3. Para cada  $s \in D$ ,  $Z(s)$  es una variable aleatoria, que será modelada para poder hacer un análisis de los datos espaciales. Empezamos por asumir que para cada  $s \in D$  existen  $E[Z(s)]$  y  $Var[Z(s)]$ .

Se sigue entonces que se puede aplicar la descomposición anterior a campos aleatorios, de donde resulta la forma general

$$Z(s) = \mu(s) + e(s) \tag{1.17}$$

donde  $\mu(s)$  es la función de la media de  $Z(s)$ , i.e.,

$$E[Z(s)] = \mu(s)$$

y  $e(s)$  es un proceso estocástico de error donde

$$E[e(s)] = 0.$$

Para modelar  $Z(s)$  debemos modelar  $\mu(s)$  y  $e(s)$ . Se empieza por asumir una estructura lineal para  $\mu(s)$ ; en Geoestadística a este se le conoce como el modelo *Kriging Universal*. En particular asúmase que se tienen  $p$  funciones conocidas de  $s$ ,  $x_1(s), x_2(s), \dots, x_p(s)$ , de tal forma que la función de la media cumple

$$\mu(s) = \sum_{j=1}^p \beta_j x_j(s)$$

para parámetros desconocidos  $\beta_1, \dots, \beta_p$ .

En la modelación del proceso de error, se centrará en gran parte el interés en sus propiedades de segundo orden, es decir, en la covarianza entre errores en distintos sitios  $Cov(e(s), e(r))$   $s, r \in D$ . La función de covarianza es

$$\sigma(s, r) = Cov(e(s), e(r)).$$

Nótese también que

$$\sigma(s, r) = Cov(y(s), y(r)). \quad (1.18)$$

Si nuestro espacio consta de  $n$  sitios (ordenados), el modelo para nuestro campo aleatorio será

$$\mathbf{Z} = \mathbf{X}\beta + \mathbf{e}. \quad (1.19)$$

donde  $\mathbf{X}$  es una matriz ( $n \times p$ ) donde el renglón  $i$  es el vector de funciones  $x_1(i), \dots, x_p(i)$  para el sitio  $i$ ,  $\beta$  es un vector de parámetros desconocidos de dimensión  $p$  y  $\mathbf{e}$  es el vector de errores.

En este caso, la estructura espacial que se le dará al vector de errores es la de un GMRF( $\mathbf{0}, \mathbf{Q}$ ). Es por (1.18) que se obtiene entonces que  $\mathbf{Z}$  es un GMRF con media  $\mathbf{X}\beta$  y matriz de precisión  $\mathbf{Q}$ .

Escribiendo a  $\mathbf{Z}$  como un CAR resulta

$$\begin{aligned} \mathbb{E}(Z_i | Z_{-i}) &= \mu_i - \sum_{j:j \in \partial(i)} C_{ij}(Z_j - \mu_j) \\ Prec(Z_i | Z_{-i}) &= \kappa_i > 0 \end{aligned}$$

donde  $\mu_i = \sum_{j=0}^p x_j(i)\beta_j$  para  $i = 1, \dots, n$ .

## Capítulo 2

# Inferencia Bayesiana

Este trabajo hace uso de muchas herramientas de estadística bayesiana, por lo que es necesario definir desde los conceptos más básicos hasta la herramienta más importante que se usa en este trabajo, los métodos Monte Carlo de cadenas de Markov.

### 2.1. Conceptos Básicos

Desde un enfoque bayesiano, en un modelo estadístico no hay distinción entre las variables observables y los parámetros: todos son considerados cantidades aleatorias. Sea  $y$  el vector de datos observados y  $\theta$  el vector de parámetros del modelo y los datos faltantes. La inferencia Bayesiana está compuesta por 2 elementos: la distribución de muestra  $f(y|\theta)$  y la distribución  $p(\theta)$ . Esta última distribución puede ser definida con la ayuda de constantes. Para distinguir entre el parámetro de interés  $\theta$  y estas constantes se les da el nombre de hiperparámetros, ya que son los parámetros de la distribución de los parámetros. Inicialmente los hiperparámetros se suponen conocidos.

Viendo al primer componente de la inferencia Bayesiana como una función de  $\theta$  se tiene la función de verosimilitud de  $\theta$ ,  $l(\theta) = f(y|\theta)$ . El segundo componente es llamado distribución a priori, ya que contiene la distribución de probabilidad de  $\theta$  **antes** de observar el valor de  $y$ . Es natural que la inferencia esté basada en la distribución de probabilidad de  $\theta$  **después** de haber observado el valor de  $y$ . Esta distribución se conoce como distribución posterior.

La distribución posterior se puede obtener a través del teorema de Bayes, atribuido a

Thomas Bayes (1763) y lo que resulta es:

$$p(\theta|y) = \frac{f(y|\theta)p(\theta)}{f(y)} = \frac{l(\theta)p(\theta)}{f(y)}, \quad (2.1)$$

donde

$$f(y) = \int f(y|\theta)p(\theta)d\theta.$$

Dado que el lado izquierdo es una densidad para  $\theta$ , la observación  $y$  es simplemente una constante, así como  $f(y)$ . De ahora en adelante la distribución posterior se denotará  $\pi(\theta)$ , dejando implícito el valor de la observación por el cual fue condicionado. De manera más compacta se tiene

$$\pi(\theta) \propto l(\theta)p(\theta). \quad (2.2)$$

**Ejemplo 2.1.** Supóngase que se tiene una observación de una distribución normal  $Y \sim N(\theta, \sigma^2)$  con  $\sigma^2$  conocida, de tal forma que la función de verosimilitud sea  $l(\theta) = f(y|\theta) = N(y|\theta, \sigma^2) \equiv \frac{1}{\sigma\sqrt{2\pi}}\exp(-\frac{(y-\theta)^2}{2\sigma^2})$ ,  $y \in \mathbb{R}$ ,  $\theta \in \mathbb{R}$  y  $\sigma > 0$ . Si se especifica la distribución a priori como  $p(\theta) = N(\theta|\mu, \tau^2)$  con  $\lambda = (\mu, \tau^2)$  fija, entonces de (2.1) se puede calcular la distribución posterior como

$$\begin{aligned} p(\theta|y) &= \frac{N(y|\theta, \sigma^2)N(\theta|\mu, \tau^2)}{f(y)} \\ &\propto N(y|\theta, \sigma^2)N(\theta|\mu, \tau^2) \\ &= N(\theta|\frac{\sigma^2}{\sigma^2 + \tau^2}\mu + \frac{\tau^2}{\sigma^2 + \tau^2}y, \frac{\sigma^2\tau^2}{\sigma^2 + \tau^2}). \end{aligned} \quad (2.3)$$

La proporcionalidad en la segunda línea surge de (2.2). La última igualdad resulta de juntar todos los términos de  $\theta$  y  $\theta^2$  en el exponencial y de luego completar el cuadrado. Por lo tanto, la distribución posterior de  $\theta$  dado  $y$  también es normal con media  $B\mu + (1 - B)y$ , donde  $B = \sigma^2/(\sigma^2 + \tau^2)$ .

Nótese que la media posterior  $E(\theta|y)$  es un promedio ponderado de la media a priori  $\mu$  y el valor del  $y$  con pesos que son inversamente proporcionales a las varianzas correspondientes; por esta razón,  $B$  es algunas veces llamado *factor de encogimiento* ya que da la proporción de la distancia en que la media posterior se ha «encogido» desde  $y$  hacia  $\mu$ . También, la precisión es igual a  $1/\sigma^2 + 1/\tau^2$ , que es la suma de las precisiones de la función de verosimilitud y de la distribución a priori. Así, viendo a la precisión como «información», se ve que en el modelo normal/normal la información en la posterior es el total de la información en la distribución a priori y en la verosimilitud.

## 2.2. Modelos Jerárquicos Bayesianos

El término *modelos jerárquicos* se refiere más a un conjunto general de principios de modelación que a una familia específica de modelos.

A los modelos Bayesianos se les puede dar una estructura jerárquica. La distribución a priori  $p(\theta) = f(\theta|a)$  de los parámetros del modelo con parámetros a priori  $a$  puede ser considerada como un nivel de la jerarquía, con la función de verosimilitud como el escalón final de un modelo Bayesiano resultando en una distribución posterior  $f(\theta|x) \propto l(\theta)f(\theta; a)$ . En la Figura 2.1 se da una representación gráfica de la estructura jerárquica de un típico modelo Bayesiano.

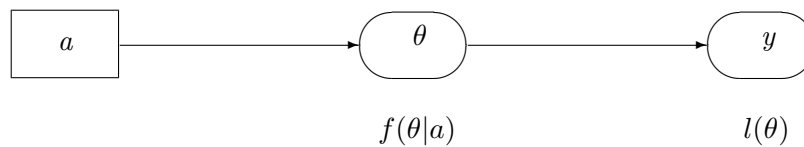


Figura 2.1: Representación gráfica de un modelo Bayesiano estándar. Los nodos cuadrados se refieren a parámetros constantes, nodos ovalados se refieren a componente estocásticos del modelo.

