



**UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO**  
PROGRAMA DE MAESTRÍA Y DOCTORADO EN CIENCIAS MATEMÁTICAS Y  
DE LA ESPECIALIZACIÓN EN ESTADÍSTICA APLICADA.

MAPAS DE MORTALIDAD BAYESIANOS

TESIS  
QUE PARA OPTAR POR EL GRADO DE:  
MAESTRO (A) EN CIENCIAS

PRESENTA:  
ROMÁN AGUIRRE PÉREZ

DIRECTOR:  
EDUARDO GUTIÉRREZ PEÑA  
INSTITUTO DE INVESTIGACIONES EN MATEMÁTICAS APLICADAS Y EN  
SISTEMAS

CIUDAD DE MÉXICO, **MAYO DE 2016.**



Universidad Nacional  
Autónoma de México



**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.



# *Agradecimientos*

Al Consejo Nacional de Ciencia y Tecnología (CONACYT), por haberme brindado el apoyo económico para realizar mis estudios de maestría.

Investigación realizada gracias al Programa de Apoyo a Proyectos de Investigación e Innovación Tecnológica (PAPIIT) de la UNAM IN106114 “Análisis Bayesiano Aplicado”. Agradezco a la DGAPA-UNAM la beca recibida.



# Índice general

<b>1. Introducción</b>	<b>1</b>
<b>2. Preliminares</b>	<b>7</b>
2.1. Epidemiología . . . . .	7
2.1.1. Conceptos básicos . . . . .	8
2.1.2. Estandarización por edad . . . . .	9
2.2. El paradigma Bayesiano . . . . .	13
2.2.1. Inferencia Bayesiana . . . . .	14
2.2.2. Monte Carlo vía Cadenas de Markov . . . . .	16
2.2.3. Modelos Jerárquicos . . . . .	21
2.3. Estadística Espacial . . . . .	22
2.3.1. Campos Aleatorios de Markov y el Teorema de Hammersley-Clifford . . . . .	24
<b>3. Modelos para Mapas de Mortalidad</b>	<b>27</b>
3.1. Enfoques tradicionales . . . . .	31
3.2. Efectos correlacionados . . . . .	34
3.2.1. Distribuciones conjuntas . . . . .	34
3.2.2. Distribuciones marginales . . . . .	35
3.3. Modelos semi-parámétricos . . . . .	38
3.3.1. Mezcla finita de distribuciones . . . . .	39
3.3.2. Particiones . . . . .	40
3.4. Extensiones . . . . .	42
<b>4. Mapas de mortalidad por cáncer de mama para el sur del Estado de México</b>	<b>45</b>
4.1. Cáncer de mama en México . . . . .	46
4.2. Distribución espacial de la mortalidad por cáncer de mama en el sur del Estado de México: 2010-2013 . . . . .	48
4.2.1. Metodología . . . . .	48
4.2.2. Resultados . . . . .	52
<b>5. Conclusiones</b>	<b>59</b>

---

<b>A. Códigos</b>	<b>61</b>
A.1. Mapa de mortalidad estandarizada . . . . .	61
A.2. Mapa de mortalidad: BYM . . . . .	62
A.3. Mapa de mortalidad: GR . . . . .	63
<b>Bibliografía</b>	<b>71</b>

# Índice de figuras

4.1. Media ergódica de $k$ . . . . .	53
4.2. Tasa de Mortalidad Estandarizada por cáncer de mama para los 14 municipios del sur del Estado de México: 2010-2013. . . . .	55
4.3. Estimación del riesgo relativo, vía modelo BYM, por cáncer de mama para los 14 municipios del sur del Estado de México: 2010-2013. . . . .	56
4.4. Estimación del riesgo relativo, vía modelo GR, por cáncer de mama para los 14 municipios del sur del Estado de México: 2010-2013. . . . .	57



# Índice de tablas

4.1. Casos observados y esperados de los 14 municipios del sur del Estado de México: 2010-2013. . . . .	58
4.2. Estimaciones del riesgo relativo de los 14 municipios del sur del Estado de México: 2010-2013. . . . .	58



# *Resumen*

En estudios epidemiológicos, el empleo de mapas para representar la distribución geográfica de enfermedades o decesos permite generar hipótesis acerca de las posibles causas en la variación del riesgo relativo así como la identificación de regiones con riesgo alto o bajo.

Los enfoques tradicionales para la construcción de mapas de mortalidad se basan en la distribución de las tasas de mortalidad estandarizadas. Sin embargo, estos enfoques no permiten estimar la distribución de probabilidad subyacente. Recientemente, los trabajos se enfocan en plantear métodos estadísticos para proporcionar una estimación confiable del riesgo relativo tomando en cuenta la variación “extra-Poisson”, la cual puede deberse a la existencia de conglomerados y/o a la correlación espacial.

En el presente trabajo de tesis abordamos el problema de estimar la distribución espacial de una enfermedad rara desde una perspectiva bayesiana. En particular, se analiza la mortalidad por cáncer de mama en el sur del Estado de México para el periodo 2010-2013.



# Capítulo 1

## Introducción

La Epidemiología surgió de los esfuerzos por comprender la naturaleza de los fenómenos infecciosos colectivos que aparecieron de forma inexplicable, conocidos como plagas, pestes o epidemias. Ya en el siglo XX los estudios epidemiológicos se extendieron a las enfermedades y problemas de salud en general. Hoy en día podemos definir a la Epidemiología como la disciplina científica que estudia la distribución, la frecuencia, los factores determinantes, las predicciones y el control de los factores relacionados con la salud y con las distintas enfermedades existentes en poblaciones humanas definidas (López-Moreno *et al.*, 2000). En vista de esto, y debido a la naturaleza de dichas actividades, surge el uso de metodología estadística como herramienta fundamental para la formulación y validación de hipótesis sobre las características del fenómeno de interés. En términos generales, los procedimientos estadísticos tratan de dar sentido a un conjunto de datos y, como consecuencia natural de la simbiosis modelo-datos, surge el proceso de inferencia.

Dentro del quehacer epidemiológico, como consecuencia de que un componente causal de la enfermedad es el medio ambiente, los factores ambientales y el análisis de los patrones espaciales de la enfermedad son muy útiles en la formulación y validación de hipótesis etiológicas. Esto conlleva el uso de técnicas estadísticas para la modelación de las posibles interacciones con factores geográficos presentes en la región de estudio. Cabe resaltar que, en la práctica, además de los datos

observados no siempre se cuenta con informaci on adicional para relacionar alg un factor ambiental con las causas. Este es el caso en las enfermedades raras, referidas as ı por su baja incidencia en la poblaci on. El panorama anterior deja claro que no basta la formulaci on matem atica para el an alisis, as ı que se involucran m as herramientas de otra  ındole; por ejemplo, el mapa de la regi on en donde se estudia el fen omeno de inter es ofrece una fuente m as de informaci on.

El empleo de mapas para representar algunas realidades y acontecimientos relevantes en materia de salud p blica, proporciona una imagen m as elocuente acerca de la distribuci on espacial que la simple observaci on de datos en tablas ([Silva-Aycaguer et al., 2003](#)). En este sentido, el valor de los atlas radica en que proporcionan gran cantidad de informaci on acerca de la relaci on entre causas de la enfermedad y lugar de ocurrencia de  sta. Por ejemplo, [Snow \(1955\)](#) formul  la hip tesis de la transmisi on del c lera por el agua contaminada en el barrio Soho de Londres, en donde un reciente brote epid mico hab a matado a m as de 500 personas en un lapso de diez d as. Snow (a quien se considera el precursor de la epidemiolog a contempor nea) marc  en el mapa del barrio las defunciones con l neas de color negro que se iban apilando unas sobre las otras a medida que el n mero de decesos aumentaba; adem s georreferenci  los pozos de agua. Con esta simple representaci on el mapa trasmit  un claro mensaje visual al conectar incidencia con concentraci on, la mayor a de las v ctimas se concentraban en torno a la bomba de agua de Broad Street.

En estudios epidemiol gicos, el empleo de mapas para representar la distribuci on geogr fica de enfermedades o decesos permite generar hip tesis acerca de las posibles causas en la variaci on del riesgo e identificar regiones caracterizadas por un riesgo alto o bajo. Sin embargo, no debe perderse de vista que, detr s del uso de mapas, el an alisis del fen omeno epidemiol gico est  directamente relacionado con el ajuste del modelo matem tico, tarea que desde el contexto estad stico frecuentemente resulta complicada.

Para ejemplificar la problem atica anterior, consideremos el caso en donde el total de casos observados de una enfermedad rara es asociado al  rea en la cual se llev  el conteo de casos. Su contexto es el siguiente: la regi n de estudio en donde hay prevalencia de la enfermedad est  o puede ser dividida en  $n$  subregiones y se asocia una variable aleatoria Poisson al n mero de casos que se ha registrado en cada subregi n. Entonces, el objetivo es obtener estimaciones estad sticas precisas y confiables de las tasas de riesgo local basadas en el n mero de casos observados dentro de las subregiones junto con la informaci n potencialmente relevante que est  disponible (este fen meno es conocido dentro del argot estad stico como problema de estimaci n en  reas peque as).

El valor m s com n para estimar la tasa de riesgo es la raz n de morbilidad estandarizada (con respecto a un factor de confusi n). Sin embargo, esta medida puede resultar seriamente enga osa, ya que es notablemente dependiente de la reducida magnitud de los valores involucrados en el c culo. Por tanto, los mapas de tasas estandarizadas, que se basan en la categorizaci n de los valores observados de las tasas de morbilidad estandarizadas para la construcci n del mapa, no permiten estimar la distribuci n de probabilidad subyacente del riesgo. Por otro lado, los mapas de significancia estad stica tienen la ventaja de proporcionar estimadores de los par metros de inter s pero plantean dos problemas. El primero, la variaci n observada en los eventos excede la variaci n esperada por la inferencia Poisson (fen meno conocido como sobredispersi n). El segundo, la mayor a de los modelos no toma en cuenta ning n patr n espacial en la enfermedad ([Molli , 1996](#)).

Haciendo hincapi  en la naturaleza del problema, las soluciones basadas en dise os cl sicos para este problema a menudo no son viables, ya que los tama os de las muestras locales de cada subregi n necesarios para los niveles deseados de precisi n estad stica no est n disponibles o son inalcanzables. En contraste, enfoques basados en modelos que asumen correlaci n positiva entre las observaciones y un mecanismo de “pedir prestada fuerza” (como considerar m s informaci n de regiones vecinas que de regiones m s lejanas, o comportamientos espaciales de otro

fen omeno aleatorio registrado en la regi on) permiten la suavizaci on de las tasas extremas ocasionadas por los tama os de muestra peque os. Estos enfoques se remontan al trabajo seminal para la estimaci on en  reas peque as de [Fay y Herriot \(1979\)](#) y frecuentemente son expresados como modelos lineales con efectos mixtos ([Waller y Carlin, 2010](#)). Aunque tradicionalmente estos modelos estadisticos se consideran m as flexibles en la captura de la dependencia espacial cuando un gran n umero de observaciones puede medirse en diferentes sitios, la desventaja de  stos se presenta en la carga computacional involucrada, as  como la limitaci on de los modelos disponibles para las estructuras de covarianza.

Estas consideraciones han llevado al desarrollo de modelos bajo el enfoque bayesiano, que proporcionen un mejor ajuste y que sean computacionalmente eficientes. A diferencia de los m etodos estadisticos cl asicos, los m etodos estadisticos bayesianos no se reducen a operar con la informaci on conseguida emp iricamente sino que la combinan con conocimiento *a priori* que posee el investigador, producto de estudios previos o de reflexiones y juicios racionales. Como resultado de tal integraci on, que se realiza por conducto del teorema de Bayes, se obtiene una visi on *a posteriori*. A pesar de que la Estadistica Bayesiana floreci o en el Siglo XIX, su desarrollo e implementaci on se ha visto truncado por la complejidad del c alculo matem atico involucrado. Afortunadamente, el avance en materia de c omputo, a partir de los a os 80 del siglo pasado, ha venido a resolver en gran parte este problema. En particular, el costo tiempo-m aquina (relativamente bajo) necesario para la implementaci on de m etodos de simulaci on estoc astica como las t ecnicas de Monte Carlo v ia Cadenas de Markov.

En el presente trabajo de tesis abordamos el problema de estimar la distribuci on espacial del riesgo de una enfermedad rara desde una perspectiva bayesiana. El esquema de trabajo es el siguiente. En el Cap itulo 2 se presentan los conceptos epidemiol ogicos b asicos involucrados en el problema de inter es, se resume de manera breve el paradigma bayesiano y se introducen las estructuras y herramientas estadisticas que facilitar an el proceso de inferencia. En el Cap itulo 3 se presenta

una revisi on de los principales enfoques para analizar la construcci on del mapa de una enfermedad rara cuando los datos son conteos de casos observados. En el Cap itulo 4 se construyen los mapas (i.e., se analiza la distribuci on espacial) de la mortalidad por c ancer de mama para el sur del Estado de M exico. Por  ultimo, en el Cap itulo 5 se presentan algunas conclusiones generales respecto a los modelos empleados en el cap itulo anterior.



# Capítulo 2

## Preliminares

El propósito de este capítulo es brindar al lector un bosquejo de los elementos teóricos involucrados en el tema principal de esta tesis. En la Sección 2.1 se presentan los conceptos epidemiológicos básicos necesarios para describir e interpretar los modelos estudiados en este trabajo. Debido a que se trabajará bajo el enfoque bayesiano, en la Sección 2.2 se resume de manera breve el paradigma bayesiano y se introducen las herramientas estadísticas que facilitarán el proceso de inferencia. El desarrollo de dicha sección está basado en [Gutiérrez-Peña \(1997, 1998\)](#), así como en las notas de cursos del mismo autor. Además, en la Sección 2.3 se introduce metodología estadística para analizar datos espaciales, pues al fenómeno de interés también se le puede asociar una estructura de dependencia espacial.

### 2.1. Epidemiología

La Epidemiología surgió de los esfuerzos por comprender la naturaleza de los fenómenos infecciosos colectivos que aparecieron de forma inexplicable, conocidos como plagas, pestes o epidemias. Ya en el siglo XX los estudios epidemiológicos se extendieron a las enfermedades y problemas de salud en general, de donde se propone la definición técnica de Epidemiología como la disciplina que estudia la

enfermedad como un fenómeno colectivo en poblaciones humanas.

La palabra epidemiología, que proviene de los términos griegos *e*pi (encima), *demo*s (pueblo) y *logos* (estudio), etimológicamente significa el estudio de lo que “está sobre las poblaciones”. Hoy en día podemos definir a la Epidemiología como la disciplina científica que estudia la distribución, la frecuencia, los factores determinantes, las predicciones y el control de los factores relacionados con la salud y con las distintas enfermedades existentes en poblaciones humanas definidas (López-Moreno *et al.*, 2000).

### 2.1.1. Conceptos básicos

Los principales objetivos de la investigación epidemiológica son, por un lado, describir la distribución de las enfermedades y, por otro, la caracterización de las leyes que gobiernan o influyen en dicha distribución. Lo último se logra a partir de proponer una o varias hipótesis etiológicas que posteriormente son sometidas a contrastación empírica.

Como lo discuten Moreno-Altamirano *et al.* (2000), un rasgo característico de la contrastación es que las relaciones causales postuladas entre las variables se describen en términos probabilísticos; es decir, se trata de establecer si la probabilidad de que un evento ocurra se debe precisamente a los factores que se sospecha intervienen en su génesis y no al azar.