Para capturar la estructura complicada de algunos datos, la distribución a priori es frecuentemente estructurada haciendo uso de una serie de distribuciones condicionales llamadas *niveles jerárquicos* de la distribución a priori. Entonces, un modelo jerárquico Bayesiano se define cuando se asigna una distribución a priori para los parámetros  $a$  asociados a los parámetros  $\theta$  de la función de verosimilitud. Así, la distribución posterior puede escribirse como

$$\begin{aligned} f(\theta|x) &\propto f(x|\theta)f(\theta; a)f(a; b) \\ &\propto l(\theta)f(\theta|a)f(a|b). \end{aligned}$$

La distribución a priori en este modelo se caracteriza por dos niveles de jerarquía:  $f(\theta|a)$  como primer nivel y  $f(a|b)$  como segundo nivel. La estructura puede extenderse añadiendo más niveles de jerarquía si es necesario. En la Figura 2.2 se muestra gráficamente el modelo jerárquico de dos niveles descrito anteriormente.

**Ejemplo 2.2.** Considérense observaciones  $y_{ij} \sim N(\beta_i, \sigma^2), j = 1, \dots, n_i, i = 1, \dots, d$ , colectadas de  $d$  grupos con distintas medias  $\beta_i$  pero la misma dispersión. Este modelo es un caso



Figura 2.2: Representación gráfica de un modelo jerárquico Bayesiano de dos niveles. Los nodos cuadrados se refieren a parámetros constantes, nodos ovalados se refieren a componentes estocásticos del modelo.

especial de un modelo de regresión con un vector de observaciones  $y = (y_{11}, \dots, y_{1n_1}, \dots, y_{d1}, \dots, y_{dn_d})'$  y una matriz de diseño  $X = \text{diag}(1_{n_1}, \dots, 1_{n_d})$ , donde  $1_m$  es el vector de dimensión  $m$  cuyas entradas son unos. El modelo queda completo con una distribución a priori para  $(\beta, \sigma^2)$ . Una posibilidad es asumir dependencia a priori entre las medias  $\beta_i$ ,  $i = 1, \dots, d$ . Si los  $d$  grupos son similares en algún sentido, una alternativa válida es asumir que las medias son una muestra de una población de medias. Asumiendo una población normal,  $\beta_1, \dots, \beta_d$  son una muestra de una  $N(\mu, \tau^2)$ , donde  $\mu$  es la media y  $\tau^2$  mide la dispersión de la población de medias. Se agrega al modelo una distribución a priori para  $(\mu, \tau^2)$  y así la especificación completa a priori es

$$\begin{aligned} 1er. \text{ nivel} & : \beta | \mu, \tau^2 \sim N(1_d \mu, \tau^2 Id) \\ 2o. \text{ nivel} & : \mu \sim N(b_0, B_0) \\ & \sigma^2 \sim F_\sigma \quad y \quad \tau^2 \sim F_\tau \end{aligned}$$

para distribuciones  $F_\sigma$  y  $F_\tau$  independientes. La densidad a priori para los parámetros del modelo es

$$p(\beta, \mu, \sigma^2, \tau^2) = p(\beta | \mu, \tau^2) p(\mu) p(\sigma^2) p(\tau^2).$$

Finalmente, el modelo completo se escribe como sigue:

$$\begin{aligned} y | \beta, \sigma^2 & \sim N(X\beta, \sigma^2 I_n) \\ \beta | \mu, \tau^2 & \sim N(1_d \mu, \tau^2 Id) \\ \mu & \sim N(b_0, B_0) \\ \sigma^2 & \sim F_\sigma \\ \tau^2 & \sim F_\tau \end{aligned} \tag{2.4}$$



Generalmente, cualquier modelo Bayesiano con parámetros  $\theta$  y  $\phi$ , con distribución a priori  $f(\theta, \phi)$  puede ser descrito con una estructura jerárquica si la distribución a priori conjunta se descompone en una serie de distribuciones condicionales, i.e.,  $f(\theta, \phi) = f(\theta|\phi)f(\phi)$ . En modelos jerárquicos, los hiperparámetros  $\phi$  rara vez están involucrados en la función de verosimilitud del modelo.

En general, los modelos jerárquicos son más flexibles que los típicos modelos no jerárquicos ya que se le puede dar una estructura más complicada al modelo. Por esta razón, estos modelos describen mejor los datos, especialmente cuando el tamaño de la muestra es grande.

Una característica importante de los modelos jerárquicos es que cada parámetro referente a un grupo específico o unidad toma fuerza de los parámetros correspondientes de otros grupos con características similares. En otras palabras, se presenta un efecto de contracción o encogimiento hacia la media poblacional con el uso de un modelo jerárquico.

Robert (2007, sec. 10.2.2) provee una serie de justificaciones y ventajas en el uso de modelos jerárquicos, incluyendo el hecho de que la estructura jerárquica nos lleva a un análisis más fuerte, lo que reduce el subjetivismo ya que los resultados posteriores son promediados a través de distintas opciones a priori de los parámetros de interés.

Finalmente, esta estructura simplifica tanto la interpretación como los cálculos del modelo dado que la distribución posterior correspondiente se simplifica, resultando en distribuciones condicionales con una forma más simple. Esto permite la implementación de esquemas de muestreo de Gibbs mucho más simples, que serán nuestra herramienta más importante en este trabajo (véase sección 1.3).

### 2.3. Distribuciones Condicionales Totales

Sean  $\mathbf{y}$  el vector de observaciones y  $\theta$  el vector de parámetros del modelo. La función de densidad condicional total a priori de la componente  $i$  de  $\theta$  se define como:

$$p(\theta_i|\theta_{-i}),$$

mientras que la densidad condicional total posterior de  $\theta_i$  es:

$$p(\theta_i|\theta_{-i}, \mathbf{y}).$$

La función de densidad condicional total de  $\theta_i$  puede ser obtenida de la función de densidad conjunta de las observaciones y los parámetros,  $l(\theta)p(\theta)$ , tomando sólo los términos

que contienen a  $\theta_i$  y tratando a todos los demás como constantes. (George et al., 1993; Gilks, 1996).

**Ejemplo 2.3.** Se tienen datos de los pesos ganados por  $N$  mujeres embarazadas en 5 a 7 visitas a un consultorio de ginecología. Se propone el siguiente modelo jerárquico de regresión en el tiempo:

$$\begin{aligned} y_{ij} | \alpha_i, \beta_i, \phi &\sim N(\alpha_i + \beta_i t_{ij}, \phi^{-1}), \\ (\alpha_i, \beta_i)^T | \alpha, \beta &\sim N((\alpha, \beta)^T, \text{diag}(\tau_\alpha^{-1}, \tau_\beta^{-1})), \\ (\alpha, \beta)^T &\sim N((0, 0)^T, \text{diag}(P_\alpha^{-1}, P_\beta^{-1})), \end{aligned}$$

donde  $\phi$ ,  $\tau_\alpha$ ,  $\tau_\beta$  son parámetros independientes a priori y tienen una distribución  $G(a, b)$  y  $y_{ij}$  y  $t_{ij}$  son el  $j$ -ésimo pesaje y tiempo de visita respectivamente de la  $i$ -ésima mujer,  $j = 1, \dots, n_i$ ,  $i = 1, \dots, N$ . Aquí,  $n = \sum_{i=1}^N n_i$ ,  $P_\alpha = P_\beta = 1/1000$  y  $a = b = 0.001$ .

La función de densidad conjunta de todos los parámetros y de las observaciones es la siguiente:

$$\begin{aligned} P(\mathbf{y}, \alpha, \beta, \alpha, \beta, \phi, \tau_\alpha, \tau_\beta) &\propto \phi^{\frac{n}{2}} \prod_{i=1}^N \prod_{j=1}^{n_i} \exp \left\{ -\frac{\phi}{2} (y_{ij} - \alpha_i - \beta_i t_{ij})^2 \right\} \\ &\times \tau_\alpha^{\frac{N}{2}} \prod_{i=1}^N \exp \left\{ -\frac{\tau_\alpha}{2} (\alpha_i - \alpha)^2 \right\} \tau_\beta^{\frac{N}{2}} \prod_{i=1}^N \exp \left\{ -\frac{\tau_\beta}{2} (\beta_i - \beta)^2 \right\} \\ &\times \phi^{a-1} \exp\{-b\phi\} \tau_\alpha^{a-1} \exp\{-b\tau_\alpha\} \tau_\beta^{a-1} \exp\{-b\tau_\beta\} \\ &\times \exp \left\{ -\frac{P_\alpha}{2} \alpha^2 \right\} \exp \left\{ -\frac{P_\beta}{2} \beta^2 \right\} \end{aligned} \quad (2.5)$$

Entonces, para obtener las distribuciones condicionales totales de cada uno de los parámetros se deben tomar sólo los términos en (2.5) que los contengan. Así,

$$\begin{aligned} P(\alpha | \cdot) &\propto \prod_{i=1}^N \exp \left\{ -\frac{\tau_\alpha}{2} (\alpha_i - \alpha)^2 \right\} \exp \left\{ -\frac{P_\alpha}{2} \alpha^2 \right\} \\ &\propto \exp \left\{ -\frac{\tau_\alpha}{2} \sum_{i=1}^N (\alpha_i - \alpha)^2 - \frac{P_\alpha}{2} \alpha^2 \right\} \end{aligned}$$

$$\begin{aligned}
&\propto \exp \left\{ -\frac{1}{2} \left( (N\tau_\alpha + P_\alpha)\alpha^2 - 2\alpha\tau_\alpha \sum_{i=1}^N \alpha_i \right) \right\} \\
&\propto \exp \left\{ -\frac{N\tau_\alpha + P_\alpha}{2} \left( \alpha - \frac{\tau_\alpha \sum_{i=1}^N \alpha_i}{N\tau_\alpha + P_\alpha} \right)^2 \right\}.
\end{aligned} \tag{2.6}$$

Análogamente,

$$P(\beta|\cdot) \propto \exp \left\{ -\frac{N\tau_\beta + P_\beta}{2} \left( \beta - \frac{\tau_\beta \sum_{i=1}^N \beta_i}{N\tau_\beta + P_\beta} \right)^2 \right\}. \tag{2.7}$$

Procediendo de manera similar para  $\alpha_i$  y  $\beta_i$  se obtiene:

$$\begin{aligned}
P(\alpha_i|\cdot) &\propto \exp \left\{ -\frac{\phi}{2} \sum_{j=1}^{n_i} (y_{ij} - \alpha_i - \beta_i t_{ij})^2 \right\} \exp \left\{ -\frac{\tau_\alpha}{2} (\alpha_i - \alpha)^2 \right\} \\
&\propto \exp \left\{ -\frac{\phi}{2} \sum_{j=1}^{n_i} (y_{ij} - \alpha_i - \beta_i t_{ij})^2 - \frac{\tau_\alpha}{2} (\alpha_i - \alpha)^2 \right\} \\
&\propto \exp \left\{ -\frac{1}{2} \left[ (n_i\phi + \tau_\alpha)\alpha_i^2 - 2 \left( \phi \sum_{j=1}^{n_i} (y_{ij} - \beta_i t_{ij}) + \alpha\tau_\alpha \right) \alpha_i \right] \right\} \\
&\propto \exp \left\{ -\frac{n_i\phi + \tau_\alpha}{2} \left( \alpha_i - \frac{\phi \sum_{j=1}^{n_i} (y_{ij} - \beta_i t_{ij}) + \alpha\tau_\alpha}{n_i\phi + \tau_\alpha} \right)^2 \right\}.
\end{aligned} \tag{2.8}$$

$$\begin{aligned}
P(\beta_i|\cdot) &\propto \exp \left\{ -\frac{\phi}{2} \sum_{j=1}^{n_i} (y_{ij} - \alpha_i - \beta_i t_{ij})^2 \right\} \exp \left\{ -\frac{\tau_\beta}{2} (\beta_i - \beta)^2 \right\} \\
&\propto \exp \left\{ -\frac{\phi}{2} \sum_{j=1}^{n_i} (y_{ij} - \alpha_i - \beta_i t_{ij})^2 - \frac{\tau_\beta}{2(\beta_i - \beta)^2} \right\} \\
&\propto \exp \left\{ -\frac{1}{2} \left[ \left( \phi \sum_{j=1}^{n_i} t_{ij}^2 + \tau_\beta \right) \beta_i^2 - 2 \left( \phi \sum_{j=1}^{n_i} t_{ij} (y_{ij} - \alpha_i) + \tau_\beta \beta \right) \beta_i \right] \right\} \\
&\propto \exp \left\{ -\frac{\phi \sum_{j=1}^{n_i} t_{ij}^2 + \tau_\beta}{2} \left( \beta_i - \frac{\phi \sum_{j=1}^{n_i} t_{ij} (y_{ij} - \alpha_i) + \tau_\beta \beta}{\phi \sum_{j=1}^{n_i} t_{ij}^2 + \tau_\beta} \right)^2 \right\}.
\end{aligned} \tag{2.9}$$

Haciéndolo ahora para  $\phi$ ,  $\tau_\alpha$  y  $\tau_\beta$  se tiene

$$\begin{aligned}
P(\phi|\cdot) &\propto \phi^{\frac{n}{2}} \exp \left\{ -\frac{\phi}{2} \sum_{i=1}^N \sum_{j=1}^{n_i} (y_{ij} - \alpha_i - \beta_i t_{ij})^2 \right\} \phi^{a-1} \exp\{-b\phi\} \\
&\propto \phi^{\frac{n}{2}+a-1} \exp \left\{ -\left( \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^{n_i} (y_{ij} - \alpha_i - \beta_i t_{ij})^2 + b \right) \phi \right\}, \tag{2.10}
\end{aligned}$$

$$\begin{aligned}
P(\tau_\alpha|\cdot) &\propto \tau_\alpha^{\frac{n}{2}} \exp \left\{ -\frac{\tau_\alpha}{2} \sum_{i=1}^N (\alpha_i - \alpha)^2 \right\} \tau_\alpha^{a-1} \exp\{-b\tau_\alpha\} \\
&\propto \tau_\alpha^{\frac{n}{2}+a-1} \exp \left\{ -\left( \frac{1}{2} \sum_{i=1}^N (\alpha_i - \alpha)^2 + b \right) \tau_\alpha \right\}, \tag{2.11}
\end{aligned}$$

y de forma análoga para  $\tau_\beta$

$$P(\tau_\beta|\cdot) \propto \tau_\beta^{\frac{n}{2}+a-1} \exp \left\{ -\left( \frac{1}{2} \sum_{i=1}^N (\beta_i - \beta)^2 + b \right) \tau_\beta \right\}. \tag{2.12}$$

Observando (2.6), (2.7), (2.8), (2.9), (2.10), (2.11) y (2.12) se tiene que las distribuciones condicionales totales de todos los parámetros son las siguientes:

$$\begin{aligned}
\alpha|\cdot &\sim N \left( \frac{\tau_\alpha \sum_{i=1}^N \alpha_i}{N\tau_\alpha + P_\alpha}, (N\tau_\alpha + P_\alpha)^{-1} \right) \\
\beta|\cdot &\sim N \left( \frac{\tau_\beta \sum_{i=1}^N \beta_i}{N\tau_\beta + P_\beta}, (N\tau_\beta + P_\beta)^{-1} \right) \\
\alpha_i|\cdot &\sim N \left( \frac{\phi \sum_{j=1}^{n_i} (y_{ij} - \beta_i t_{ij}) + \tau_\alpha \alpha}{n_i \phi + \tau_\alpha}, (n_i \phi + \tau_\alpha)^{-1} \right) \\
\beta_i|\cdot &\sim N \left( \frac{\phi \sum_{j=1}^{n_i} t_{ij} (y_{ij} - \alpha_i) + \tau_\beta \beta}{\phi \sum_{j=1}^{n_i} t_{ij}^2 + \tau_\beta}, \left( \phi \sum_{j=1}^{n_i} t_{ij}^2 + \tau_\beta \right)^{-1} \right)
\end{aligned}$$

$$\begin{aligned}
\phi|\cdot &\sim G\left(n/2 + a, b + \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^{n_i} (y_{ij} - \alpha_i - \beta_i t_{ij})^2\right) \\
\tau_\alpha|\cdot &\sim G\left(N/2 + a, b + \frac{1}{2} \sum_{i=1}^N (\alpha_i - \alpha)^2\right) \\
\tau_\beta|\cdot &\sim G\left(N/2 + a, b + \frac{1}{2} \sum_{i=1}^N (\beta_i - \beta)^2\right).
\end{aligned} \tag{2.13}$$

Dado que se pueden tomar muestras directamente de las distribuciones anteriores, es sencillo implementar un método Monte Carlo como el muestreo de Gibbs para obtener muestras de la distribución posterior. Los métodos Monte Carlo se verán en la siguiente sección.

## 2.4. Simulación Monte Carlo de Cadenas de Markov

Los métodos Monte Carlo de Cadenas de Markov (MCMC) son útiles, entre otras áreas, en el análisis de imágenes, ya que permite simular a partir de campos aleatorios y optimizar funcionales de energía. Como el nombre lo indica, estos métodos se basan en la simulación de cadenas de Markov.

**Definición 2.1.** *Un método Monte Carlo de Cadenas de Markov (MCMC) para la simulación de una distribución  $\pi$  es cualquier método que produce una cadena de Markov ergódica  $(X^{(t)})$  cuya distribución estacionaria es  $\pi$ . (Robert, C.P. and G. Casella p.268)*

Construir tal cadena de Markov es bastante fácil. Este método se describe gracias a Hastings (1970), que es una generalización del método propuesto inicialmente por Metropolis *et al.* (1953).