Para llevar a cabo dicho proceso, se construyen medidas basadas en operaciones aritméticas simples, conocidas como *razones* y *tasas*. Las tasas expresan la dinámica de un suceso en una población a lo largo del tiempo. Se pueden definir como la magnitud de cambio de una variable (enfermedad o muerte) por unidad de cambio de otra (usualmente el tiempo) en relación con el tamaño de la población que se encuentra en riesgo de experimentar el suceso. Las razones pueden definirse

como magnitudes que expresan la relaci on aritm tica existente entre dos eventos en una misma poblaci n, o un solo evento en dos poblaciones.

Las medidas en Epidemiolog a se refieren a la medici n del n mero de individuos que enferman e individuos que fallecen debido a una enfermedad en un espacio y tiempo determinado, definidos como *morbilidad* y *mortalidad*, respectivamente. Las medidas m s frecuentes son la *prevalencia* y la *incidencia*. La prevalencia se refiere al n mero de individuos que, en relaci n con la poblaci n total, padecen la enfermedad en un momento espec fico. La incidencia, por su parte, expresa el volumen de casos nuevos que aparecen en un periodo definido. Es decir, expresa la probabilidad con la que los individuos de una poblaci n desarrollar n la enfermedad en el periodo.

### **2.1.2. Estandarizaci n por edad**

Los epidemi logos siempre tienen muy presente la diversidad en la poblaci n. Literalmente, toda poblaci n grande es heterog nea con respecto a aspectos sociodemogr ficos (edad, g nero, educaci n), geogr ficos, gen ticos y con respecto a otros innumerables atributos personales y factores ambientales relacionados con la salud. Una poblaci n puede ser concebida como una composici n de diversos subgrupos mutuamente ajenos (en  ltima instancia, subgrupos de tama o uno, es decir, individuos, pero las medidas epidemiol gicas no tienen sentido a ese nivel), conocidos en la literatura epidemiol gica como estratos.

Como lo plantean en [Schoenbach \(1999\)](#), las tablas que contienen medidas por grupos no se interpretan f cilmente por la gran cantidad de datos que suelen contener; esto explica la necesidad de buscar medidas de resumen globales que sean pr cticas para el an lisis de la poblaci n de estudio. Cualquier medida general refleja el valor de esa medici n para cada uno de los subgrupos que componen la

poblaci on. Una medida global que no tiene expl icitamente en cuenta la composici on de la poblaci on se denomina *cruda*. Su valor ser a el promedio de los valores para los subgrupos individuales ponderados por sus tama os relativos (por ejemplo, la tasa de mortalidad cruda de una poblaci on es el promedio ponderado de las tasas de mortalidad de los grupos de edad). Cuanto m as grande el subgrupo, mayor influencia tendr a en la medida cruda.

La tasa cruda es la forma de resumen m as sencilla y directa de la experiencia poblacional con respecto a una enfermedad. En el caso de la mortalidad (la cual est a fuertemente relacionada con la edad) las tasas espec ficas por edad ser an muy distintas una de otra. El resumen que produce la tasa cruda oculta la heterogeneidad de las tasas de mortalidad espec fica por grupo de edad. Esto se exhibe cuando se comparan las tasas entre poblaciones, porque si las poblaciones difieren en composici on, por lo menos una parte de lo observado puede ser atribuible a estas diferencias. Por otra parte, operar con tasas espec ficas por grupos de edad, es una alternativa para contemplar esta realidad que no permite un juicio sint tico. Por ejemplo, si se cotejan dos unidades, tendr an que establecerse comparaciones entre tantos pares de valores como grupos de edad se consideren, y las diferencias entre los valores de cada par podr an ser de distinta magnitud y signo. Entonces, lo m as conveniente es tener una medida resumen que permitiera una comparaci on global “justa” que no considere las diferencias entre los subgrupos generadas por los factores que influyen en las medidas de inter es, pero que no son el foco de atenci on, como lo es la edad en las tasas de mortalidad. Dicho mecanismo es conocido como *ajuste*.

La *estandarizaci on* es un m etodo de ajuste basado en promedios ponderados en el que los pesos son seleccionados de manera que presenten una base para la comparaci on, generalmente con respecto a una poblaci on “est andar”. La mayor parte de los textos de epidemiolog a presentan el tema de estandarizaci on en relaci on al ajuste por edad. Esta tendencia no es coincidencia, dado que todos los eventos m orbidos ocurren con distinta frecuencia en los diferentes grupos de edad.

De modo que, con el objetivo de neutralizar el efecto de las distintas composiciones por edad en el c alculo de las tasas de morbilidad bruta, se pueden utilizar dos m etodos bien conocidos: *m etodo directo* y *m etodo indirecto*. Para facilitar la conceptualizaci on algebraica de ambos m etodos, se introduce la siguiente notaci on y se trabajar a bajo el supuesto de que la poblaci on de estudio se divide en  $m$  estratos:

$n_k$  = n umero de personas en el  $k$ - esimo estrato de la poblaci on de estudio

$d_k$  = n umero de enfermos en el  $k$ - esimo estrato de la poblaci on de estudio

$r_k$  = tasa de morbilidad en el  $k$ - esimo estrato de la poblaci on de estudio

$w_k$  = peso del  $k$ - esimo estrato de la poblaci on de estudio

$N_k$  = n umero de personas en el  $k$ - esimo estrato de la poblaci on est andar

$D_k$  = n umero de enfermos en el  $k$ - esimo estrato de la poblaci on est andar

$R_k$  = tasa de morbilidad en el  $k$ - esimo estrato de la poblaci on est andar

$W_k$  = peso del  $k$ - esimo estrato de la poblaci on est andar

Por un lado, en el m etodo directo, las tasas espec ificas por estrato de la poblaci on de estudio se aplican a la distribuci on por estrato de la poblaci on est andar. En otras palabras, la estandarizaci on directa aplica el mismo conjunto de pesos a las tasas espec ificas por estrato. Es decir, la tasa directamente estandarizada es igual a

$$\sum_{k=1}^m r_k W_k.$$

Este m etodo garantiza la comparabilidad de las tasas para distintas poblaciones, pero para que las comparaciones tengan sentido deben haber n umeros suficientemente grandes en todos los estratos importantes. Si no es as ı, las estimaciones de las tasas espec ificas por estrato ser an demasiado imprecisas. Adem as, est a sujeto al supuesto de que la composici on por estratos es similar.

Por otro lado, cuando los n umeros por estrato son peque os, como suele ocurrir

en las enfermedades raras, se recurre a un procedimiento indirecto y se calcula una “raz on estandarizada”. En la estandarizaci on indirecta, contrario a la directa, la poblaci on est andar provee las tasas y la poblaci on de estudio provee los pesos. Se puede concebir a la estandarizaci on indirecta como el tomar el n umero de eventos de la poblaci on de estudio y compararlos con un n umero “esperado” de eventos en la poblaci on de estudio si su experiencia fuera la misma que para la poblaci on est andar ( $\sum_{k=1}^m R_k n_k$ ). La raz on de casos observados a esperados se denomina “Raz on de Morbilidad Estandarizada” (SMR, por sus siglas en ingl es), y ella, m as que las tasas estandarizadas, es el producto habitual de la estandarizaci on indirecta.

$$\begin{aligned} SMR &= \frac{\sum_{k=1}^m r_k n_k}{\sum_{k=1}^m R_k n_k} \\ &= \frac{\sum_{k=1}^m r_k w_k}{\sum_{k=1}^m R_k w_k} \end{aligned}$$

La tasa indirectamente estandarizada, llam emosle  $R_{ie}$ , puede ser obtenida de una SMR a trav es de la siguiente relaci on:

$$R_{ie} = SMR \times R,$$

donde  $R$  es la tasa cruda de la poblaci on est andar. La l ogica de esta relaci on es la SMR produce una comparaci on estandarizada de la experiencia de morbilidad en la poblaci on de estudio con la experiencia en la poblaci on est andar. Aunque se ha resuelto el problema de evitar ocultar la configuraci on de los estratos, la  unica comparaci on permitida es entre la poblaci on de estudio y la est andar.

Entre estas dos variantes, con frecuencia se prefiere la indirecta, en buena medida porque permiten una interpretaci on  gil y sencilla: por ejemplo, una SMR mayor que uno para una unidad geogr fica dada, significa que la morbilidad para esa unidad es mayor que la que le corresponder a si su patr on de morbilidad fuese similar al de la poblaci on de referencia. Sin embargo, este m etodo ha sido a su

vez cuestionado debido a que las SMR pueden resultar notablemente dependientes de los tama os poblacionales, lo cual implica un impacto sobre la variabilidad de las estimaciones. Este fen omeno ocurre con m as frecuencia cuando los datos son escasos, donde se pueden producir estimaciones extremas para las SMR en cada regi on gracias a la reducida magnitud de los valores involucrados en el c alculo.

## 2.2. El paradigma Bayesiano

Quiz as la mejor manera de definir una disciplina sea en t erminos de su objeto de estudio. En el caso de la estad stica, podemos decir que su objeto de estudio son los fen omenos aleatorios; es decir, aquellos que no se pueden predecir con certeza. Para describir adecuadamente la incertidumbre asociada a este tipo de fen omenos es necesario echar mano de la teor a de la probabilidad. Aunque la definici n de medida de probabilidad es una, existen varias interpretaciones de la probabilidad:

- *Cl sica*

Supone que el experimento aleatorio produce resultados igualmente veros imiles y mutuamente excluyentes y propone como medida de probabilidad el cociente entre los casos favorables y los casos totales.

- *Frecuentista*

Supone que un experimento aleatorio puede ser repetido un n umero infinito de veces bajo condiciones similares y propone como medida de probabilidad la proporci n de veces que ocurri  el evento de inter es.

- *Subjetiva*

Es simplemente una medida de la incertidumbre, asociada a uno de los posibles resultados del experimento, asignada por un decisor. En otras palabras, es un juicio personal sobre la verosimilitud de que ocurra un resultado.

Es com n concebir a la Estad stica (Frecuentista) como un conjunto de reglas, m etodos y algoritmos. Particularmente entre los usuarios, se suele visualizar a la

Estadística como una basta colecci on de algoritmos cuyo empleo es apropiado en forma casuística. Cabe mencionar que, aunque cada uno tiene sus propios m eritos y ventajas, no necesariamente constituyen una teor a coherente debido a que cada m etodo est a disenado para satisfacer criterios espec ficos y a menudo carecen de la flexibilidad necesaria para adaptarse a problemas distintos a aquellos para los cuales fueron propuestos. M as a un, tanto la bibliograf a general como los programas docentes de estadística suelen eludir toda menci on a las contradicciones inherentes a los m etodos cl asicos de inferencia y al intenso debate desarrollado durante casi todo el Siglo XX sobre el tema. Sin embargo, en la d ecada de los anos 50 Leonard J. Savage public o su libro *The Foundations of Statistics*, en donde se construy o una teor a axiom tica de la Inferencia Estadística a trav es de la Teor a de la Decisi on, ahora conocida como Estadística Bayesiana.

### 2.2.1. Inferencia Bayesiana

A diferencia de lo que ocurre con los m etodos cl asicos, los m etodos bayesianos permiten incorporar de manera natural conocimiento externo (basado en creencias personales, opini on de expertos y datos previos acerca del fen meno de inter es) a la realizaci on del experimento sobre las cantidades desconocidas de inter es,  $\theta$ . Esta informaci on *a priori* se cuantifica en t erminos de una distribuci on de probabilidad que llamaremos distribuci on inicial. Esta conceptualizaci on permite, de forma impl cita, tratar a  $\theta$  como objeto aleatorio. Por ende, extender el proceso de inferencia a eventos que no podemos medir directamente como observaciones futuras.

El mecanismo para hacer inferencias bajo el enfoque bayesiano es el siguiente:

- *Proposici on de una distribuci on inicial,  $p(\theta)$*

A pesar de la pol mica que existe por la subjetividad a la que est a sujeta la informaci on apriori, siempre debe asignarse una distribuci on inicial consistente con nuestra informaci on a priori. Tambi en debemos considerar que, en caso de carecer de informaci on inicial acerca del par metro  $\theta$ , debemos

ser capaces de reflejar esta ignorancia. Esto se hace a trav es de las llamadas distribuciones iniciales no informativas.

■ *Especificaci on de un modelo*

Con el fin de recolectar la informaci on que los datos aportan, se ajusta un modelo estoc astico para tratar de describir las caracter sticas del fen omeno que gener  los datos muestrales,  $\mathbf{x}$ . La cuantificaci on de la muestra bajo el modelo sugerido,  $p(\mathbf{x}|\boldsymbol{\theta})$ , es conocido como la funci on de verosimilitud.

■ *Distribuci on final*

Una vez obtenida una muestra  $\mathbf{x}$ , podemos condicionar en su valor observado. Por lo tanto, nuestro conocimiento acerca de  $\boldsymbol{\theta}$  queda descrito por medio de la distribuci on final  $p(\boldsymbol{\theta}|\mathbf{x})$ , obtenida a trav es del Teorema de Bayes:

$$p(\boldsymbol{\theta}|\mathbf{x}) = \frac{p(\boldsymbol{\theta}) p(\mathbf{x}|\boldsymbol{\theta})}{\int p(\boldsymbol{\theta}) p(\mathbf{x}|\boldsymbol{\theta}) d\boldsymbol{\theta}}$$

El denominador  $p(\mathbf{x}) = \int p(\boldsymbol{\theta}) p(\mathbf{x}|\boldsymbol{\theta})$ , no depende de  $\boldsymbol{\theta}$ , por lo que es com un escribir

$$p(\boldsymbol{\theta}|\mathbf{x}) \propto p(\boldsymbol{\theta}) p(\mathbf{x}|\boldsymbol{\theta}).$$

Entonces, dicho mecanismo es una herramienta capaz de, a modo de caja negra, combinar de forma coherente la informaci on a priori con informaci on muestral para producir informaci on *a posteriori* en la cual se basar  el investigador para dar sus conclusiones. Por esta raz on, el Teorema de Bayes se interpreta como un proceso de aprendizaje y de actualizaci on de conocimientos sobre el fen omeno de inter es. Adem as, este artificio no s olo combina las dos fuentes de informaci on sino que reconoce la fuerza de cada curva. Por tanto, la distribuci on final estar  m as influenciada por la distribuci on m as informativa.

Habiendo especificado la distribuci on final  $p(\boldsymbol{\theta}|\mathbf{x})$ , se puede presumir que hemos resuelto el problema de inferencia; ya que toda la informaci on disponible sobre el valor de  $\boldsymbol{\theta}$  queda contenida en la distribuci on final. Aunque a n es necesario

resumir esta informaci on para poder interpretar los resultados y responder las preguntas de inter es. Esto se puede hacer a trav es de procedimientos de estimaci on, prueba de hip otesis o pron osticos por ejemplificar algunos casos. Sin embargo, en la pr actica, el c alculo de la distribuci on final puede ser un asunto complicado, especialmente si la dimensi on del par ametro de inter es es grande. Para resolver dicho problema se utilizar an t ecnicas computacionales relativamente sofisticadas.

### 2.2.2. Monte Carlo v ia Cadenas de Markov

Com unmente, para proporcionar una soluci on al problema planteado, se recurre al c alculo de ciertas caracter sticas de la distribuci on final del par ametro de inter es que permitan resumir de alguna forma la informaci on contenida en ella. En muchos casos estos “resumenes inferenciales” se reducen a integrales. Por ejemplo, para pasar de una distribuci on conjunta a una colecci on de distribuciones y momentos marginales que sean  utiles para hacer inferencias sobre subconjuntos de par ametros. Desafortunadamente, en la pr actica, la mayor ia de las implementaciones con m etodos bayesianos no se pueden llevar a cabo de forma anal tica. Alternativamente, se puede hacer uso de m etodos computacionales para aproximar el valor de las integrales. Dentro de los m as usados se encuentra el m etodo de Monte Carlo.