### 2.4.1. Algoritmo Metropolis-Hastings

Según el algoritmo *Metropolis-Hastings* se construye la cadena de Markov como sigue: a cada tiempo  $t$  el siguiente estado  $X^{(t+1)}$  se escoge tomando una muestra, un punto candidato  $Y$ , de una distribución propuesta  $q(\cdot|X^{(t)})$ . El punto candidato  $Y$  es entonces *aceptado* con probabilidad  $\alpha(X^{(t)}, Y)$  donde

$$\alpha(X, Y) = \min \left( 1, \frac{\pi(Y)q(X|Y)}{\pi(X)q(Y|X)} \right) \quad (2.14)$$

Si el punto candidato es aceptado, el siguiente estado se vuelve  $X^{(t+1)} = Y$ . Si el punto candidato es rechazado, la cadena no se mueve, i.e. el siguiente estado se convierte en  $X^{(t+1)} = X^{(t)}$

Con el siguiente argumento se puede ver que la distribución propuesta  $q(\cdot|\cdot)$  puede tener cualquier forma y aun así la distribución estacionaria de la cadena será  $\pi(\cdot)$ . El kernel de transición para el algoritmo Metropolis-Hastings es

$$\begin{aligned} P(X^{(t+1)}|X^{(t)}) &= q(X^{(t+1)}|X^{(t)})\alpha(X^{(t)}, X^{(t+1)}) \\ &\quad + \mathbb{I}(X^{(t+1)} = X^{(t)})[1 - \int q(Y|X^{(t)})\alpha(X^{(t)}, Y)dY], \end{aligned} \quad (2.15)$$

donde  $\mathbb{I}(\cdot)$  denota la función indicadora (tomando el valor 1 cuando su argumento es cierto y 0 cuando es falso). El primer término de (2.15) viene de la aceptación de un candidato  $Y = X^{(t+1)}$ , y el segundo término surge del rechazo, para todos los posibles candidatos  $Y$ . Por (2.14) se puede ver que

$$\pi(X^{(t)})q(X^{(t+1)}|X^{(t)})\alpha(X^{(t)}, X^{(t+1)}) = \pi(X^{(t+1)})q(X^{(t)}|X^{(t+1)})\alpha(X^{(t+1)}, X^{(t)})$$

ya que si

$$\alpha(X^{(t)}, X^{(t+1)}) = 1 \implies \alpha(X^{(t+1)}, X^{(t)}) = \frac{\pi(X^{(t)})q(X^{(t+1)}|X^{(t)})}{\pi(X^{(t+1)})q(X^{(t)}|X^{(t+1)})},$$

o si

$$\alpha(X^{(t)}, X^{(t+1)}) = \frac{\pi(X^{(t+1)})q(X^{(t)}|X^{(t+1)})}{\pi(X^{(t)})q(X^{(t+1)}|X^{(t)})} \implies \alpha(X^{(t+1)}, X^{(t)}) = 1.$$

Así se obtiene la ecuación de *balance detallado*:

$$\begin{aligned}
\pi(X^{(t)})P(X^{(t+1)}|X^{(t)}) &= \pi(X^{(t)})q(X^{(t+1)}|X^{(t)})\alpha(X^{(t)}, X^{(t+1)}) \\
&\quad + I(X^{(t+1)} = X^{(t)})\pi(X^{(t)})[1 - \int q(Y|X^{(t)})\alpha(X^{(t)}, Y)] \\
&= \pi(X^{(t+1)})q(X^{(t)}|X^{(t+1)})\alpha(X^{(t+1)}, X^{(t)}) \\
&\quad + I(X^{(t)} = X^{(t+1)})\pi(X^{(t+1)})[1 - \int q(Y|X^{(t+1)})\alpha(X^{(t+1)}, Y)] \\
\pi(X^{(t)})P(X^{(t+1)}|X^{(t)}) &= \pi(X^{(t+1)})P(X^{(t)}|X^{(t+1)}) \tag{2.16}
\end{aligned}$$

Integrando ambos lados de (2.16) con respecto a  $X_t$  se tiene:

$$\int \pi(X^{(t)})P(X^{(t+1)}|X^{(t)})dX^{(t)} = \pi(X^{(t+1)}) \tag{2.17}$$

Si se asume que  $X^{(t)}$  es de  $\pi(\cdot)$ , entonces el lado izquierdo de la ecuación (2.17) resulta la distribución marginal de  $X^{(t+1)}$ . Así (2.17) dice que si  $X^{(t)}$  es de  $\pi$ , entonces también lo será  $X^{(t+1)}$ . Por lo tanto, una vez que una muestra de la distribución estacionaria haya sido obtenida, todas las muestras posteriores serán de esta distribución. Esto sólo prueba que  $\pi$  es la distribución estacionaria y no es una justificación completa para el algoritmo Metropolis-Hastings. Una justificación requiere probar que  $P^{(t)}(X^{(t)}|X_0)$  converge a la distribución estacionaria.

Existe una variante de este algoritmo en la que en lugar de actualizar todo  $X$  en bloque, se divide a  $X$  en componentes  $\{X_{.1}, X_{.2}, \dots, X_{.h}\}$  que posiblemente difieran de dimensión, y entonces actualizar estas componentes una por una. Esto resulta a veces más conveniente y computacionalmente eficiente. Este marco para MCMC fue el originalmente propuesto por Metropolis *et al.* (1953).

Una iteración de este algoritmo comprende  $h$  pasos de actualización, como sigue. Sea  $X_{t,i}$  el estado de  $X_{.i}$  al final de la iteración  $t$ . Para el paso  $i$  de iteración  $t + 1$ ,  $X_{.i}$  es actualizado usando Metropolis-Hastings. El candidato  $Y_{.i}$  es generado de una distribución propuesta  $q_i(Y_{.i}|X_{t,i}, X_{t,-i})$ , donde  $X_{t,-i}$  denota el valor de  $X_{.-i}$  después de completar el paso  $i - 1$  de la iteración  $t + 1$ :

$$X_{t,-i} = \{X_{t+1,1}, \dots, X_{t+1,i-1}, X_{t,i+1}, \dots, X_{t,h}\}$$

donde las componentes  $1, 2, \dots, i - 1$  ya han sido actualizadas. Por lo tanto la  $i$ -ésima distribución propuesta  $q_i(\cdot|\cdot, \cdot)$  genera un candidato solo para la  $i$ -ésima componente de  $X$ , y puede depende en los valores *actuales* de cualquiera de las componentes de  $X$ . El candidato es aceptado entonces con probabilidad  $\alpha(X_{t,-i}, X_{t,i}, Y_{.i})$  donde

$$\alpha(X_{t,-i}, X_{t,i}, Y_i) = \min \left( 1, \frac{\pi(Y_i|X_{,-i})q_i(X_i|Y_i, X_{,-i})}{\pi(X_i|X_{,-i})q_i(Y_i|X_i, X_{,-i})} \right) \quad (2.18)$$

Aquí  $\pi(X_i|X_{,-i})$  es la distribución condicional total de  $X_i$  bajo  $\pi(\cdot)$ . Si  $Y_i$  es aceptado, se establece  $X_{t+1,i} = Y_i$ ; de otro modo, se pone  $X_{t+1,i} = X_{t,i}$ . Las componentes restantes no se modifican en el paso  $i$ .

### 2.4.2. Muestreo de Gibbs

Un caso especial del algoritmo anterior es el *muestreo de Gibbs*, que fue nombrado por Geman y Geman (1984), quienes lo usaron para analizar distribuciones de Gibbs en mallas. Sin embargo sus aplicaciones no se limitan a distribuciones de Gibbs.

Para el muestreo de Gibbs, la propuesta de distribución para actualizar la  $i$ -ésima componente de  $X$  es

$$q_i(Y_i|X_i, X_{,-i}) = \pi(Y_i|X_{,-i}) \quad (2.19)$$

Sustituyendo (2.19) en (2.18) nos da una probabilidad de aceptación de 1, esto es, los candidatos del muestreo de Gibbs son siempre aceptados. Por lo tanto, el muestreo de Gibbs consiste simplemente en tomar muestras de distribuciones condicionales totales.

Los pasos a seguir son los siguientes:

1. Inicialice el contador de la cadena en  $j=1$  y establezca valores iniciales  $X^{(0)} = (X_{0,1}, \dots, X_{0,h})'$ .
2. Obtenga un nuevo valor  $X^{(j)} = (X_{j,1}, \dots, X_{j,h})'$  a partir de  $X^{(j-1)}$  mediante la generación sucesiva de valores

$$\begin{aligned} X_{j,1} &\sim \pi(X_{,1}|X_{j-1,2}, \dots, X_{j-1,h}) \\ X_{j,2} &\sim \pi(X_{,2}|X_{j,1}, X_{j-1,3}, \dots, X_{j-1,h}) \\ &\vdots \\ X_{j,h} &\sim \pi(X_{,h}|X_{j,1}, \dots, X_{j,d-1}) \end{aligned}$$



3. Incrementar el contador  $j$  y regresar al paso 2 hasta que se alcance la convergencia.

### 2.4.3. Implementación

Como se vio en las secciones anteriores, los métodos MCMC son métodos numéricos e iterativos, por lo que estrategias para mejorar la eficiencia de éstos pueden tener un gran impacto en su costo computacional.

La forma más obvia de obtener una muestra de tamaño  $n$  de la distribución posterior  $\pi$  es procesar  $n$  cadenas en paralelo hasta la convergencia, dígase  $m$  iteraciones, y tómesese como muestra el  $m$ -ésimo elemento de cada una de las  $n$  cadenas. Este procedimiento requiere generar  $nm$  valores de la cadena. Si las cadenas son inicializadas independientemente, la muestra consiste de valores independientes de  $\pi$ . Es más sencillo establecer la independencia si los valores iniciales son distintos.