La idea b asica del m etodo consiste en escribir la integral requerida de una funci on arbitraria,  $h$ ,

$$I = \int_{\Theta} h(\theta) d\theta,$$

como valor esperado de una funci on con respecto a alguna distribuci on de probabilidad, lo cual sugiere una soluci on estad stica al problema de integraci on. Escribamos a  $I$  como un valor esperado

$$I = \int_{\Theta} \frac{h(\theta)}{s(\theta)} s(\theta) d\theta = E_{s(\theta)} \left( \frac{h(\theta)}{s(\theta)} \right),$$

donde  $s(\theta)$  es una funci3n de probabilidad f acil de simular, conocida como la distribuci3n de muestreo por importancia. Despu3s se genera una muestra de  $s(\theta)$ ,  $\theta^{(1)}, \theta^{(2)}, \dots, \theta^{(M)}$ , con  $M$  grande. Finalmente se aproxima  $I$  a partir de la media muestral

$$\hat{I} = \frac{1}{M} \sum_{j=1}^M \frac{h(\theta^{(j)})}{s(\theta^{(j)})}.$$

El m3todo de Monte Carlo permite calcular algunos res3menes inferenciales elementales tales como valores esperados de ciertas funciones del par ametro respecto a la distribuci3n final. Es posible, adem as, aprovechar la muestra simulada con el fin de aproximar cualquier otra caracter istica de inter3s de la distribuci3n final. Sin embargo, la precisi3n de este estimador depende tanto del tama o de la muestra, como de la forma de la distribuci3n de muestreo por importancia. M as a3n, no siempre es posible encontrar una distribuci3n adecuada.

Las t3cnicas de Monte Carlo v ia Cadenas de Markov (MCMC, por sus siglas en ingl3s) permiten generar, de manera iterativa, observaciones de distribuciones que dif cilmente podr an simularse usando m3todos directos. La idea b asica es simple (basada en el Teorema 2.1), tenemos que construir una cadena de Markov homog3nea, irreducible y aperi3dica que sea f acil de simular y cuya distribuci3n de equilibrio corresponda a la distribuci3n final que nos interesa. El lector puede encontrar todos los detalles de esta teor a en [Meyn y Tweedie \(2009\)](#).

**Teorema 2.1.** *Sea  $\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}, \dots$  una cadena de Markov homog3nea, irreducible y aperi3dica, con espacio de estados  $\Theta$  y distribuci3n de equilibrio  $p(\boldsymbol{\theta}|\mathbf{x})$ . Entonces,*

$$(i) \quad \boldsymbol{\theta}^{(t)} \xrightarrow[t \rightarrow \infty]{d} \boldsymbol{\theta}, \quad \boldsymbol{\theta} \sim p(\boldsymbol{\theta}|\mathbf{x});$$

$$(ii) \quad \frac{1}{N} \sum_{i=1}^N g(\boldsymbol{\theta}^{(i)}) \xrightarrow[t \rightarrow \infty]{} \mathbb{E}(g(\boldsymbol{\theta})|\mathbf{x}).$$

El algoritmo b asico propuesto en [Metropolis \*et al.\* \(1953\)](#) fue generalizado por [Hastings \(1970\)](#). Este algoritmo es conocido como Metropolis-Hastings y constituye la versi3n m as general de la familia de algoritmos MCMC. M as tarde, [Geman](#)

y Geman (1984) elaboran un m etodo de simulaci on, que tambi en genera una cadena de Markov y que, tras ser enaltecido en Gelfand y Smith (1990), pasa a ser conocido en la literatura como muestreo de Gibbs. A continuaci on, se describen ambos algoritmos.

## Metropolis-Hastings

Este m etodo permite generar el siguiente estado de la cadena de una distribuci on objetivo,  $\pi$ . La invarianza de la distribuci on objetivo se obtiene por medio de un paso de aceptaci on o rechazo. Supongamos que la cadena se encuentra actualmente en  $\psi$ . Entonces el valor propuesto es  $\tilde{\psi}$ , que se gener o por medio de la distribuci on propuesta  $q(\cdot|\psi)$ . El valor propuesto  $\tilde{\psi}$  se acepta como un nuevo estado de la cadena con probabilidad

$$\alpha(\psi, \tilde{\psi}) = \min \left\{ 1, \frac{\pi(\tilde{\psi})q(\psi|\tilde{\psi})}{\pi(\psi)q(\tilde{\psi}|\psi)} \right\}. \quad (2.1)$$

Si el valor propuesto se rechaza, la cadena permanece en el estado  $\psi$ . El siguiente pseudo-algoritmo describe los pasos a seguir.

---

### Algoritmo 1 Metropolis-Hastings

---

0. Inicia: fijamos  $\psi^{(0)}$ .
  1. Para  $j = 1, \dots, N$ :
    - 1.1) genera  $\tilde{\psi}^{(j)}$  de  $q(\cdot|\psi^{(j-1)})$ ;
    - 1.2) calcula  $\alpha = \alpha(\psi^{(j-1)}, \tilde{\psi}^{(j)})$  de acuerdo a (2.1);
    - 1.3) genera una variable aleatoria independiente  $U_j \sim \text{Bernoulli}(\alpha)$ ;
    - 1.4) si  $U_j = 1$  hacemos  $\psi^{(j)} = \tilde{\psi}^{(j)}$ , en otro caso hacemos  $\psi^{(j)} = \psi^{(j-1)}$ .
- 

## Muestreo de Gibbs

Supongamos que el par ametro es multidimensional,  $\psi = (\psi_1, \psi_2, \dots, \psi_k)$ , y que existen todas las distribuciones condicionales completas de la distribuci on objetivo,  $\pi(\psi_i|\psi_{(-i)})$ , donde  $\psi_{(-i)} = (\psi_1, \dots, \psi_{i-1}, \psi_{i+1}, \dots, \psi_k)$  es el vector que se obtiene de eliminar la  $i$ - esima entrada,  $i = 1, \dots, k$ . El muestreo de Gibbs inicia en un valor

arbitrario  $\psi^{(0)} = (\psi_1^{(0)}, \dots, \psi_k^{(0)})$  y actualiza un componente a la vez,  $\psi_i$ , a trav es de la distribuci on condicional  $\pi(\psi_i | \psi_{(-i)})$ , como se describe en el siguiente pseudo-algoritmo. Un punto importante es que este algoritmo funciona incluso cuando uno o m as componentes  $\psi_i$  por s  mismos son multidimensionales. En cuyo caso, el algoritmo funciona por bloques de los componentes de  $\psi$ . El muestreo de Gibbs puede verse como un caso particular del algoritmo de Metropolis-Hastings.

---

**Algoritmo 2** Muestreo de Gibbs

---

0. Inicia: fijamos  $\psi^{(0)} = (\psi_1^{(0)}, \dots, \psi_k^{(0)})$ .
  1. Para  $j = 1, \dots, N$ :
    - 1.1) genera  $\psi_1^{(j)}$  de  $\pi(\psi_1 | \psi_2 = \psi_2^{(j-1)}, \dots, \psi_k = \psi_k^{(j-1)})$ ;
    - 1.2) genera  $\psi_2^{(j)}$  de  $\pi(\psi_2 | \psi_1 = \psi_1^{(j)}, \psi_3 = \psi_3^{(j-1)}, \dots, \psi_k = \psi_k^{(j-1)})$ ;
    - $\vdots$
    - 1.k) genera  $\psi_k^{(j)}$  de  $\pi(\psi_k | \psi_1 = \psi_1^{(j)}, \dots, \psi_{k-1} = \psi_{k-1}^{(j)})$ .
- 

El algoritmo de Metropolis-Hastings y el muestreo de Gibbs, forman los dos esquemas b asicos de la metodolog a MCMC. A partir de  stos se han creado variantes aprovechando sus caracter sticas y ventajas. Supongamos que dentro de un muestreo de Gibbs queremos simular de la distribuci on condicional  $\pi(\psi_i | \psi_{(-i)})$ , pero  sta no tiene una forma conocida o es demasiado complicado simularla. En este caso, en lugar de generar  $\psi_i$  a partir de  $\pi(\psi_i | \psi_{(-i)})$ , podemos hacerlo por medio de un paso del algoritmo de Metropolis-Hastings sin que la distribuci on de la cadena de Markov deje de ser invariante.

Cabe mencionar que, por la propiedades de la cadena, se garantiza que la cadena converger  sin importar el valor inicial. Sin embargo, no es f cil determinar en qu  momento la cadena ha convergido, por lo que usualmente se utilizan m todos emp ricos para verificarlo. Por otro lado, podemos usar las funciones de autocorrelaci on para determinar si la muestra que generamos presenta una autocorrelaci on suficientemente baja como para no afectar la precisi on de las aproximaciones. Con frecuencia se omite un ‘‘periodo de calentamiento’’ (primeros valores de la cadena)

con la finalidad de permitir que la cadena salga de una primera fase de inestabilidad y as ı aplicar las propiedades erg odicas que ofrece el Teorema 2.1.

## RJMCMC

Una de las extensiones m as importantes del algoritmo Metropolis-Hastings fue introducida en el trabajo seminal de Green (1995), en el cual se propus o un mecanismo para generar observaciones de una distribuci on objetivo cuyo espacio parametral puede ser escrito como la uni on de subespacios, cada uno no necesariamente de la misma dimensi on. Este algoritmo fue denominado MCMC con saltos reversibles (RJMCMC, por sus siglas en ingl es) y su conceptualizaci on matem tica es la siguiente: sea  $\mathcal{M} = \{M_1, M_2, \dots\}$  una colecci on numerable de posibles modelos que puedan ser ajustados y  $\boldsymbol{\theta}_k \in \mathbb{R}^{n_k}$  el espacio parametral del modelo  $M_k$ ,  $k = 1, 2, \dots$ . El algoritmo RJMCMC muestrea en todos los diferentes modelos como sigue. Primero, necesitamos una medida de probabilidad sobre  $\mathcal{M}$ , denotada por  $p$ , y una probabilidad de saltar del modelo  $M_i$  al modelo  $M_j$ , denotada como  $p(i, j)$ ,  $i, j \in \{1, 2, \dots\}$ . Segundo, sea  $\pi_k$  la distribuci on final bajo  $M_k$ . Finalmente, sea  $q_{ij}$  una densidad propuesta para generar  $\mathbf{u}_i$  y  $g_{ij}$  una funci on predeterminada que mapea uno a uno los espacios  $(\boldsymbol{\theta}_i, \mathbf{u}_i)$  y  $(\boldsymbol{\theta}_j, \mathbf{u}_j)$  de tal forma que  $\dim(\boldsymbol{\theta}_i) + \dim(\mathbf{u}_i) = \dim(\boldsymbol{\theta}_j) + \dim(\mathbf{u}_j)$ . El siguiente pseudo-c odigo describe la metodolog ıa para llevar a cabo su implementaci on.

---

### Algoritmo 3 RJMCMC

---

0. Inicia: Eligir  $j \in \{1, 2, \dots\}$  y generar  $\mathbf{u}_j$ .
1. Proponer un nuevo modelo con probabilidad  $p(i, j)$ .
2. Generar  $\mathbf{u}_i$  de la densidad propuesta  $q_{ij}(\mathbf{u}|\boldsymbol{\theta}_i)$ .
3. Hacer  $(\boldsymbol{\theta}_j, \mathbf{u}_j) = g_{ij}(\boldsymbol{\theta}_i, \mathbf{u}_i)$ .
4. Aceptar el movimiento propuesto con probabilidad

$$\min \left\{ 1, \frac{\pi_j(\boldsymbol{\theta}_j|\mathbf{y})p(j, i)q_{ji}(\mathbf{u}_j|\boldsymbol{\theta}_j)}{\pi_i(\boldsymbol{\theta}_i|\mathbf{y})p(i, j)q_{ij}(\mathbf{u}_i|\boldsymbol{\theta}_i)} \left| \frac{\partial g_{ij}(\boldsymbol{\theta}_i, \mathbf{u}_i)}{\partial(\boldsymbol{\theta}_i, \mathbf{u}_i)} \right| \right\}$$


---

A pesar de la enorme utilidad de este m etodo, a un no ha sido tan ampliamente implementado, debido a la complejidad te orica y computacional que conlleva. De hecho, las complicaciones se presentan en la elicitaci on de las densidades propuestas y las funciones de mapeo que garantizan los saltos entre los modelos.

### 2.2.3. Modelos Jer arquicos

Muchas aplicaciones estad isticas involucran colecciones de par ametros que pueden considerarse como relacionados o conectados entre s ı debido a la estructura del problema. Por ejemplo, en un estudio sobre la efectividad de un tratamiento, si los pacientes del hospital  $i$  tienen una probabilidad de supervivencia  $\theta_i$  entonces parece razonable suponer que las estimaciones de las probabilidades de supervivencia en  $k$  hospitales est an relacionadas entre s ı.

Una manera natural de modelar esta situaci on consiste en usar una distribuci on inicial bajo la cual  $\{\theta_i\}_{i=1}^k$  puedan verse como una muestra de una distribuci on poblacional com un, que a su vez depende de un hiperpar ametro desconocido  $\phi$ . Una caracter istica importante de este enfoque es que los datos observados,  $x_i$  ( $i = 1, \dots, k$ ), pueden utilizarse para estimar ciertos aspectos de la distribuci on poblacional de  $\{\theta_i\}_{i=1}^k$  a pesar de que los valores de  $\{\theta_i\}_{i=1}^k$  no son observables.

Es importante destacar que, en general, si el modelo tiene pocos par ametros no pueden ajustar adecuadamente conjuntos de datos relativamente grandes. Por el contrario, si tiene muchos par ametros entonces tender a a “sobreajustar” los datos, en el sentido de que ajustar a bien a los datos observados pero no necesariamente producir a buenas predicciones. En contraste, los modelos jer arquicos pueden tener suficientes par ametros para ajustar bien los datos, evitando problemas de sobreajuste al modelar la estructura de dependencia entre los par ametros a trav es de la distribuci on poblacional.

La flexibilidad en la modelizaci on que proporcionan los modelos jer arquicos bayesianos han contribuido a la soluci on de una gran variedad de problemas en las ciencias aplicadas. Los factores m as importantes que han facilitado su expansi on son el desarrollo de herramientas de simulaci on y el auge computacional, que permiten hacer inferencias sobre los par metros de inter es.

### 2.3. Estad stica Espacial

La creciente necesidad del uso de herramientas estad sticas para el an lisis de datos en las ciencias aplicadas, ha generado el desarrollo de nuevas metodolog as que, aunque son construidas a partir de los fundamentos probabil sticos, son espec ficas para cada  rea del saber. Por ejemplo, la estad stica espacial.

La estad stica espacial se puede definir como el conjunto de t cnicas aptas para el an lisis de datos que corresponden a la medici on de variables aleatorias en diversos sitios de una regi on geogr fica, cuyo objetivo principal es analizar la distribuci on espacial mediante la detecci on de patrones, procesos y relaciones espaciales entre los sitios de la regi on de estudio (dependencia espacial). En t rminos matem ticos, se puede decir que la estad stica espacial consiste en el an lisis de realizaciones de un campo aleatorio<sup>1</sup>,  $\{X_s : s \in D\}$ , donde  $s$  representa un sitio en la regi on de estudio  $D$  y  $X_s$  una variable aleatoria asociada a  $s$ .

Conceptualmente, la estad stica espacial se divide en tres  reas. La definici on de cada una de ellas est  determinada por las caracter sticas del fen meno de inter es en la regi on de estudio  $D$ . A continuaci on se describen las caracter sticas de cada  rea.

---

<sup>1</sup>Un proceso estoc stico es una colecci on de variables aleatorias  $\{X_t : t \in T\}$ , indexadas por el conjunto  $T$ , en donde las variables toman valores en un conjunto  $S$  (espacio de estados). Un campo aleatorio es un proceso estoc stico con conjunto de  ndices  $T \subset \mathbb{R}^d$ .

- *Geoestadística*

$D$  es un conjunto continuo y fijo. Por ejemplo, el problema de estimar los valores de precipitaci on en un pa s con base a los registros de diferentes estaciones meteorol gicas en un mes dado, puede ser tratado con esta metodolog a.

- *Rejillas*

$D$  es un conjunto discreto y fijo, donde los  ndices representan agregaciones espaciales (regulares o irregulares) de la regi on de inter s. Por ejemplo, estimar la tasa de mortalidad por estados de un pa s.

- *Patrones puntuales*

$D$  es un conjunto aleatorio que puede ser discreto o continuo. Por ejemplo, el epicentro de un temblor.

Existen diversas formas de inferir la distribuci on espacial subyacente de un fen meno aleatorio (v ase [Cressie \(1993\)](#) para una introducci on general a la metodolog a estadística para procesos espaciales). En general,  stas se clasifican en dos clases gen ricas, las cuales est n determinadas por la especificaci on de la medida de probabilidad del modelo estoc stico asociado, por medio de la distribuci on conjunta o las distribuciones condicionales. Formalmente, la especificaci on natural de cualquier proceso estoc stico es en t rminos de la distribuci on conjunta de las variables aleatorias involucradas, pues una vez definida la medida de probabilidad subyacente se puede inferir cualquier característica de inter s. No obstante, no se deben ignorar las dificultades que conlleva el proceso de proponer dicha medida. Por otro lado, a pesar de que las distribuciones condicionales determinan de forma  nica la distribuci on conjunta y la formulaci on de la distribuci on condicional goza de una interpretaci on intuitiva,  stos atributos se ven empa ados por una serie de desventajas. La m s preocupante, desde el punto de vista matem tico, es que no existe un proceso obvio para identificar la estructura de la distribuci on conjunta a partir de las distribuciones condicionales.

Sin embargo,  este no es el caso cuando se ajustan modelos a rejillas finitas. Usualmente, a una rejilla finita se le asocia un grafo simple<sup>2</sup>,  $G = (V, A)$ , donde el conjunto de v ertices  $V$  representa las regiones y el conjunto de aristas  $A$  representa las relaciones existentes entre cada par de v ertices. Despu es, extendiendo la propiedad de Markov al campo aleatorio  $V$  y usando el Teorema de Hammersley-Clifford, se estima la distribuci on espacial (medida de probabilidad) subyacente del proceso estoc astico. La teor a de los Campos Aleatorios de Markov (MRF, por sus siglas en ingl es) proporcionan una forma conveniente para modelar la dependencia espacial en t erminos de las probabilidades condicionales. El uso pr actico de los modelos de los MRF se debe en gran parte a la equivalencia entre la distribuci on de probabilidad y la distribuci on de Gibbs de un MRF, establecidas por [Hammersley y Clifford \(1971\)](#) y posteriormente desarrolladas por [Besag \(1974\)](#). A continuaci on, se introduce la conceptualizaci on matem atica de dicha teor a.

### 2.3.1. Campos Aleatorios de Markov y el Teorema de Hammersley-Clifford

Supongamos que estamos interesados en modelar un sistema finito de variables aleatorias,  $X_1, \dots, X_n$ , las cuales est an asociadas a sitios con etiquetas  $1, \dots, n$ , respectivamente. Denotemos por  $\mathbf{x} = (x_1, \dots, x_n)$  a una posible realizaci on de  $\mathbf{X} = (X_1, \dots, X_n)$  y  $\Omega$  al espacio muestral de todas las posibles realizaciones de  $\mathbf{X}$ . Es decir,  $\Omega = \{\mathbf{x} : p(\mathbf{x}) > 0\}$ . En este sentido, trabajaremos bajo el siguiente supuesto: si  $x_1, \dots, x_n$  pueden ocurrir individualmente en los sitios  $1, \dots, n$ , respectivamente, entonces pueden ocurrir de manera conjunta. Matem aticamente; si  $p(x_i) > 0 \quad \forall i \in \{1, \dots, n\}$ , entonces  $p(\mathbf{x}) > 0$ . Esta propiedad fue llamada *condici on de positividad* por Hammersley y Clifford. Ahora, considerando dos realizaciones,  $\mathbf{x}$  y  $\mathbf{y}$ , se puede obtener la distribuci on conjunta de  $\mathbf{x}$  por medio de las

---

<sup>2</sup>Un grafo  $G$  es un par ordenado  $G = (V, A)$ , donde  $V$  es un conjunto de v ertices o nodos y  $A$  es un conjunto de aristas o arcos que relacionan estos nodos. Un grafo es simple si a lo m as existe una arista uniendo dos v ertices cualesquiera.

distribuciones condicionales como sigue:

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

La ecuaci on anterior deja en claro las dos principales dificultades con respecto a la especificaci on de un sistema mediante su estructura de distribuciones condicionales. En primer lugar, el etiquetado arbitrario de los sitios individuales implica la existencia de muchas factorizaciones. Esto, a su vez, con el fin de lograr una estructura de probabilidad conjunta matematicamente consistente, implica la existencia de las severas restricciones a las formas funcionales disponibles de las distribuciones de probabilidad condicionales. En segundo lugar, tenemos que calcular la constante de normalizaci on extremadamente compleja  $p(\mathbf{y})$ . No obstante, el Teorema de Hammersley-Clifford evita estos problemas y permite construir esquemas espaciales v alidos. Comenzaremos dando una serie de definiciones y resultados para despu es enunciar dicho teorema.

**Definici on 2.2.** Sea  $\mathbf{x} = \{x_1, \dots, x_n\}$  a una realizaci on del sistema asociado a los sitios  $\{1, \dots, n\}$ .

1. Un sitio  $j \neq i$  se dice que es *vecino* del sitio  $i$  si y s olo si la distribuci on condicional  $p(x_i|x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n)$  depende funcionalmente de  $x_j$ .
2. La *vecindad* del sitio  $i$  es definida y denotada por  $N_i = \{j|j \text{ es vecino de } i\}$ .
3. Se dice que un campo aleatorio es de *Markov* si  $p(x_i|x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n) = p(x_i|x_j : j \in N_i) \quad i \in \{1, \dots, n\}$ .
4. Cualquier subconjunto de sitios que cumpla con la condici on: ‘‘cualquier sitio es vecino del resto de los sitios del subconjunto’’, es llamado *clique*.

Sin p erdida de generalidad, supongamos que el valor cero puede ocurrir en cada sitio. Este supuesto (hecho por razones puramente t ecnicas), bajo la condici on de positividad, asegura que  $\mathbf{0} \in \Omega$  y legitimized la definici on de

$$Q(\mathbf{x}) = \log(p(\mathbf{x})/p(\mathbf{0})) \quad \forall \mathbf{x} \in \Omega.$$

Entonces, conocer a  $Q(\cdot)$  (de aqu ı en adelante, nos referiremos a  $Q$  como la *funci on potencial*) es equivalente a conocer a  $p(\cdot)$ , pues  $p(\mathbf{x}) = \exp(Q(\mathbf{x})) / \int_{\mathbf{y} \in \Omega} \exp(Q(\mathbf{y}))$ . Hammersley y Clifford se plantearon la pregunta: dados los vecinos de cada sitio,  cu al es la forma m as general que puede tomar la funci on potencial de tal manera que se defina una estructura de probabilidad conjunta v alida para el sistema? Dado que

$$\begin{aligned} \exp(Q(\mathbf{x}) - Q(\mathbf{x}_i)) &= \frac{p(\mathbf{x})}{p(\mathbf{x}_i)} \\ &= \frac{p(x_i | x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n)}{p(0 | x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n)}, \end{aligned}$$

donde  $\mathbf{x}_i = (x_1, \dots, x_{i-1}, 0, x_{i+1}, \dots, x_n)$ . La soluci on a este problema inmediatamente da la forma m as general que puede ser tomada por la distribuci on de probabilidad condicional en cada sitio. Una prueba simple descansa en la observaci on de que, para cualquier distribuci on de probabilidad  $p(\mathbf{x})$  sujeta a las condiciones anteriores, existe una expansi on  unica de  $Q(\mathbf{x})$  tal que

$$Q(\mathbf{x}) = \sum_{i=1}^n x_i G_i(x_i) + \sum_{i=1}^n \sum_{j=1}^n x_i x_j G_{i,j}(x_i, x_j) + \dots + x_1 \dots x_n G_{1,\dots,n}(x_1, \dots, x_n),$$

donde  $G_{1\dots t}$  es una funci on que da las relaciones existentes entre las observaciones  $x_1, \dots, x_t$ . Entonces, el resultado de Hammersley y Clifford se puede enunciar de la siguiente manera:

**Teorema 2.3** (Hammersley-Clifford). *Supongamos que  $\mathbf{X}$  es un campo aleatorio de Markov tal que satisface la condici on de positividad. Entonces, la funci on potencial  $Q(\cdot)$  debe satisfacer la siguiente propiedad: si los sitios  $i, j, \dots, s$  no forman un clique, entonces  $G_{i,j,\dots,s}(\cdot) = 0$ , donde los cliques est an definidos por el conjunto de vecindades  $\{N_i : i = 1, \dots, n\}$ .*

*Demostraci on.* La prueba de este resultado se encuentra en Besag (1974). □

## Capítulo 3

# Modelos para Mapas de Mortalidad

Como lo sugiere [Silva-Aycaguer \*et al.\* \(2003\)](#), la representación mediante mapas de ciertos acontecimientos de salud relevantes (como la distribución de la mortalidad) proporciona una imagen más elocuente de su configuración espacial que la mera observación de datos organizados en tablas. En su vertiente descriptiva, este tipo de análisis permite detectar la existencia de patrones espaciales del fenómeno epidemiológico, y contribuye a cuantificar en términos absolutos y relativos la importancia de las diferencias presentes en una región (típicamente áreas producto de una división política). Es decir, permite identificar áreas con una tasa alta o baja con respecto a una medida estándar de una manera visual fácil de entender. En la vertiente exploratoria, ayuda a sugerir posibles explicaciones para las distribuciones identificadas y generar hipótesis acerca de las posibles causas en la variación de las tasas. En efecto, el estudio de la variación espacial del riesgo para cierta enfermedad puede permitir la formulación de hipótesis acerca de la relación entre las causas y lugar de ocurrencia, y revelar indicios para orientar futuros estudios de investigación que procuren explicaciones más refinadas.

En particular, el empleo de mapas para representar la distribuci3n geogr fica de enfermedades o decesos permite generar hip3tesis acerca de las posibles causas en la variaci3n del riesgo e identificar regiones caracterizadas por un riesgo alto o bajo. Por ejemplo, [Snow \(1955\)](#) formul3 la hip3tesis de la transmisi3n del c3lera por el agua contaminada en el barrio Soho de Londres, en donde un reciente brote epid3mico hab a matado a m s de 500 personas en un lapso de diez d as. Snow marc3 en el mapa del barrio las defunciones con l neas de color negro que se iban apilando unas sobre las otras a medida que el n mero de decesos aumentaba; adem s georreferenci3 los pozos de agua. Con esta simple representaci3n el mapa trasmit a un claro mensaje visual al conectar incidencia con concentraci3n, la mayor a de las v ctimas se concentraban en torno a la bomba de agua de Broad Street.

En este sentido, se debe tener muy presente el formato de los datos epidemiol3gicos disponibles. De hecho, la “geocodificaci3n” de los datos definir a naturalmente las diferentes estructuras espaciales disponibles para el an lisis de la distribuci3n espacial del riesgo. Cuando se registra un caso de ocurrencia, existe una dualidad con respecto a la disponibilidad de la localizaci3n exacta del lugar de ocurrencia, pues  sta puede o no estar en la base de datos en la que se basar a el an lisis. En el primer caso, lo m s apropiado es describir la distribuci3n espacial a trav s de modelos que consideren la estructura de un patr3n puntual. El lector puede encontrar una introducci3n sobre el tema en [Elliott \*et al.\* \(2000\)](#) y [Lawson \(2013\)](#). En el segundo caso, como primer paso se considera el total de casos observados y se asocia al  rea en la cual se observaron los casos. Despu s, se modela la distribuci3n espacial mediante el uso de un proceso de rejillas.

Considerando el formato de los datos que se analizar n en el Cap tulo 4, en este trabajo de tesis s3lo abordaremos los enfoques estad sticos m s usados para modelar datos espacialmente agrupados.

El surgimiento de datos asociados a un  rea “peque a” (tama os de muestra y/o poblaci n peque a) plantean problemas espec ficos de modelado y an lisis. Este escenario fue reconocido en la d cada de los 90 como una clase gen rica de problemas estad sticos donde existe un proceso espacial (latente) subyacente de inter s contaminado por un ruido observacional, y se denomina problema de estimaci n en  reas peque as. En estos modelos, tanto el an lisis como la inferencia son a nivel grupo, donde los grupos est n compuestos de individuos que viven en  reas geogr ficas definidas. Aunque en principio esta estructura genera una p rdida de informaci n espacial, la informaci n fidedigna disponible como tama o poblacional por estratos y medidas socio-econ micas proporcionan un contrapeso a dicha p rdida.

Dentro de este contexto, la manera m s sencilla de representar la variaci n espacial del riesgo de una enfermedad parte de suponer que la regi n geogr fica de inter s  $S$  est  o puede ser dividida en  $n$  subregiones  $\{S_1, S_2, \dots, S_n\}$  tales que:

1.  $S_i \neq \emptyset \quad \forall i = 1, 2, \dots, n.$
2.  $S_i \cap S_j = \emptyset \quad \forall i \neq j.$
3.  $\bigcup_{i=1}^n S_i = S.$

Y para las cuales es posible obtener los pares

$$(o_i, n_i) \quad i = 1, 2, \dots, n,$$

donde  $n_i$  es el n mero de personas en riesgo o expuestas y  $o_i$  es el n mero de individuos enfermos en la  $i$ - sima regi n. A la representaci n gr fica del patr n espacial de la enfermedad le llamaremos *mapeo de la enfermedad*<sup>1</sup>.

---

<sup>1</sup>El t rmino mapeo de la enfermedad, fue introducido por primera vez en [Clayton y Kaldor \(1987\)](#).

Desde el punto de vista estad stico, para modelar este fen meno se parte de suponer que el riesgo a nivel individual var a aleatoriamente dentro de cada regi n (no se agrupan espacialmente) y el riesgo asociado a cada regi n act a proporcional a los riesgos de cada estrato (el riesgo del estrato de una regi n es el producto del riesgo de la regi n y del riesgo del estrato). Es decir, el n mero de enfermos en la  $i$ - sima regi n,  $O_i$ , se puede ver como una variable aleatoria con distribuci n Binomial de par metros  $n_i$  y una probabilidad desconocida  $\pi_i$ , denotado como

$$O_i \sim Bin(n_i, \pi_i).$$

Sin embargo, bajo el supuesto de una enfermedad rara ( $n_i$  grande,  $\pi_i$  peque a), el modelo Binomial puede aproximarse por el modelo Poisson, es decir,

$$O_i \sim Poi(n_i \pi_i).$$

Recordemos que, en las enfermedades raras, se recurre a un proceso de estandarizaci n ya que las tasas crudas ocultan la heterogeneidad de la poblaci n. Debido a esto, se propone el uso de la tasa de mortalidad de la poblaci n de referencia con la que se estandariz  (llam mosle  $p > 0$ ), para reescribir el modelo como

$$O_i \sim Poi(e_i \lambda_i), \tag{3.1}$$

donde  $e_i$  es el n mero esperado de casos ( $e_i = n_i p$ ) y  $\lambda_i$  es el cociente de las tasas de mortalidad de la  $i$ - sima regi n de estudio y de la poblaci n de referencia ( $\lambda_i = \frac{\pi_i}{p}$ ), conocido como el *riesgo relativo* con respecto a  $p$ .