Otra forma es considerar una sola cadena, en donde una muestra de tamaño  $n$  de  $\pi$  puede ser formada por  $n$  valores consecutivos de la cadena, después de  $m$  iteraciones con las que se logró la convergencia. Este proceso requerirá generar  $m+n$  valores de la cadena, mucho menos que el método anterior. El problema es que los elementos de la muestra no son independientes, pero sigue siendo válido hacer inferencia sobre esta muestra gracias a teoremas ergódicos.

Un enfoque alternativo pensando en la independencia es tomar valores para la muestra cada  $k$ -ésimo valor de la cadena después de las  $m$  iteraciones para la convergencia. Los valores de la muestra se vuelven cada vez menos correlacionados y son virtualmente independientes para valores grandes de  $k$ . Para este procedimiento se requieren  $m+kn$  iteraciones. Con esta estrategia se mejora con respecto al muestreo independiente. Sin embargo, la eficiencia no aumenta y la estimación resulta ser menos precisa que tomando para la muestra todos los valores de la cadena.

Cualquiera que sea el esquema usado para generar la muestra de tamaño  $n$ , se pueden usar los valores obtenidos  $\theta_1, \dots, \theta_n$  para encontrar estimadores de una función real cualquiera del vector de parámetros  $\theta$ ,  $\psi = t(\theta)$ . La esperanza posterior de  $\psi$  se estima por  $\hat{E}(\psi) = \hat{\psi} = (1/n) \sum_{j=1}^n \psi_j$ , donde  $\psi_j = t(\theta_j)$ ,  $j=1, \dots, n$ . La varianza se estima de forma similar,  $Var(\psi) = \hat{\sigma}_\psi^2 = \hat{E}(\psi^2) - [\hat{E}(\psi)]^2 = 1/n \sum_{j=1}^n (\psi_j - \hat{\psi})^2$ .

Los intervalos de confianza se obtienen similarmente estimando los intervalos límites por los respectivos cuantiles de la muestra, es decir, un intervalo del  $100(1 - \alpha)\%$  de confianza para  $\psi$  está aproximadamente dado por  $[\psi_{(n\alpha/2)}, \psi_{(n(1 - \alpha/2))}]$  donde  $\psi_{(j)}$

es el  $j$ -ésimo valor de la muestra ordenada y  $[x]$  es el mayor entero menor que  $x$ . Por ejemplo, si  $n=1000$  y  $\alpha=.05$ , el intervalo de confianza estará definido por los valores 25o. y el 975o. de  $\psi$ .

#### 2.4.4. Diagnósticos de Convergencia

La demostración de la convergencia de la distribución de  $X^{(t)}$  a  $\pi$  cuando  $t \rightarrow \infty$  fue dada por Geman, S. y Geman, D. (1984). Esta prueba de la convergencia se hizo sin importar el valor inicial  $X^{(0)}$ .

A pesar de que se sabe que la cadena converge, la parte más problemática de los cálculos con un método Monte Carlo de cadenas de Markov es decidir cuándo es seguro detener el algoritmo y resumir los resultados. Esto significa que debemos adivinar a partir de qué iteración  $t_0$  todo valor arrojado pueda pensarse que proviene de la verdadera distribución estacionaria de la cadena de Markov, es decir, la verdadera distribución posterior.

Existen dos formas de atacar este problema. El primero es más teórico e intenta medir distancias y establecer límites en funciones de distribuciones generadas de la cadena.

El estudio de la convergencia de la cadena también puede hacerse analizando las propiedades estadísticas de la muestra. Este es un enfoque totalmente empírico y obviamente es mucho más práctico, pero la dificultad que se presenta es que este método no garantiza la convergencia, ya que es un método basado en observaciones de la cadena.

Gelfand y Smith (1990) sugieren realizar chequeos informales de convergencia basados en técnicas gráficas. Después de  $m$  iteraciones en  $n$  cadenas paralelas, se grafica el histograma de los  $n$  valores de la  $m$ -ésima corrida. Este procedimiento se repite  $k$  veces, este valor no es necesariamente grande si uno sospecha que la convergencia se alcanza después de  $m$  iteraciones; normalmente, valores entre 10 y 50 son razonables. Entonces la convergencia es aceptada si los histogramas no pueden ser distinguidos.

Ideas parecidas pueden ser usadas con una sola cadena. Una trayectoria de la cadena que exhibe el mismo comportamiento a través de varias iteraciones es un indicativo de convergencia. De manera similar, se puede graficar la trayectoria de los promedios ergódicos  $\bar{y}$ , si ésta tiene un comportamiento asintótico después de varias iteraciones, indica la convergencia.

El siguiente ejemplo muestra la estrategia gráfica para inferir a partir de qué iteración los datos arrojados ya pertenecen a una muestra de la distribución posterior en la que se ve si la trayectoria de una sola cadena exhibe el mismo comportamiento.

**Ejemplo 2.4.** Retómese el ejemplo(2.3) de la sección anterior, donde los datos que se tienen son los pesos ganados de  $N=68$  mujeres en 6 visitas a una clínica ginecológica. Los tiempos a los cuales las observaciones fueron realizadas son los siguientes:

$$t_{ij} = \begin{cases} 5j & i \leq 34, j \leq 3 \\ 17 + 5j & i \leq 34, j > 3 \\ 2 + 5j & i \geq 34, j \leq 3 \\ 17 + 4j & i \geq 34, j > 3 \end{cases}$$

Se realizaron 5000 iteraciones de 2 cadenas paralelas con distintos valores iniciales del algoritmo de muestreo de Gibbs con las distribuciones condicionales totales en (2.13).

En la Figura 2.3 se muestran las trayectorias de 4 parámetros del modelo para las 2 cadenas. Estas gráficas parecen indicar que la convergencia se alcanza alrededor de las 2000 iteraciones.

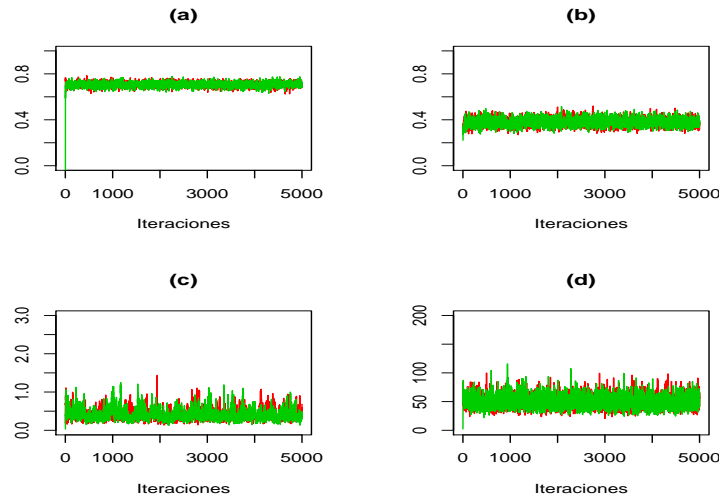


Figura 2.3: Trayectorias de parámetros del modelo con número de iteraciones para ambas cadenas. Los parámetros son: (a)  $\beta$ , el crecimiento poblacional; (b)  $\phi$ , la precisión de las observaciones; (c)  $\tau_\alpha$ , la precisión de la población de interceptos; (d) la precisión de la población de coeficientes de regresión. Los valores iniciales para la primera cadena son:  $\alpha_i = \alpha = -5$ ,  $\beta_i = \beta = 0.5$  para  $i = 1, \dots, N$ ,  $\phi = 1, \tau_\alpha = 15$ ,  $\tau_\beta = 1$ ; para la segunda cadena los valores iniciales son:  $\alpha_i = \alpha = -4$  y  $\beta_i = \beta = 0$  para  $i = 1, \dots, N$ ,  $\phi = 1$ ,  $\tau_\alpha = 10$ ,  $\tau_\beta = 1$ .

## Capítulo 3

# Reconstrucción de Imágenes

### 3.1. El Modelo

El primer paso de nuestro trabajo es “reconstruir” las imágenes de temperatura y clorofila. Supongamos que se tiene una imagen de  $n$  pixeles donde la variable de interés  $Y_i$  es la temperatura (clorofila) en el pixel  $i$ . Como algunos pixeles no fueron observados, se obtendrán valores estimados a partir de los que sí fueron observados, por lo que el valor esperado de temperatura (clorofila) en el pixel  $i$  dada la temperatura (clorofila) en los pixeles vecinos será el promedio de éstas.

Entonces,

$$\begin{aligned} E(Y_i|Y_{-i}) &= \frac{1}{n_i} \sum_{j \in \partial(i)} (Y_j) \\ Prec(Y_i|Y_{-i}, W) &= \kappa_i > 0, \end{aligned}$$

donde  $n_i = \sum_{j \in \partial(i)} 1$ .

Esto lo podemos reescribir como sigue:

$$\begin{aligned} E(Y_i|Y_{-i}) &= - \sum_{j=1}^n \beta_{ij}(Y_j) \\ Prec(Y_i|Y_{-i}, W) &= \kappa_i, \end{aligned}$$

con  $\beta_{ij} = -\frac{1}{n_i}\mathbb{I}\{j \in \partial(i)\}$ . Para poder usar el Teorema 1.4 y decir que estas esperanzas y precisiones condicionales totales definen un GMRF se necesita que  $\kappa_i\beta_{ij} = \kappa_j\beta_{ji}$ . Por cómo está definido  $\beta_{ij}$ , sólo interesan los valores distintos de cero, es decir, cuando  $j \in \partial(i)$ ; se necesita entonces que  $\frac{\kappa_i}{n_i} = \frac{\kappa_j}{n_j}$ . Comencemos con el pixel 1, entonces

$$\frac{\kappa_1}{n_1} = \frac{\kappa_j}{n_j} \quad \forall j \in \partial(1) \quad \Rightarrow \quad \kappa_1 = n_1 \frac{\kappa_j}{n_j} \quad \forall j \in \partial(1),$$

así

$$\frac{\kappa_j}{n_j} = \frac{\kappa_i}{n_i} \quad \forall j \neq i, j, i \in \partial(1) \quad \Rightarrow \quad \frac{\kappa_j}{n_j} = \phi \quad \forall j \in \partial(1).$$

Sea  $j \in \partial(1)$ , entonces  $\frac{\kappa_i}{n_i} = \frac{\kappa_j}{n_j} = \phi \quad \forall i \in \partial(j)$ . Así se procede transitivamente hasta que  $\frac{\kappa_i}{n_i} = \phi \quad \forall i = 1, \dots, n$ , de donde se obtiene que  $\kappa_i = n_i\phi \quad \forall i = 1, \dots, n$ .

Entonces ya es válido usar el Teorema 1.4, con el cual se obtiene  $\mathbf{Q}$  como sigue:

$$Q_{ij} = \begin{cases} -\phi & i \in \partial(j) \\ n_i\phi & i = j \\ 0 & e.o.c \end{cases}$$

El modelo queda completo con una distribución a priori para  $\phi$ :

$$\begin{aligned} \mathbf{Y}|\phi &\sim \text{GMRF}(0, \mathbf{Q}^{-1}) \\ \phi &\sim G(a_\phi, b_\phi) \end{aligned} \tag{3.1}$$

La función de densidad conjunta  $\pi(\mathbf{Y}, \phi)$  resulta:

$$\pi(\mathbf{Y}, \phi) \propto \phi^{\frac{n}{2}} \exp\left\{-\frac{1}{2}\mathbf{Y}^T \mathbf{Q} \mathbf{Y}\right\} \phi^{a_\phi-1} \exp\{-b_\phi\phi\} \tag{3.2}$$

Ahora se obtendrá la distribución condicional total de  $\phi$  tomando sólo los términos de (3.2) que lo contienen.

$$\begin{aligned}
\pi(\phi | \cdot) &\propto \phi^{\frac{n}{2}} \phi^{a_\phi - 1} \exp \left\{ -\frac{\phi}{2} \mathbf{Y}^T \mathbf{Q}' \mathbf{Y} - b_\phi \phi \right\} \\
&\propto \phi^{a_\phi + \frac{n}{2} - 1} \exp \left\{ -\left( \frac{1}{2} \mathbf{Y}^T \mathbf{Q}' \mathbf{Y} + b_\phi \right) \phi \right\}
\end{aligned} \tag{3.3}$$

donde  $\mathbf{Q}' = \phi^{-1} \mathbf{Q}$

De (3.3) concluimos que:

$$\phi | \cdot \sim G(a'_\phi, b'_\phi) \tag{3.4}$$

donde  $a'_\phi = a_\phi + \frac{n}{2}$ , y  $b'_\phi = \frac{1}{2} \mathbf{Y}^T \mathbf{Q}' \mathbf{Y} + b_\phi$ .

### 3.2. Muestreo de Gibbs en la Reconstrucción de Imágenes

Para ejemplificar el método de reconstrucción de imágenes se tiene una imagen con datos de temperaturas tomada del 9 al 16 de enero de 2009.

Usando el modelo en (3.1) y las distribuciones condicionales totales obtenidas en la sección anterior se procede a aplicar el método de muestreo de Gibbs.

Para saber a partir de qué iteración del algoritmo ya se obtienen muestras de la distribución posterior usamos el método de chequeo gráfico. Por el tamaño de la imagen, 2251422 pixeles, es necesario realizar 200000 iteraciones de 2 cadenas paralelas con distintos valores iniciales, a la primera cadena se le asignaron valores iniciales de 0 a los pixeles no observados y a la segunda valores de 10, mientras que a  $\phi$  se le dio un valor inicial de 1 en ambas cadenas. En las figuras 3.1 y 3.2 se observan las trayectorias de  $\phi$  y de la temperatura en 6 pixeles tomados aleatoriamente; puede verse que la convergencia se alcanza aproximadamente a las 120000 iteraciones.

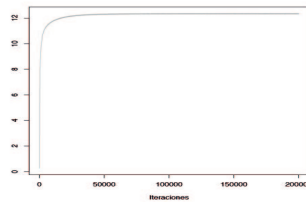


Figura 3.1: Trayectoria de  $\phi$  en el modelo de la temperatura.

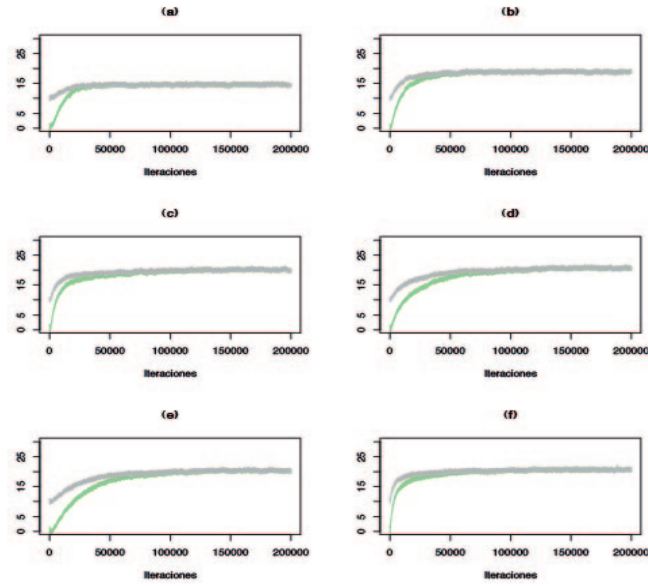


Figura 3.2: Trayectoria de 2 cadenas paralelas de la temperatura en 6 pixeles: (a)pixel 59500, (b)pixel 59800, (c)pixel 546163, (d)pixel 326472, (e)pixel 312371, (f)pixel 477495

Para reconstruir una imagen de clorofila se sigue el mismo procedimiento, usando el mismo modelo ahora con los datos de clorofila tomados del 9 al 16 de enero de 2009. Se realizan 180000 iteraciones de 2 cadenas con valores iniciales distintos para saber a partir de qué iteración ya se tienen muestras de la distribución posterior. Las trayectorias de algunos pixeles y de la clorofila en 6 pixeles tomados al azar pueden verse en las Figuras 3.3 y 3.4, de donde se puede concluir que la convergencia se alcanza aproximadamente después de 40000 iteraciones.

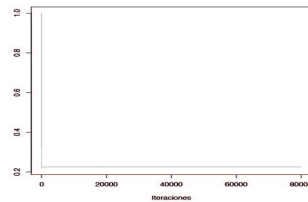


Figura 3.3: Trayectoria de  $\phi$  en el modelo de la clorofila.

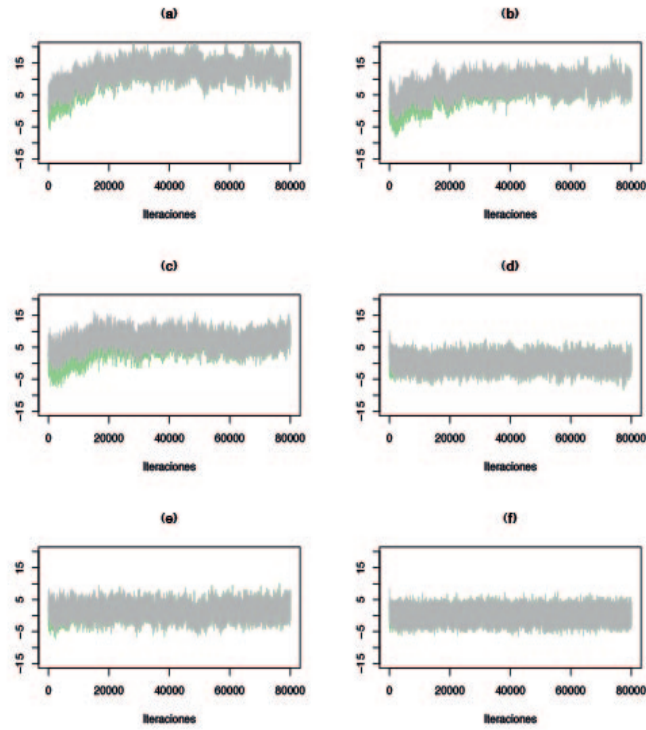


Figura 3.4: Trayectoria de 2 cadenas paralelas de la temperatura en 6 pixeles: (a)pixel 59500, (b)pixel 59800, (c)pixel 546163, (d)pixel 326472, (e)pixel 312371, (f)pixel 477495

### 3.3. Resultados

Después de la implementación del muestreo de Gibbs se obtiene una muestra de tamaño  $n=5000$  de la distribución posterior. Con esta muestra se pueden obtener los valores esperados de temperatura y clorofila para cada pixel de la imagen.