La elecci n usual para la probabilidad de referencia  $p$  es el riesgo global  $\frac{\sum o_i}{\sum n_i}$ , lo cual implica que  $\sum e_i = \sum o_i$ . Este es un caso simple de la llamada estandarizaci n interna, es decir, cuando  $p$  se deriva  nicamente de los datos obtenidos, mientras que en la llamada estandarizaci n externa la probabilidad de referencia

se obtiene de una fuente externa ajena a los datos. En el caso m as general, los datos son estratificados por factores confusores (uno de ellos es usualmente la edad). Supongamos que los datos est an ahora dados como  $(o_{ij}, n_{ij})$ , donde los sub ındices  $i = 1, 2, \dots, n$  denotan las regiones y  $j = 1, 2, \dots, m$  denota a los grupos de edad. Entonces  $O_{ij} \sim Poi(n_{ij}\pi_{ij})$  y, bajo el supuesto de no interacci on entre la regi on y la edad,  $\pi_{ij}$  puede ser escrito como el producto de la probabilidad de referencia  $p_j$  (relativa al grupo de edad  $j$ ) y el riesgo relativo  $\lambda_i$  en la regi on  $i$ . Bajo el supuesto de independencia condicional en las respuestas  $o_{ij}$ , el modelo puede escribirse ahora como  $O_i \sim Poi(e_i\lambda_i)$  donde  $O_i = \sum_j O_{ij}$  y  $e_i = \sum_j e_{ij} = \sum_j n_{ij}p_j$ . De aqu ı que el modelo se reduzca al caso sin factores de estratificaci on ([Knorr-Held y Becker, 2000](#)).

Dado que el objetivo principal de este trabajo es el an alisis de una enfermedad rara (el cual se abordar a como un problema de estimaci on en  reas peque nas), a continuaci on se presentan los principales enfoques para la soluci on del problema.

### 3.1. Enfoques tradicionales

Como lo discute [Molli e \(1996\)](#), en la literatura del mapeo de la enfermedad existen dos enfoques tradicionales. El primero consiste en considerar al cociente  $\frac{o_i}{e_i}$  como el riesgo local de la  $i$ - sima regi on, el cual resulta ser el estimador de m axima verosimilitud de  $\lambda_i$  y la SMR (con respecto al factor de confusi on edad). Despu es, simplemente se proponen puntos de corte para la categorizaci on de los valores de las SMR para la construcci on del mapa. Este m etodo ha sido cuestionado debido a varios inconvenientes; el principal se presenta cuando se calcula la SMR para  reas peque nas en las cuales los casos observados y esperados son escasos, las tasas brutas de mortalidad o morbilidad son muy vol tiles y su representaci on a trav es de un mapa exhibe una variaci on amplia y es poco indicativa de los patrones vigentes. Por tanto, se suelen producir estimaciones del riesgo relativo muy extremas (muy bajas o muy altas en relaci on con las dem as) que van ubic andose de manera

ca otica en el mapa hasta el punto de obstaculizar una interpretaci on epidemiol ogica  util. Otro problema es la dif cil interpretaci on de la SMR. Espec ficamente, la varianza de  $\frac{O_i}{e_i}$  es igual a  $\frac{\lambda_i}{n_i p}$ , y por lo tanto, inversamente proporcional al tama o de la poblaci on en la regi on  $i$ . De modo que, a medida que la poblaci on en riesgo sea menor, mayor ser  la variabilidad en la SMR y viceversa. Tambi en es posible encontrar algunas dificultades al calcular la SMR, por ejemplo cuando  $o_i = 0$ , el cual es un caso com n en enfermedades raras. El segundo enfoque, con el objetivo de obtener estimaciones estad sticas m s precisas y confiables de las tasas de riesgo local, se basa en modelos de significancia estad stica basados en el supuesto (3.1), los cuales tienen la ventaja de proporcionar estimadores de los par metros de inter s, pero plantea dos problemas. El primero, la variaci on observada en los eventos excede la esperada por la inferencia Poisson (sobredispersi on). En un  rea, la variaci on observada se debe en parte al muestreo Poisson pero tambi en a la variaci on extra-Poisson resultante de la variabilidad en la tasa de enfermedad dentro del  rea (la cual resulta de la heterogeneidad) y la dependencia espacial. De hecho, una proporci on desconocida de la variaci on observada puede ser causada por factores de riesgo no observables que var an geogr ficamente. El segundo, la mayor a de los modelos no toma en cuenta ning n patr n espacial en la enfermedad; por ejemplo, es usual en estos casos incorporar informaci on de regiones adyacentes a la de inter s como la tendencia a que regiones vecinas tengan tasas similares (dependencias locales).

En contraste, los enfoques basados en modelos que asumen correlaci on positiva entre las observaciones y un mecanismo de “pedir prestada fuerza” (como considerar m s informaci on de  reas vecinas que de  reas m s lejanas o comportamientos espaciales de otro fen meno aleatorio registrado en la regi on), permiten el suavizado de las tasas extremas ocasionadas por los tama os de muestra local peque os mediante la identificaci on de la heterogeneidad no observada que genera la variaci on muestral adicional. Frecuentemente estos enfoques son planteados como modelos lineales generalizados mixtos y se remontan al trabajo seminal de [Fay y](#)

Herriot (1979), quienes propusieron el uso de efectos aleatorios para agregar informaci n espacial (Waller y Carlin, 2010). Aunque tradicionalmente estos modelos geoestad sticos se consideran m s flexibles en la captura de la dependencia espacial cuando un gran n mero de observaciones puede medirse en diferentes sitios, la desventaja de estos se presenta en la carga computacional involucrada, as  como la limitaci n de los modelos disponibles para las estructuras de covarianza. Estas consideraciones han llevado a desarrollar modelos bayesianos que a su vez proporcionen un mejor ajuste al problema y que sean computacionalmente eficientes.

El modelo gen rico m s utilizado para la construcci n del mapeo de la enfermedad, planteado de una forma natural dentro del enfoque bayesiano, se basa en el siguiente modelo jer rquico de tres niveles:

$$\begin{aligned} O_i &\sim Poi(e_i \lambda_i), \quad i = 1, \dots, n \\ \log(\lambda_i) &= \mathbf{x}'_i \boldsymbol{\beta} + R_i \\ R_i &\sim p(\cdot | \boldsymbol{\theta}) \\ \boldsymbol{\theta} &\sim g(\cdot) \end{aligned}$$

donde  $\mathbf{x}'_i$  es el vector de valores correspondiente a la medici n de las covariables con coeficientes  $\boldsymbol{\beta}$  y  $R_i$  es el efecto aleatorio del riesgo relativo en la  $i$ - sima regi n,  $p(\cdot | \boldsymbol{\theta})$  es una distribuci n inicial apropiada para  $R_i$  con hiperpar metros  $\boldsymbol{\theta}$  y  $g$  es la distribuci n inicial de los hiperpar metros.

Como se ha mencionado, adem s de los datos observados, no siempre se cuenta con informaci n adicional disponible. Por ello, en lo que resta de este trabajo (basado en Best *et al.* (2005)) se considerar  el modelo jer rquico sin covariables.

## 3.2. Efectos correlacionados

Retomando las nociones antes mencionadas, una forma de mejorar la estimaci on del riesgo relativo local es suponer correlaci on positiva entre las observaciones. Sin p erdida de generalidad, existen dos formas de plantearlo; con distribuciones conjuntas y distribuciones condicionales. De hecho, los modelos que se presentan en esta secci on se clasifican por dicha especificaci on para la distribuci on inicial del modelo jer arquico propuesto.

Es bien sabido que la distribuci on normal multivariada es una de las distribuciones m as flexibles para representar variables aleatorias correlacionadas. Sea  $\mathbf{R} = \{R_1, \dots, R_n\}$  el vector de los log-riesgos relativos de  area.

### 3.2.1. Distribuciones conjuntas

Por un lado, podemos especificar la estructura de dependencia en t erminos de una matriz de covarianzas de una distribuci on Normal Multivariada como sigue:

$$\mathbf{R} \sim N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}),$$

donde  $\boldsymbol{\Sigma} = \sigma^2 \boldsymbol{\Omega}$ . Por razones de parsimonia, es usual especificar los elementos de la matriz de correlaci on como una funci on de la distancia entre los centroides de cada par de regiones  $i$  y  $j$ , es decir,  $\Omega_{ij} = f(d_{ij}; \phi)$ . Es claro que se debe asegurar que la funci on elegida genere una matriz definida positiva, as ı como evitar la confusi on entre los par ametros de correlaci on y la media. Una de las elecciones m as usuales para  $f(\cdot)$  es la funci on exponencial decreciente,  $f(d_{ij}; \phi) = \exp(-\phi d_{ij})$ , donde  $\phi > 0$  controla la tasa de decrecimiento.

Sin embargo, en el contexto del mapeo de la enfermedad, en muy pocas ocasiones existe informaci on suficiente para la proposici on de la forma funcional. M as

a un, la escasez de informaci on en la muestra observada para estimar los par ametros de la funci on de correlaci on, as ı como la costosa carga computacional (debido a la inversi on de la matriz de covarianzas en cada iteraci on) que conlleva su implementaci on v ıa algoritmos MCMC, limitan la pr actica de esta clase de modelos.

### 3.2.2. Distribuciones marginales

Por otro lado, podemos especificar la estructura de dependencia en t erminos de una serie de distribuciones normales condicionales las cuales, en su forma m as general se pueden escribir como:

$$\mathbb{E}(R_i | \mathbf{r}_{(-i)}) = \mu_i + \sum_{j \neq i} a_{ij}(R_j - \mu_j)$$

$$\mathbb{V}(R_i | \mathbf{r}_{(-i)}) = \kappa_i,$$

donde  $\mathbf{r}_{(-i)}$  denota los valores de todos los log-riesgos excepto el de la  $i$ - esima regi on. El par ametro  $\mu_i$  representa la tendencia espacial en la  $i$ - esima regi on (usualmente se asume constante en toda la regi on de estudio o especificada por una funci on de covarianza) y las  $a_{ij}$  son coeficientes que reflejan la dependencia espacial local entre las regiones  $i$  y  $j$ . Bajo ciertas condiciones de  $\mu_i$ ,  $a_{ij}$  y  $\kappa_i$  se puede probar que  $\mathbf{R} \sim N_n(\mathbf{0}, \mathbf{P}^{-1})$ , donde  $P_{ii} = \frac{1}{\kappa_i}$  y  $P_{ij} = -\frac{a_{ij}}{\kappa_i}$ .

Este enfoque permite una mejor estimaci on de los log-riesgos relativos considerando en cada regi on toda la informaci on global disponible para posteriormente obtener una estimaci on del riesgo relativo, que es una combinaci on entre la informaci on del lugar y de las regiones a las que se les asocio dependencia espacial. De este modo, las fluctuaciones de la SMR se reducen y el mapa se suaviza, es decir, se filtra la variaci on extra-Poisson. Estos modelos son conocidos como Campos Aleatorios de Markov Gaussianos (GMRF) o Condicionales Auto-Regresivos Gaussianos (GCAR).

Una formulaci on muy com un es

$$a_{ij} = \frac{\gamma\omega_{ij}}{n_i}$$

$$\kappa_i = \frac{\sigma^2}{n_i},$$

donde  $\omega_{ij}$  se considera como el peso que mide la dependencia espacial entre las regiones  $i$  y  $j$ , asignado al  $ij$ - esimo elemento de la matriz sim etrica de pesos  $\mathbf{W}$ , y  $n_i = \sum_{j=1}^n \omega_{ij}$ . La elecci on t ipica para la matriz  $\mathbf{W}$  est a dada por

$$\omega_{ij} = \begin{cases} 1 & \text{si las regiones } i \text{ y } j \text{ comparten frontera} \\ 0 & \text{en otro caso} \end{cases}$$

De este modo,  $n_i$  es el n umero de vecinos de la  $i$ - esima regi on.

Esta parametrizaci on permite una intuitiva interpretaci on de la media condicional  $\mathbb{E}(r_i|\mathbf{r}_{(-i)})$  como una media ponderada de los efectos aleatorios de las regiones adyacentes a la  $i$ - esima regi on pero fuerza a que la varianza condicional  $\mathbb{V}(R_i|\mathbf{r}_{(-i)})$  no sea constante para todas las regiones.

El modelo cl asico de esta clase gen erica fue introducido por Clayton y Kaldor (1987), quienes propusieron m etodos bayesianos emp iricos construidos a partir de una regi on Poisson con interceptos aleatorios con correlaci on espacial. M as tarde, Besag *et al.* (1991) extendieron el modelo dentro de una estructura bayesiana completa, el cual fue nombrado modelo de convoluci on por Molli e (1996). Con respecto al planteamiento (3.1), su formulaci on es la siguiente: el logaritmo del riesgo relativo para cada regi on puede modelarse como

$$\log(\lambda_i) = u_i + v_i,$$

donde los par ametros  $u_i$  y  $v_i$  sirven para representar la heterogeneidad local de la regi on y la dependencia espacial con sus regiones vecinas, respectivamente. Las

distribuciones iniciales sobre estos par ametros pueden ser especificadas como sigue: la distribuci on del modelo para la heterogeneidad local se supone

$$u_i \sim N(0, \sigma_u^2) \quad i = 1, 2, \dots, n.$$

Para la correlaci on espacial, en donde las estimaciones del riesgo relativo en una regi on dependen de las estimaciones en las regiones vecinas se recurre a un modelo condicional auto-regresivo gaussiano (GCAR):

$$v_i | \mathbf{v}_{(-i)} \sim N(\bar{v}_i, \tau_i),$$

donde

$$\bar{v}_i = \frac{1}{n_i} \sum_{j=1}^n v_j \omega_{ij}$$

$$\tau_i = \frac{\sigma_v^2}{n_i}.$$

Los hiperpar ametros  $\sigma_u^2$  y  $\sigma_v^2$ , que controlan la variabilidad de  $u$  y  $v$ , se consideran variables aleatorias con distribuci on gamma inversa para ambos tal y como se sugiere en [Bernardinelli et al. \(1995\)](#).

Con excepci on de la distribuci on inicial gamma independientes de [Clayton y Kaldor \(1987\)](#), la mayor a de la literatura estad stica se concentra en definir distribuciones iniciales para los log-riesgos relativos y, en particular, encontrar una distribuci on normal con estructura de correlaci on operable. No obstante, [Nieto-Barajas \(2008\)](#) sugiere trabajar directamente con el espacio parametral de los riesgos relativos y modelar su distribuci on espacial por medio de un Campo Aleatorio de Markov Gamma.

### 3.3. Modelos semi-par ametricos

Con cualquiera de la especificaciones param etricas descritas anteriormente, el nivel de suavizamiento conseguido es afectado globalmente por todas las regiones. En consecuencia, si estamos interesados en identificar grupos de regiones geogr aficamente contiguas con riesgo elevado o bajo, dichas parametrizaciones pueden sobreajustar el valor del riesgo relativo y ocultar las discontinuidades del espacio parametral. Aunque la detecci on de conglomerados en enfermedades es a primera vista un problema aparte, en enfermedades raras donde los factores de riesgo tienen una estructura espacial puede presentarse dicho comportamiento. Adem as, tambi en se podr a estar interesado en la detecci on de discontinuidades en el mapa, es decir, diferencias sospechosas en el riesgo relativo entre regiones adyacentes. Por ello, con la finalidad de ofrecer una alternativa estad istica para estas situaciones, varios autores han desarrollado modelos espaciales semi-par ametricos, los cuales tienen en com un la idea de reemplazar la distribuci on espacial continuamente variable de  $\{\lambda_1, \dots, \lambda_n\}$  por modelos de agrupaci on tales que

$$\lambda_i = \lambda_{z_i},$$

donde  $\{z_1, \dots, z_n\}$  son variables de asignaci on que toman valores en  $\{1, \dots, k\}$ . Es decir,  $\{\lambda_1, \dots, \lambda_n\}$  forman  $k$  diferentes grupos a los que llamaremos *cl usters* y dentro de los cuales el riesgo relativo es constante para todas las subregiones que lo conforman.

Los modelos propuestos bajo esta clase difieren en el mecanismo de asignaci on de las subregiones a los cl usters, pero son similares en el sentido de permitir discontinuidades en el espacio parametral y de relajar los supuestos distribucionales. Sin embargo, su m erito principal radica en la flexibilidad para tratar al n umero de cl usters desconocido (y, por tanto, considerarlo como una variable a estimar). En consecuencia, la implementaci on de estos modelos es a trav es de un algoritmo RJMCMC (ver Secci on 2.2.2). Por otro lado, en diversos enfoques frecuentistas se

ha sugerido la determinaci on del n umero de cl usters o la detecci on de cl usters bajo un esquema de validaci on de hip otesis. Desafortunadamente, estos procedimientos adolecen de una serie de inconvenientes dentro de los cuales destaca la dificultad de la especificaci on del  $p$ -valor.

Como se mencion o, los modelos difieren en el mecanismo de asignaci on, resultado de la elecci on del modelo espacial para el proceso  $\{z_1, \dots, z_n\}$ . De hecho, considerando la contiguidad de las subregiones de cada cl uster, esta clase se puede dividir en t erminos de dos especificaciones param etricas: mezcla finita de distribuciones y particiones, las cuales se presentan a continuaci on.

### 3.3.1. Mezcla finita de distribuciones

Green y Richardson (2002) proponen un modelo de mezclas finitas de distribuciones para analizar la heterogeneidad espacial del riesgo, en el cual se supone que el mecanismo de asignaci on est a dado por un proceso espacialmente correlacionado. La formulaci on de dicho proceso est a basada en la construcci on del grafo simple asociado al campo aleatorio oculto  $\{z_1, \dots, z_n\}$  con estructura markoviana (este modelo se puede ver como una extensi on de los modelos de Markov ocultos a campos aleatorios de Markov con estados discretos ocultos). Adem as, considera un par ametro de interacci on no negativo,  $\psi$ , que controla el grado de dependencia espacial mediante el artilugio de beneficiar a aquellos patrones de asignaci on en los cuales incrementa el n umero de vecinos que pertenecen al mismo cl uster cuando se incrementa el valor de  $\psi$ . Este mecanismo de asignaci on es conocido como el modelo Potts. En contraste con el planteamiento jer arquico del modelo de mezclas propuesto por Richardson y Green (1997) (de hecho, con la mayor a de modelos de mezclas), el modelo Potts no hace uso de pesos expl ıcitos para los componentes, as ı que modela conjuntamente al campo aleatorio  $\{z_1, \dots, z_n\}$ . Para completar la especificaci on del modelo, definiremos las distribuciones iniciales para  $\lambda$ ,  $\psi$  y  $k$ . Por parsimonia, los riesgos relativos se suponen independientes e id enticamente distribuidos con distribuci on inicial Gamma. Para la distribuci on inicial de  $\psi$  se propone

una distribuci3n uniforme discreta sobre los valores de un intervalo discretizado y acotado por un valor m aximo determinado. Por  ultimo, dado que consideraremos el n mero de componentes de la mezcla como variable aleatoria, usualmente la distribuci3n inicial para  $k$  es una distribuci3n uniforme discreta sobre los valores  $\{1, \dots, k_{max}\}$ . En resumen,

$$\begin{aligned} \lambda_i &= \lambda_{z_i}, \quad i = 1, \dots, n \\ z_i &\in \{1, \dots, k\} \\ \lambda_{z_i} &\sim \text{Gamma}(\alpha, \beta), \quad i = 1, \dots, n \\ p(\mathbf{z}|\psi, k) &= e^{\psi U(\mathbf{z}) - \theta_k(\psi)} \\ k &\sim U(1, k_{max}) \end{aligned}$$

donde  $U(\mathbf{z}) = \sum_{j \neq i} \mathbb{I}(z_i = z_j)$  es el n mero de vecinos de la regi3n  $i$  que fueron asignados al mismo cl ster en la configuraci3n  $\mathbf{z}$ . Este mecanismo de asignaci3n no s3lo captura las creencias a priori acerca de la similaridad espacial del riesgo en regiones vecinas sino que tambi3n permite que regiones no contiguas pertenezcan al mismo cl ster. El  ltimo t rmino del modelo Potts,  $\theta_k(\psi) = \log \left( \sum_{\mathbf{z} \in \{1, \dots, k\}^n} e^{\psi U(\mathbf{z})} \right)$ , es la constante que normaliza la distribuci3n para  $\{z_1, \dots, z_n\}$ , donde la suma es sobre todas las posibles configuraciones de asignaci3n de las  $n$  subregiones.

Otra formulaci3n que se basa en un modelo de mezcla finita de distribuciones para modelar la heterogeneidad espacial se presenta en [Fern andez y Green \(2002\)](#). En los dos modelos propuestos se supone que la distribuci3n conjunta de los pesos posee la estructura de un MRF.

### 3.3.2. Particiones

Una clase que comparte el mismo esp ritu de los modelos semi-par mtricos son los modelos de partici3n espacial introducidos por [Knorr-Held y Ra er \(2000\)](#) y [Denison y Holmes \(2001\)](#). Como en cualquier modelo de particiones, se asume

que la regi3n de estudio puede dividirse en varios cl usters, cada uno con riesgo relativo constante e independientes entre s ı, y con el n umero de clusters considerado como desconocido. Estos modelos difieren t ecnicamente en la forma como se definen los cl usters y en la especificaci3n de sus respectivas distribuciones iniciales.

En el modelo de [Knorr-Held y Ra er \(2000\)](#), para un n umero fijo de cl usters  $k$ , como primer paso se eligen al azar (equiprobablemente)  $k$  de las  $n$  regiones y se definen como ‘‘centros de cl uster’’, llam emosles  $g_1, \dots, g_k$ . Cada centro de cl uster  $g_j$ , definir a al cl uster  $C_j$ . Dado el conjunto de centros de cl usters,  $G_k = \{g_1, \dots, g_k\}$ , el mecanismo de asignaci3n est a dado como sigue: la regi3n  $i$  ( $i \notin G_k$ ) es asignada al cl uster  $C_j$  si  $g_j$  es el m as cercano en el sentido del n umero m ınimo de regiones que se tendr ıan que cruzar para llegar de  $i$  a  $g_j$ . Sin embargo, esta construcci3n puede permitir que algunas regiones tengan la misma distancia con dos o m as centros de cl uster. Para asegurar la unicidad, asignaremos esas regiones en conflicto al cl uster con  ındice m as peque no que cumpla la condici3n anterior. Como suposici3n previa para los riesgos relativos, parece natural asumir que se distribuyen sim etricamente en la escala logar ıtmica, por lo que el logaritmo de los riesgos relativos siguen una distribuci3n Normal. Para finalizar, la distribuci3n inicial para  $k$  es proporcional a  $(1 - c)^k$ , donde  $c \in [0, 1)$ . El caso l ımite  $c = 0$  corresponde a la distribuci3n uniforme discreta sobre  $\{1, \dots, n\}$ , mientras que el caso  $c > 0$  corresponde a la distribuci3n geom etrica truncada. En resumen,

$$\begin{aligned} \log(\lambda_i) &= \eta_{z_i}, \quad i = 1, \dots, n \\ z_i &\in \{1, \dots, k\} \\ \eta_{z_i} &\sim \text{Normal}(\mu, \sigma^2), \quad i = 1, \dots, n \\ k &\sim U(1, k_{max}) \end{aligned}$$

Por otro lado, en [Denison y Holmes \(2001\)](#) hacen uso de la idea de los mosaicos de Voronoi como mecanismo de agrupaci3n. De hecho, la construcci3n para los cl uster adoptado en el modelo KHR se puede ver como una modificaci3n del

modelo del mosaico de Voronoi para espacios irregulares discretos, y por tanto posiblemente m as apropiado para datos asociados a regiones peque nas. Cada regi n es considerada como un punto de realizaci n y representada por las coordenadas de su centroide. Se supone que los puntos generados mediante el mosaico de Voronoi pueden ser asignados a cualquier punto en la regi n de inter s. Condicional en el conjunto de los  $k$  puntos generados mediante el mosaico de Voronoi (an logos a los  $g_j$  en el modelo KHR), la  $j$ - sima partici n est  compuesta por las regiones con centroide m s cercano (en distancia euclidiana) al  $j$ - simo punto generado que a cualquier otro punto generado. Otra diferencia entre estos dos modelos de particiones es que [Denison y Holmes \(2001\)](#) usan iniciales gammas para  $\eta_j$ , en lugar de la distribuci n lognormal, con el fin de explotar el conjugamiento Poisson-Gamma y as  lograr una implementaci n eficiente del modelo. Notemos que en esta clase de modelos todas las regiones en cada cl ster son contiguas, mientras que en el modelo espacial de mezclas descrito anteriormente existen regiones que pertenecen al mismo cl ster pero no son contiguas. Debido a esto, el n mero de cl sters en estos modelos tender  a ser un poco m s grande que el n mero de componentes distribucionales de la mezcla.

Finalmente, se resalta el hecho de que los modelos de particiones no poseen una estructura markoviana para  $\{z_1, \dots, z_n\}$ , en contraste con el modelo de mezcla de distribuciones.

### 3.4. Extensiones

Como se coment  anteriormente, el formato de los datos epidemiol gicos limitar  las especificaciones param tricas disponibles para el an lisis de los datos. No obstante, existen otras caracter sticas que definir n el enfoque estad stico bajo el cual se debe inferir la distribuci n subyacente. Por ejemplo, en el contexto del mapeo de la enfermedad, cuando la base de datos consta de observaciones recopiladas en distintos tiempos o cuando la observaci n en cada sitio es un vector cuyas

---

coordenas corresponden a las mediciones relativas a dos o m as enfermedades, se deben considerar modelos espacio-temporales y modelos vectoriales, respectivamente. Los modelos espacio-temporales tienen el objetivo de ofrecer una visi n din mica de la enfermedad, mientras que en el an lisis de m ltiples enfermedades se busca exhibir un conjunto de factores de riesgo no observables comunes. Para profundizar en ambas extensiones, v ase [Lawson \(2013\)](#).



## Capítulo 4

# Mapas de mortalidad por cáncer de mama para el sur del Estado de México

El cáncer se origina cuando las células sufren una alteración, la cual propicia el aumento descontrolado de células anormales que se propagan e invaden los tejidos (a este proceso se le conoce como *metástasis*) formando una *neoplasia*. Existen más de 100 tipos de cáncer maligno, los cuales se denominan dependiendo de la parte del cuerpo donde inician su desarrollo. Si bien en algunos casos está implícito un factor genético, su desarrollo es multifactorial, por lo que su aparición puede depender de la exposición a agentes externos. Adicionalmente, conforme la población envejece, los mecanismos de reparación celular pierden eficacia, de tal forma que la edad es uno de los factores de riesgo más importantes para el desarrollo de ciertos tipos de cáncer como el cáncer de mama. Actualmente, el cáncer es la tercer causa de muerte a nivel mundial, después de las enfermedades del corazón y la diabetes mellitus. Según la Organización Mundial de la Salud (OMS), en 2012 murieron 8.2 millones de personas por esta enfermedad; los varones principalmente por cáncer de pulmón, estómago, colon y recto, próstata e hígado; mientras que las mujeres por pulmón, estómago, colon y recto, cervical y seno (aunque este último puede presentarse también en hombres). Además, destaca que aproximadamente 70 % de

las muertes por c ancer se presentan en pa ses de ingresos bajos y medios, debido a un menor acceso a los servicios de salud para la detecci on temprana, tratamiento y control.

A pesar de los avances de la oncolog a moderna y de la reducci on en la mortalidad por c ancer, el c ancer de mama es la primera causa de muerte por neoplasia maligna en la mujer en el  mbito mundial, con m s de 500 mil muertes cada a o. El riesgo de enfermar es superior en las mujeres de pa ses con nivel socioecon mico alto, pero el riesgo de morir es mayor entre las mujeres que habitan pa ses pobres debido a la tendencia ascendente, determinadas por el envejecimiento de la poblaci on, el aumento en la prevalencia de los factores de riesgo y la falta de un programa de intervenci on oportuna. El c ancer de mama constituye una grave amenaza a la salud de las mujeres y bienestar de las familias, as  como un problema de salud p blica a nivel mundial. Contrario a lo que ha pasado en pa ses desarrollados, donde esta enfermedad es ya una prioridad, en los pa ses en v as de desarrollo se ha puesto poca atenci on a este problema sanitario cada vez m s grave. Recientemente, las agencias de salud han promovido pol ticas para mejorar los servicios de prevenci on, diagn stico, tratamiento, control y vigilancia de la enfermedad pero no han implementado la disposici on de datos confiables para la cuantificaci on del alcance de dichas actividades. En forma paralela, los trabajos de investigaci on han producido nuevas opciones terap uticas, muchas de las cuales tienen un costo elevado. Por consiguiente, la epidemia de c ancer de mama representa un desaf o para los sistemas de salud, en particular para los pa ses en desarrollo.

## 4.1. C ancer de mama en M xico

En comparaci on con el resto del mundo, M xico se mantiene con una de las tasas m s bajas de mortalidad por c ancer. La evoluci on y tendencia de la mortalidad por c ancer y por entidad federativa han sido previamente reportados en

Torres-S anchez *et al.* (2014).

En cuanto a la magnitud actual del c ancer de mama en M exico, a partir de 2006  este ocupa el primer lugar de mortalidad por tumor maligno en las mujeres mayores de 25 a os, desplazando de esa posici on al c ancer cervicouterino. La tasa de mortalidad observada por c ancer de mama durante los  ultimos a os no ha mostrado grandes variaciones; de forma general, de 2007 a la fecha, 15 de cada 100 mil mujeres de 20 a os y m as han fallecido por esta enfermedad. El an alisis de la mortalidad muestra diferencias notorias por entidad federativa, con las tasas m as altas en los estados del centro y norte. Cinco estados concentraron casi la mitad de las muertes por c ancer de mama en 2010: Distrito Federal (16.5%), M exico (9.3%), Jalisco (8.2%), Veracruz (6.2%) y Nuevo Le on (6.2%).

Con respecto a las pol ticas, uno de los principales temas es el mejoramiento y la ampliaci on del tamizaje encaminado a promover la detecci on temprana. Los datos disponibles sugieren que s olo el 10% de los casos se detecta en las fases iniciales de la enfermedad. Una iniciativa clave para conseguir el objetivo anterior fue la reforma y legislaci on de 2003 que cre  el Seguro Popular de Salud. Sin embargo, en la pr actica, no todas las mujeres tienen hoy en d a acceso a estos servicios, dado que todav a existen barreras importantes tanto del lado de la demanda como de la oferta.

En resumen, esta afecci on se ha convertido para M exico en un problema apremiante de salud p blica. Tanto la mortalidad como el n mero de casos nuevos que se presentan cada a o se han incrementando paulatinamente, siendo impostergable el an alisis de los factores de riesgo no observables, as  como el control de los factores de riesgo conocidos y la implementaci on de un programa de detecci on organizado que incluya la garant a de calidad en todos los procesos, desde la detecci on hasta el tratamiento y seguimiento de pacientes con c ancer.