En las siguientes figuras se puede ver la reconstrucción de las imágenes de temperatura y clorofila tomadas en el periodo del 17 al 24 de enero del 2009.



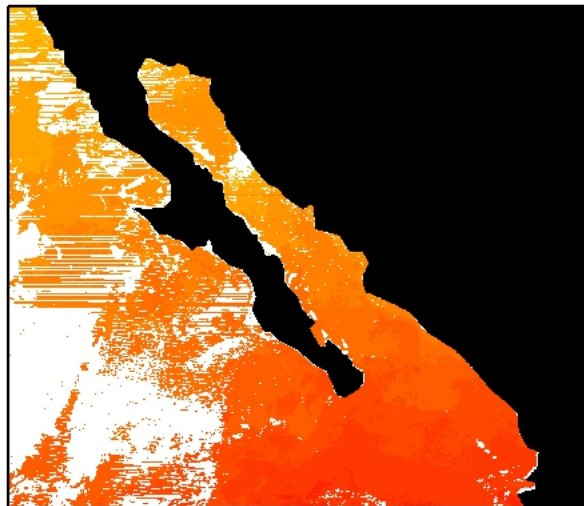


Figura 3.5: Imagen de temperatura incompleta e imagen reconstruída.

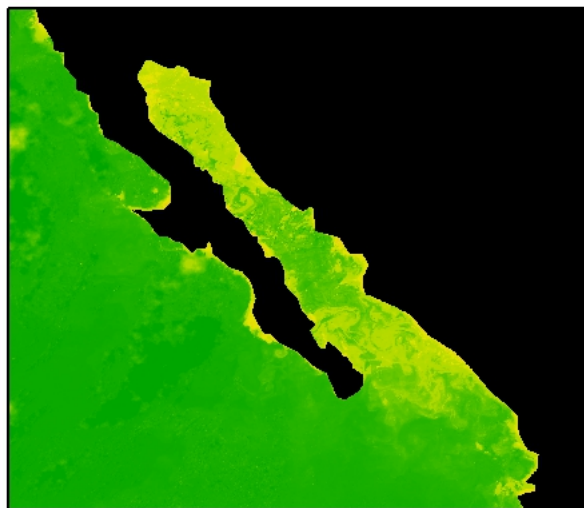
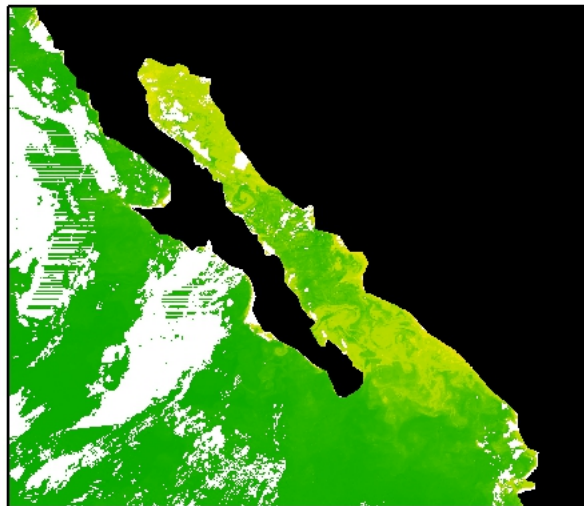


Figura 3.6: Imagen de clorofila incompleta e imagen reconstruída.

## Capítulo 4

# La Captura

### 4.1. El Modelo

Si nuestra imagen consta de  $n$  pixeles, sea  $Y_i$  el dato de captura de atún en el pixel  $i$ . Se propone el siguiente modelo de regresión lineal para  $Y_i$

$$Y_i = \beta_0 + \beta_1 T_i + \beta_2 C_i + \epsilon_i,$$

donde  $T_i$ ,  $C_i$  son la temperatura y la clorofila en el pixel  $i$  respectivamente y  $\epsilon_i$  el error. Así el modelo para el campo aleatorio  $\mathbf{Y}$  es

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \quad (4.1)$$

donde  $\boldsymbol{\epsilon}$  es un vector de errores de tamaño  $n$ , tal que  $\boldsymbol{\epsilon} \sim GMRF(\mathbf{0}, \mathbf{Q}^{-1})$ ,  $\mathbf{X}$  es la matriz de  $(n \times 3)$  cuyo renglón  $i$  es el vector  $(1, T_i, C_i)$ , y  $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)$ . De esta forma se tiene que  $\mathbf{Y} \sim N(\mathbf{X}\boldsymbol{\beta}, \mathbf{Q}^{-1})$ .

El modelo jerárquico que se utilizará es el siguiente

$$\begin{aligned} \mathbf{Y} &\sim N(\mathbf{X}\boldsymbol{\beta}, \mathbf{Q}^{-1}) \\ \boldsymbol{\beta} &\sim N(\mathbf{0}, \boldsymbol{\Sigma}) \\ \tau_i &\sim G(a_i, b_i) \\ \phi &\sim G(a_\phi, b_\phi). \end{aligned}$$

Aquí,  $\boldsymbol{\Sigma}$  es la matriz diagonal de  $(3 \times 3)$  con  $\Sigma_{ii} = \tau_{i-1}^{-1}$  y  $\mathbf{Q}$  se define como en (3.1). El siguiente paso es obtener la función de densidad conjunta, suponiendo independencia

entre los parámetros  $\boldsymbol{\tau}$  y  $\phi$ .

$$\begin{aligned}
\pi(\mathbf{Y}, \boldsymbol{\beta}, \boldsymbol{\tau}, \phi) &\propto \phi^{n/2} \exp \left\{ -\frac{1}{2} (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^T \mathbf{Q} (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) \right\} \\
&\times \prod_{i=0}^2 \tau_i^{1/2} \exp \left\{ -\frac{1}{2} (\boldsymbol{\beta}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\beta}) \right\} \\
&\times \prod_{i=0}^2 \tau_i^{a_i-1} \exp \{-b_i \tau_i\} \\
&\times \phi^{a_\phi-1} \exp \{-b_\phi \phi\}
\end{aligned} \tag{4.2}$$

Para poder obtener las distribuciones condicionales totales de  $\beta_i$ ,  $i = 0, 1, 2$ , es necesario realizar el siguiente cálculo:

$$\begin{aligned}
(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^T \mathbf{Q} (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) &= \phi (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^T \mathbf{Q}' (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) \\
&= \phi \sum_{i=1}^n [n_i (Y_i - \beta_0 - \beta_1 T_i - \beta_2 C_i) - \sum_{j \in \partial(i)} (Y_j - \beta_0 - \beta_1 T_j - \beta_2 C_j)] \\
&\times [Y_i - \beta_0 - \beta_1 T_i - \beta_2 C_i] \\
&= \phi \sum_{i=1}^n [n_i (Y_i - \beta_0 - \beta_1 T_i - \beta_2 C_i)^2 \\
&- (Y_i - \beta_0 - \beta_1 T_i - \beta_2 C_i) \sum_{j \in \partial(i)} (Y_j - \beta_0 - \beta_1 T_j - \beta_2 C_j)]
\end{aligned}$$

Con este resultado obtenemos la distribución condicional total de  $\beta_i$  tomando los términos en (4.2) que lo contienen. Comenzando con  $\beta_0$  se tiene que

$$\begin{aligned}
\pi(\beta_0 | \cdot) &\propto \exp \left\{ -\frac{1}{2} \left( \phi \sum_{i=1}^n [n_i (\beta_0^2 - 2\beta_0 (Y_i - \beta_1 T_i - \beta_2 C_i)) + \beta_0 \sum_{j \in \partial(i)} (Y_j - \beta_1 T_j - \beta_2 C_j)] \right) \right\} \\
&\times \exp \left\{ -\frac{1}{2} \left( \phi \sum_{i=1}^n [\beta_0 (Y_i - \beta_1 T_i - \beta_2 C_i) n_i - \beta_0^2 n_i] + \tau_0 \beta_0^2 \right) \right\} \\
&\propto \exp \left\{ -\frac{1}{2} \left( \tau_0 \beta_0^2 - \beta_0 \phi \sum_{i=1}^n [n_i (Y_i - \beta_1 T_i - \beta_2 C_i) - \sum_{j \in \partial(i)} (Y_j - \beta_1 T_j - \beta_2 C_j)] \right) \right\}
\end{aligned}$$

$$\propto \exp \left\{ -\frac{\tau_0}{2} \left( \beta_0 - \frac{\phi R_0}{2\tau_0} \right)^2 \right\}, \quad (4.3)$$

con

$$R_0 = \sum_{i=1}^n [n_i(Y_i - \beta_1 T_i - \beta_2 C_i) - \sum_{j \in \partial(i)} (Y_j - \beta_1 T_j - \beta_2 C_j)].$$