## 4.2. Distribución espacial de la mortalidad por cáncer de mama en el sur del Estado de México: 2010-2013

En el mundo, los patrones geográficos de la mortalidad por cáncer han sido proveedores de claves importantes acerca de las posibles causas; ver, por ejemplo, [Knorr-Held y Becker \(2000\)](#) y [Mollié \(1996\)](#). En esta sección se construyen los mapas de mortalidad por cáncer de mama para el periodo comprendido entre 2010 y 2013 en el sur del Estado de México basados en los modelos propuestos por [Besag \*et al.\* \(1991\)](#) y [Green y Richardson \(2002\)](#) (de aquí en adelante denotados por BYM y GR), y discutidos en las Secciones [3.2.2](#) y [3.3.1](#), respectivamente.

### 4.2.1. Metodología

Para identificar todas las muertes por cáncer de mama, se usaron los códigos de la Décima Revisión de la lista de la Clasificación Estadística Internacional de Enfermedades y Problemas Relacionados con la Salud (CIE-10). El número de decesos en cada municipio se obtuvo de la base de datos de los certificados de defunción, validadas y publicadas anualmente por el Instituto Nacional de Estadística y Geografía (INEGI). A su vez, la distribución de la población (a mitad del año), por municipio, edad y sexo desde 2010 hasta 2013, se obtuvo de las proyecciones de población del Consejo Nacional de Población (CONAPO). Ambas bases de datos están disponibles en sus correspondientes sitios web.

Para calcular el número esperado de decesos por cáncer de mama, se utilizó la población femenina en la República Mexicana bajo el Censo de Población y Vivienda 2010 y el número de decesos para los grupos de edad: 0 a 29 años, 30 a 44 años, 45 a 64 años y 65 años y más (para el mismo año) para establecer las tasas específicas por estrato de la población estándar. A partir de estos datos, se

emple o el m etodo de estandarizaci on indirecta (ver Secci on 2.1.2) para el c alculo de la SMR para cada uno de los 14 municipios que forman al sur del Estado de M exico desde 2005 (Almoloya de Alquisiras, Amatepec, Coatepec de Harinas, Oztoloapan, San Sim on de Guerrero, Santo Tom as, Sultepec, Tejupilco, Temascaltepec, Texcaltitlan, Tlatlaya, Zacazonapan, Zacualpan y Luvianos). Agregar los datos de cuatro a os permite una mayor estabilidad de las tasas de mortalidad por municipio.

El manejo de las bases de datos, la obtenci on de las tasas de referencia y el c alculo del n umero esperado de decesos por c ancer de mama (por a o, municipio y grupo de edad) se realiz o con la ayuda de la paqueter a de Office (Excel). La construcci on de los mapas se realiz o con el software estad stico R, desarrollado por [R Core Team \(2015\)](#).

Naturalmente, para ajustar los datos a los modelos bayesianos propuestos, es necesario implementar m etodos MCMC. La estimaci on del riesgo relativo, bajo el modelo BYM, se obtuvo con la ayuda del software de inferencia bayesiana OpenBUGS (plataforma experimental de [The BUGS project \(2014\)](#)). Particularmente, con la funci on de distribuci on *car.normal*, la cual genera una cadena cuya distribuci on objetivo es el GMRF subyacente.

Con respecto al modelo GR, se construy o una cadena de Markov mediante un algoritmo RJMCMC de dos movimientos para calcular el valor esperado de los par metros. Uno de los dos movimientos es de dimensi on fija (es decir,  $k$  fijo) y se realiza a trav es de un m etodo de Gibbs por bloques. Este movimiento actualiza los riesgos relativos  $\lambda$ , los par metros de asignaci on  $z$  y de interacci on espacial  $\psi$ . De hecho, cada  $z_i$  tiene distribuci on condicional completa proporcional a un kernel de Gibbs:

$$p(z_i = j | \dots) \propto e^{-\lambda_j e_i} \lambda_j^{o_i} e^{\psi n_{ij}},$$

donde  $n_{ij}$  es el n umero de vecinos de la regi on  $i$  asociadas actualmente al cl uster  $j$ ;  $\psi$  tiene distribuci on condicional completa proporcional a

$$p(\psi) e^{\psi U(\mathbf{z}) - \theta_k(\psi)}$$

y  $\boldsymbol{\lambda}$  tiene una distribuci on condicional completa proporcional a

$$\prod_{j=1}^k \lambda_j^{\alpha + \sum_{i:z_i=j} o_i} \exp\{-\lambda_j(\beta + \sum_{i:z_i=j} e_i)\}.$$

Para generar el siguiente estado de la cadena, cuando no existe cambio de dimensi on, se procede como sigue. Para actualizar  $\psi$  se propone una caminata aleatoria con perturbaciones equiprobables de  $\pm 1$ . La discretizaci on del soporte de  $\psi$  (la cual seg un [Green y Richardson \(2002\)](#) no tiene ning un impacto significativo en la inferencia) se debe principalmente a la complejidad combinatorica involucrada en el c alculo de la constante de normalizaci on del modelo Potts. Por ello, dicho valor ser a calculado exactamente mediante un proceso en paralelo (con la ayuda de los paquetes de R **doParallel** y **partitions**, a trav es de la funci on *constante* del Ap endice [A.3](#)) y almacenada en la matriz *thetas* para los valores de  $\psi = 0, 0.1, \dots, 1.9, 2$  y  $k = 1, 2, 3, 4$ . Para actualizar el vector  $\boldsymbol{\lambda}$ , suponiendo la restricci on de orden  $\lambda_1 < \lambda_2 < \dots < \lambda_k$ , se propone un incremento simultaneo independiente normal est andar en cada  $\log(\lambda_i)$ . Despu es, los valores modificados se ordenan para obtener el valor propuesto, digamos  $\boldsymbol{\lambda}'$ . Entonces, la probabilidad de aceptaci on Metropolis-Hastings de las tasas propuestas es

$$\min \left\{ 1, \prod_{j=1}^k \left[ \left( \frac{\lambda'_j}{\lambda_j} \right)^{\alpha + \sum_{i:z_i=j} o_i} \exp\{-(\lambda'_j - \lambda_j)(\beta + \sum_{i:z_i=j} e_i)\} \right] \right\}.$$

El segundo movimiento del algoritmo RJMCMC permite el cambio de dimensi on en el espacio parametral mediante un salto reversible. Se adoptar a el m etodo de elecci on aleatoria planteado en [Richardson y Green \(1997\)](#) para dividir un cl uster en dos cl usters o unir dos cl usters en un cl uster. Supongamos que la configuraci on actual consta de exactamente  $k$  cl usters. Por un lado, para aumentar la dimensi on parametral se elige un cl uster al azar (equiprobablemente),  $j$ . Este se elimina y

ser a remplazado por dos cl usters, denotados por “−” y “+”, con riesgos relativos  $\lambda_- = \lambda_j u^c$  y  $\lambda_+ = \lambda_j u^{-c}$ , respectivamente, donde  $u \sim U(0, 1)$  y  $c$  es un par metro de alargamiento. Si  $\lambda_- < \lambda_{j-1}$     $\lambda_{j+1} < \lambda_-$  (con sus respectivas omisiones para los casos  $j = 1$     $j = k$ ), entonces el movimiento es rechazado. En otro caso, las regiones actualmente asignadas al cl uster  $j$  son reasignadas a los cl usters “−” y “+” mediante el siguiente esquema probabil stico:  $z_i$  es asignado al cl uster “−” con probabilidad

$$\frac{e^{\psi n_- - \lambda_- e_i \lambda_-^{o_i}}}{e^{\psi n_- - \lambda_- e_i \lambda_-^{o_i}} + e^{\psi n_+ - \lambda_+ e_i \lambda_+^{o_i}}},$$

donde  $n_-$  y  $n_+$  es el n mero de vecinos de la  $i$ - sima regi n que actualmente est n asignados a los cl usters “−” y “+”; es decir,  $n_- = 0$  y  $n_+ = 0$  cuando se inicia una reasignaci n. Este mecanismo tiene el efecto de imitar al t rmino del modelo Potts en la distribuci n objetivo para favorecer asignaciones propuestas con coherencia espacial. Por tanto, la probabilidad de aceptar la reasignaci n propuesta es  $\min\{1, R\}$ , donde  $p_{k+1,k}$  es la probabilidad de unir dos cl usters en un cl uster,  $p_{k,k+1}$  es la probabilidad de dividir un cl uster en dos cl usters,  $P_{alloc}$  es el producto de las probabilidades de reasignaci n de las regiones reasignadas y

$$R = \frac{p(\boldsymbol{\lambda}^*, \mathbf{z}^*, \psi, k+1, \mathbf{o})}{p(\boldsymbol{\lambda}, \mathbf{z}, \psi, k, \mathbf{o})} \times \frac{p_{k+1,k}}{p_{k,k+1} P_{alloc}} \times \frac{2c\lambda_j}{u}.$$

Por otro lado, para reducir la dimensi n parametral se elige al azar (equiprobablemente) un par de cl usters cuyos riesgos relativos sean contiguos. Para ello, se elige un valor del conjunto de sub ndices de los  $k-1$  cl usters con riesgo relativo menor,  $j \in \{1, 2, \dots, k-1\}$ , as  como su sucesor. Estos se eliminan y ser n remplazados por un cl uster, denotado por “ ”, con riesgo relativo  $\lambda_{\pm} = \sqrt{\lambda_j \lambda_{j+1}}$ . Las regiones actualmente asignadas a los cl usters  $j$  y  $j+1$  son reasignadas al cl uster “ ”. En este caso, la probabilidad de aceptar la reasignaci n propuesta es  $\min\{1, R\}$ , donde

$$R = \frac{p(\boldsymbol{\lambda}^*, \mathbf{z}^*, \psi, k-1, \mathbf{o})}{p(\boldsymbol{\lambda}, \mathbf{z}, \psi, k, \mathbf{o})} \times \frac{p_{k-1,k}}{p_{k,k-1}} \times \frac{(\frac{\lambda_j}{\lambda_{j+1}})^{1/2c}}{2c\sqrt{\lambda_j \lambda_{j+1}}}.$$

Cabe puntualizar que la funci n propuesta para mapear uno a uno los espacios parametrales  $\{\lambda_1, \dots, \lambda_k, \psi, k, u\}$  y  $\{\lambda_1, \dots, \lambda_{k+1}, \psi, k+1\}$ , llam mosle  $g$ , est  dada por  $g(\lambda_1, \dots, \lambda_j, \dots, \lambda_k, \psi, k, u) = (\lambda_1, \dots, \lambda_{j-1}, \lambda_j u^c, \lambda_j u^{-c}, \lambda_{j+1}, \dots, \lambda_k, \psi, k+1)$ .

### 4.2.2. Resultados

Considerando el marco demogr fico definido, el n mero total de decesos registrados hasta la fecha es de 24; 7 en 2010, 8 en 2011, 4 en 2012 y 5 en 2013. La Tabla 4.1 resume los casos observados por a o y municipio, as  como los casos esperados para el periodo de estudio. Las tasas de mortalidad estandarizadas oscilan entre 0 y 1.8 (ver Tabla 4.2). La Figura 4.2 es la representaci n espacial de dichos valores.

Las estimaciones del riesgo relativo bajo el modelo BYM, construidas a partir de la simulaci n de una cadena de Markov de 100000 iteraciones con un periodo de calentamiento de 50000 (generada por la rutina del Ap ndice A.2 cuando se fija la semilla 27), se concentraron entre los valores 0.2 y 1.7. Dichos valores se muestran en la Tabla 4.2. La Figura 4.3 corresponde al mapa de mortalidad basado en dichos valores.

Por  ltimo, para estimar el riesgo relativo bajo el modelo GR, se proponen  $c = 0.1$ ,  $a = 1$  y  $\beta = \frac{\sum_{i=1}^{14} o_i}{\sum_{i=1}^{14} e_i}$  como se sugiere en la mayor a de las aplicaciones epidemiol gicas, y se consideran equiprobables las opciones de disminuir, mantener o aumentar la dimensi n (con sus respectivas analog as para los casos  $k = 1$     $k = 4$ ). Adem s, considerando el an lisis previo de la regi n presentado en L pez-D az *et al.* (2010), la distribuci n inicial para  $k$  se define como sigue:

$$p(k) = \begin{cases} 0.7 & \text{si } k = 3 \\ 0.1 & \text{si } k = 1, 2, 4. \end{cases}$$

Para diagnosticar la convergencia del algoritmo se construyeron 4 cadenas, cada una con los 4 posibles valores iniciales para  $k$ , las cuales producen los mismos estimadores puntuales.

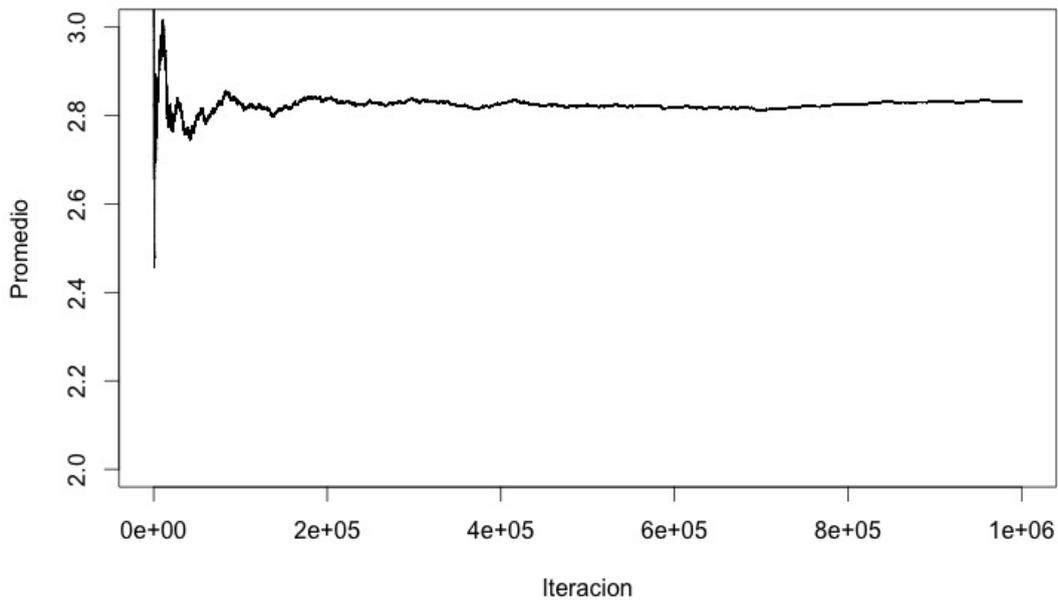


FIGURA 4.1: Media erg odica de  $k$ .

Las siguientes estimaciones se basan en la simulaci on de una cadena de Markov con saltos reversibles de 1000000 iteraciones con un periodo de calentamiento de 100000 y valores iniciales  $\lambda = (0.5, 1, 1.5)$ ,  $z = (3, 1, 1, 1, 2, 1, 1, 1, 2, 2, 1, 1, 2, 1)$ ,  $\psi = 1$  y  $k = 3$  (generada por la rutina del Ap endice A.3 cuando se fija la semilla 7). La Figura 4.1 corresponde al comportamiento de la media erg odica de  $k$  a trav es de cada iteraci on. No obstante, considerando la naturaleza del soporte de  $k$ , el estimador bayesiano de  $k$  es la moda de la distribuci on final de  $k$ . En este caso, 3. Por otro lado, las estimaciones del riesgo relativo se concentraron en un intervalo m as peque o y a la izquierda (como se esperaba), y fluct an entre 0.2 y 0.5. Nuevamente, los valores se muestran en la Tabla 4.2 y el mapa de mortalidad basado en  stos se presenta en la Figura 4.4.

Finalmente, es evidente que las estimaciones bajo el modelo GR son significativamente m as bajas que las estimaciones obtenidas con el modelo BYM. Esto se debe en buena medida a c omo est  definida la relaci on de una regi on con sus vecinos. No obstante, tanto el modelo BYM como el modelo GR conservan la

---

caracter stica espacial de asignar los riesgos relativos m as altos a vecindades donde se localizan los municipios con una SMR alta como Almoloya de Alquisiras y San Sim on de Guerrero. Adicionalmente, proporcionan estimaciones m as precisas de los riesgos relativos que los m etodos tradicionales; por ejemplo, en municipios donde no existen casos observados como ocurre en los municipios de Otzoloapan, Tlatlaya y Zacazonapan.