Haciendo lo mismo para  $\beta_1$  resulta

$$\begin{aligned} \pi(\beta_1 | \cdot) &\propto \exp \left\{ -\frac{1}{2} \left( \phi \sum_{i=1}^n [n_i(\beta_1^2 T_i^2 - 2\beta_1 T_i(Y_i - \beta_0 - \beta_2 C_i)) + \beta_1 T_i \sum_{j \in \partial(i)} (Y_j - \beta_0 - \beta_2 C_j)] \right) \right\} \\ &\times \exp \left\{ -\frac{1}{2} \left( \phi \sum_{i=1}^n [\beta_1(Y_i - \beta_0 - \beta_2 C_i) \left( \sum_{j \in \partial(i)} T_j \right) - \beta_1^2 T_i \left( \sum_{j \in \partial(i)} T_j \right)] + \tau_1 \beta_1^2 \right) \right\} \\ &\propto \exp \left\{ -\frac{1}{2} (\beta_1^2 (\tau_1 + \phi S_1) - \beta_1 \phi R_1) \right\} \\ &\propto \exp \left\{ -\frac{\tau_1 + \phi S_1}{2} \left( \beta_1 - \frac{\phi R_1}{2(\tau_1 + \phi S_1)} \right)^2 \right\}, \end{aligned} \quad (4.4)$$

donde

$$\begin{aligned} S_1 &= \sum_{i=1}^n [n_i T_i^2 - T_i \sum_{j \in \partial(i)} T_j] \quad \text{y} \\ R_1 &= \sum_{i=1}^n [(Y_i - \beta_0 - \beta_2 C_i)(2n_i T_i - \sum_{j \in \partial(i)} T_j) - T_i \sum_{j \in \partial(i)} (Y_j - \beta_0 - \beta_2 C_j)]. \end{aligned}$$

Se procede análogamente para  $\beta_2$  resultando

$$\pi(\beta_2 | \cdot) \propto \exp \left\{ -\frac{\tau_2 + \phi S_2}{2} \left( \beta_2 - \frac{\phi R_2}{2(\tau_2 + \phi S_2)} \right)^2 \right\}, \quad (4.5)$$

donde

$$S_2 = \sum_{i=1}^n [n_i C_i^2 - C_i \sum_{j \in \partial(i)} C_j] \quad y$$

$$R_2 = \sum_{i=1}^n [(Y_i - \beta_0 - \beta_1 T_i)(2n_i C_i - \sum_{j \in \partial(i)} C_j) - C_i \sum_{j \in \partial(i)} (Y_j - \beta_0 - \beta_1 T_j)].$$

Ahora se obtienen las distribuciones condicionales totales de  $\tau_i$ ,  $i = 0, 1, 2$ :

$$\begin{aligned} \pi(\tau_i | \cdot) &\propto \tau_i^{1/2} \exp\left\{-\frac{\tau_i}{2} \beta_i^2\right\} \tau_i^{a_i-1} \exp\{-b_i \tau_i\} \\ &\propto \tau_i^{(a_i+1/2)-1} \exp\left\{-\left(\frac{\beta_i^2}{2} + b_i\right) \tau_i\right\}. \end{aligned} \quad (4.6)$$

Y por último, la de  $\phi$  resulta

$$\begin{aligned} \pi(\phi | \cdot) &\propto \phi^{n/2} \exp\left\{-\frac{\phi}{2} (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^T \mathbf{Q}' (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})\right\} \phi^{a_\phi-1} \exp\{-b_\phi \phi\} \\ &\propto \phi^{(n/2+a_\phi)-1} \exp\left\{-\left[\frac{1}{2} (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^T \mathbf{Q}' (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) + b_\phi\right] \phi\right\} \end{aligned} \quad (4.7)$$

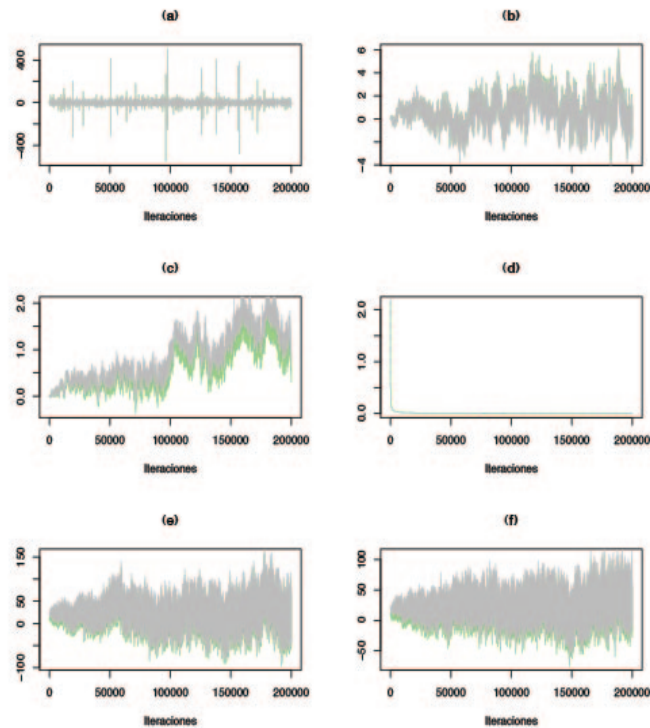
Resumiendo, las distribuciones condicionales totales que se obtuvieron son:

$$\begin{aligned} \beta_0 | \cdot &\sim N\left(\frac{\phi \sum_{i=1}^n [n_i (Y_i - \beta_1 T_i - \beta_2 C_i) - \sum_{j \in \partial(i)} (Y_j - \beta_1 T_j - \beta_2 C_j)]}{2\tau_0}, \tau_0^{-1}\right) \\ \beta_1 | \cdot &\sim N\left(\frac{\phi R_1}{2(\tau_1 + \phi S_1)}, (\tau_1 + \phi S_1)^{-1}\right) \\ \beta_2 | \cdot &\sim N\left(\frac{\phi R_2}{2(\tau_2 + \phi S_2)}, (\tau_2 + \phi S_2)^{-1}\right) \\ \tau_i | \cdot &\sim G\left(a_i + \frac{1}{2}, \frac{\beta_i^2}{2} + b_i\right) \\ \phi | \cdot &\sim G\left(a_\phi + \frac{n}{2}, \frac{1}{2} (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^T \mathbf{Q}' (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) + b_\phi\right) \end{aligned}$$

## 4.2. Muestreo de Gibbs en la Estimación de los Parámetros de la Captura

Las imágenes de temperatura y clorofila fueron tomadas en 4 periodos del mes de enero, del 1 al 8, del 9 al 16, del 16 al 24 y del 25 al 31. Dado que se tienen poco datos de captura durante todo el año y que entonces las capturas del mes de enero son aún menos, se tienen que tomar los valores de la media de todo el mes de enero de la temperatura y clorofila en cada pixel. Ya con estos datos promedio de temperatura y clorofila se procede a implementar el muestreo de Gibbs con las distribuciones condicionales obtenidas en la sección anterior.

Se realizan 200000 iteraciones del algoritmo y se siguen las trayectorias de  $\beta_i$ ,  $\phi$  y de la captura en 2 pixeles. Véase la figura 4.2.



### **4.3. Resultados**



## Capítulo 5

# Conclusiones

# Bibliografía

- [1] Gelfand, A.E. y Smith, A.F.M. (1990) Sampling-based approachesto calculating marginal densities. *Journal of the American Statistical Association*, **85**, 398-409.
- [2] Rue, H. y Held, L.(2005). *Gaussian Markov Random Fields: Theory and Applications*. Chapman and Hall/CRC
- [3] García-Borbón, J. A., Olvera, R. M., y Cerecedo, J.L. (1991). Biomasa Reproductora y Rendimiento Potencial de Tres Especies de Tunidos, Mortalidad y Supervivencia Larval en la Z.E.E. Mexicana del Golfo de México. *Ciencia Pesquera*, **8**, 41-62.
- [4] Bulgakov, S. N. y Martínez, A. (2006). Surgencia y vientos favorables en la costa oriental del Pacífico Mexicano. *Los Recursos Pesqueros y Acuícolas de Jalisco, Colima y Michoacán*. 29-40
- [5] Sabins, F. F. (1986). *Remote Sensing: Principles and Interpretation*. W. H. Freeman and Company. 2a ed.
- [6] Gamerman, D., Lopes, H. F. (2006). *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. Chapman and Hall/CRC. 2a ed.
- [7] Robert, C. P., Casella, G. (2004). *Monte Carlo Statistical Methods*
- [8] Geman, S. y Geman, D. (1984). Stochastic relaxation, Gibbs distributions and Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **6**, 721-41.
- [9] *Markov Chain Monte Carlo in Practice*
- [10] Banerjee, S., Carlin, B.P., y Gelfand, A.E. (2004) *Hierarchical Modeling and Analysis of Spatial Data*. Chapman and Hall/CRC

- [11] Christensen, R. (1990). *Linear Models for Multivariate, Time Series, and Spatial Data*. Springer