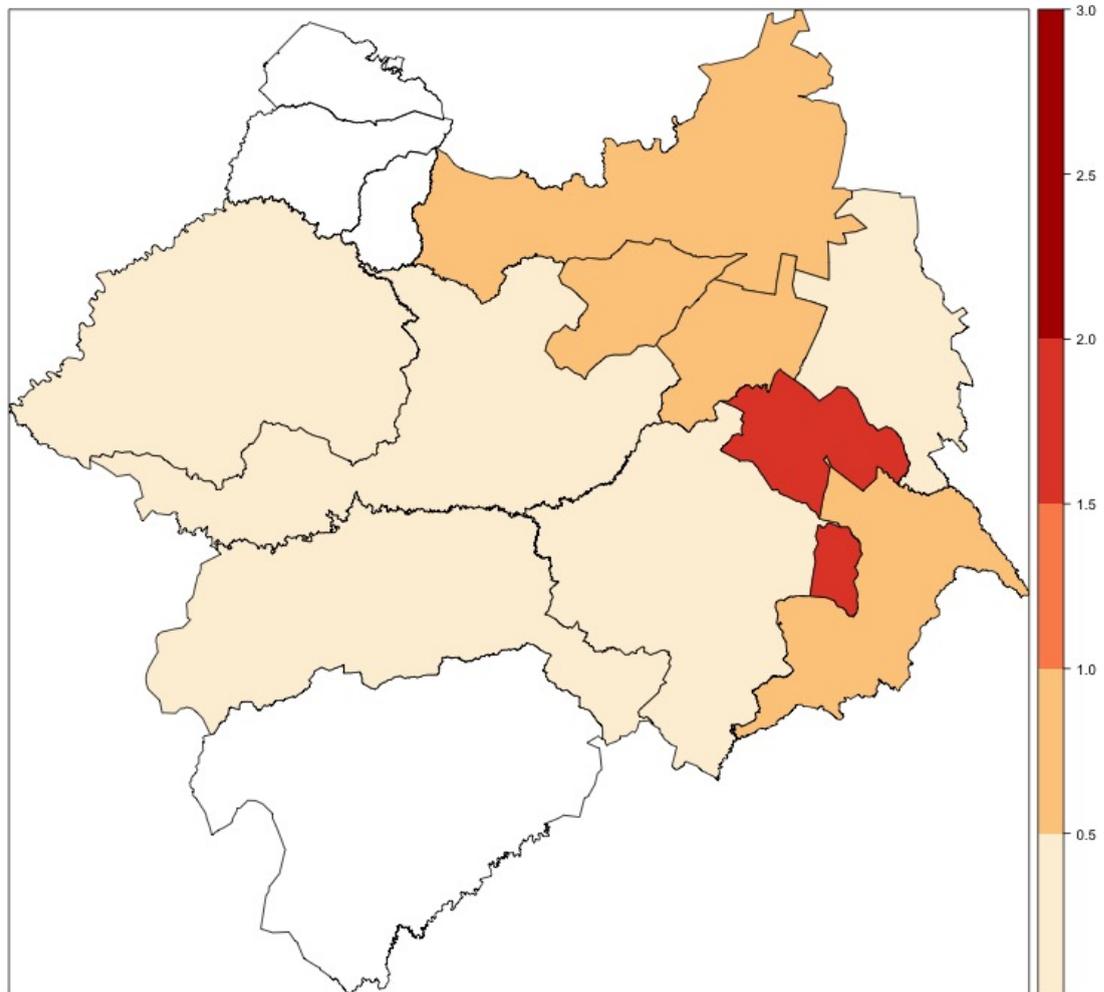


FIGURA 4.2: Tasa de Mortalidad Estandarizada por c  ncer de mama para los 14 municipios del sur del Estado de M  xico: 2010-2013.

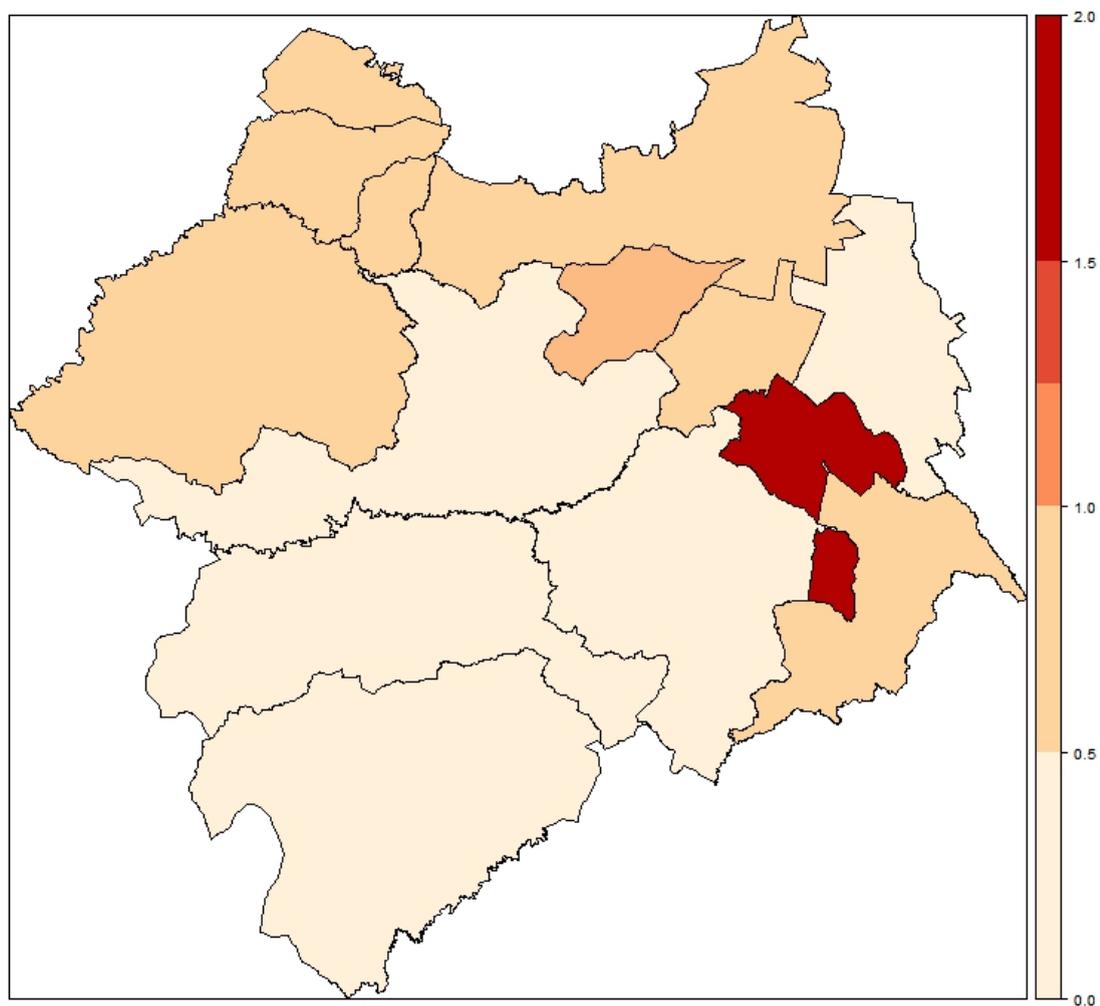


FIGURA 4.3: Estimaci n del riesgo relativo, v a modelo BYM, por c ncer de mama para los 14 municipios del sur del Estado de M xico: 2010-2013.

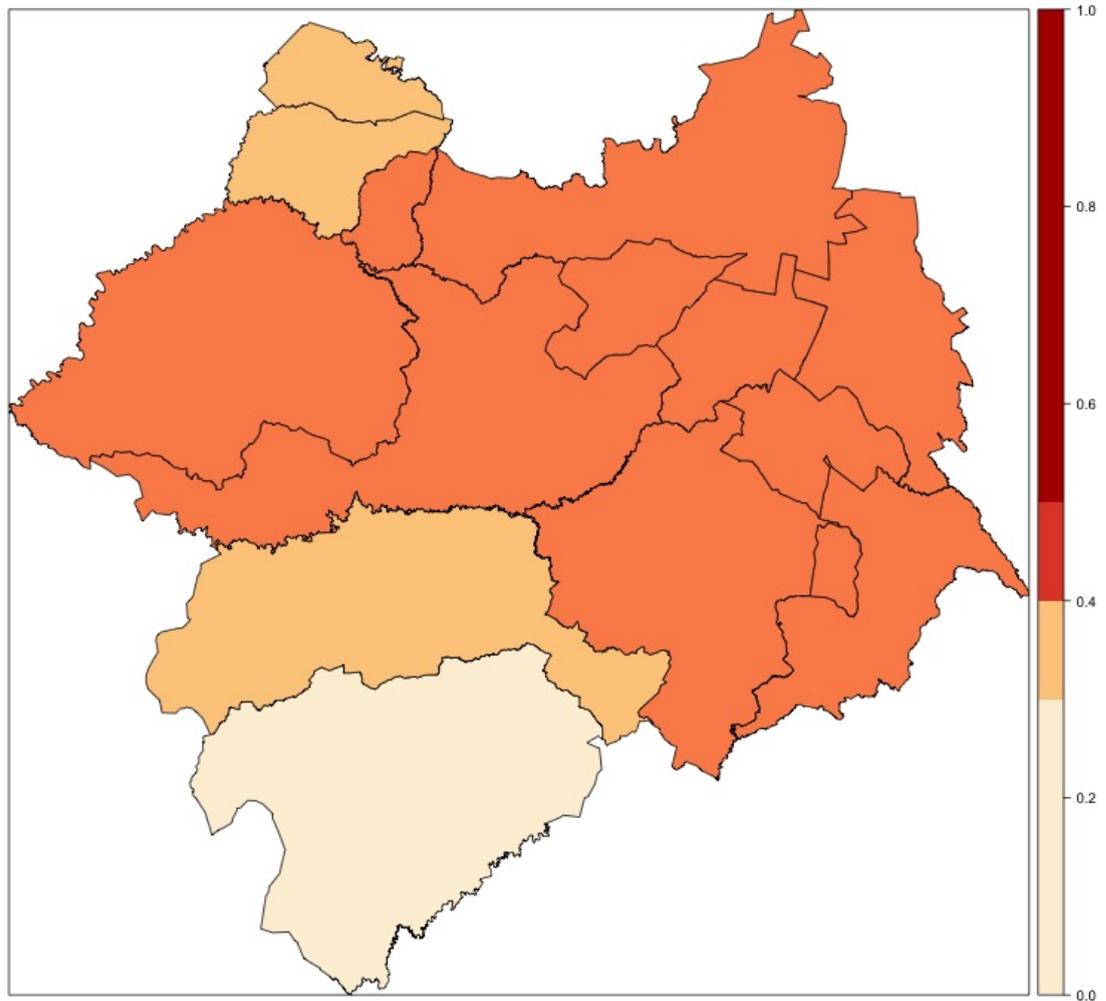


FIGURA 4.4: Estimaci  n del riesgo relativo, v  a modelo GR, por c  ncer de mama para los 14 municipios del sur del Estado de M  xico: 2010-2013.

Municipio	Observados					Esperados
	2010	2011	2012	2013	Total	
Almoloya de Alquisiras	2	1		2	5	2.80433174
Amatepec		1			1	6.029573976
Coatepec de Harinas			1		1	6.058055378
Otzoloapan					0	0.896774959
San Sim�on de Guerrero	1				1	1.210123468
Santo Tom�as					0	1.636592257
Sultepec	1				1	5.00732895
Tejupilco		2	1	2	5	13.03949145
Temascaltepec	3	1			4	5.669103324
Texcaltitl�an		1		1	2	3.023161383
Tlatlaya					0	7.394309302
Zacazonapan					0	0.646685691
Zacualpan		1	1		2	3.194877294
Luvianos		1	1		2	5.223929241

TABLA 4.1: Casos observados y esperados de los 14 municipios del sur del Estado de M exico: 2010-2013.

Municipio	Modelo		
	SMR	BYM	GR
Almoloya de Alquisiras	1.7829560	1.6381322	0.4565709
Amatepec	0.1658492	0.3764509	0.3555433
Coatepec de Harinas	0.1650695	0.3831204	0.4167290
Otzoloapan	0	0.6925201	0.3742392
San Sim�on de Guerrero	0.8263620	1.0197758	0.4303358
Santo Tom�as	0	0.5260306	0.3088592
Sultepec	0.1997073	0.4357904	0.4184128
Tejupilco	0.3834505	0.4713516	0.4253808
Temascaltepec	0.7055790	0.7858854	0.4333984
Texcaltitl�an	0.6615591	0.8104251	0.4323398
Tlatlaya	0	0.2386125	0.2108490
Zacazonapan	0	0.7996037	0.4127009
Zacualpan	0.6260021	0.7753619	0.4337196
Luvianos	0.3828536	0.5450226	0.4178582

TABLA 4.2: Estimaciones del riesgo relativo de los 14 municipios del sur del Estado de M exico: 2010-2013.

# Capítulo 5

## Conclusiones

Uno de los propósitos de este trabajo de tesis es proporcionar al lector una visión introductoria a la metodología estadística para analizar datos espacialmente agrupados de una enfermedad rara con el objetivo de inferir la distribución espacial del riesgo e identificar regiones con riesgo alto o bajo. Se han propuesto modelos bajo el enfoque bayesiano para la construcción de los mapas porque permiten resolver algunas de las dificultades que con frecuencia aparecen en los métodos clásicos. La decisión de sólo implementar los modelos basados en Campos Aleatorios de Markov (técnica que ya era muy utilizada en el campo del procesamiento de imágenes) se debe al hecho de que ambas especificaciones representan la metodología básica de las dos corrientes del mapeo de la enfermedad: el modelo BYM suaviza las tasas de riesgo a través de un conjunto de efectos aleatorios correlacionados, mientras que el modelo GR permite discontinuidades en el espacio parametral y a su vez identifica clusters. Aunque se requiere de cómputo intensivo para la implementación del algoritmo RJMCMC, para la estimación y la selección adecuada del número de componentes en la mezcla se sugiere elegir el modelo GR, pues el modelo BYM depende fuertemente del número de vecinos (y sus correspondientes tasas de riesgo) de la región.

A pesar de que los mapas construidos expresan nítidamente el efecto del suavizamiento de los riesgos relativos, es importante tener presente el alcance de los

modelos ajustados as ı como los riesgos de la interpretaci on de los resultados. En este sentido, surge la necesidad de realizar otro tipo de investigaciones en las que se consideren factores tales como el nivel socioecon mico, las condiciones clim ticas o el perfil gen tico. Sin embargo, el dise o, la implementaci on y el an lisis de estudios controlados es muy costoso, tanto en t rminos de tiempo como en t rminos de financiamiento. No obstante, es inapropiado ignorar la gran cantidad de datos observacionales disponibles.

Por  ltimo,  sta es quiz  una de las primeras veces que se aplica metodolog a bayesiana que considere el n mero de cl sters como variable aleatoria para la construcci on de mapas de mortalidad en el Estado de M xico. La idea subsiguiente es poder construir un atlas de mortalidad para los tipos de neoplasias y enfermedades que con mayor frecuencia aparecen en el Estado, as  como realizar an lisis de tipo espacio-temporal y multivariado. Los resultados mostrados pueden emplearse como una gu a para establecer alg n tipo de pol tica preventiva en materia de salud que permita reducir las tasas de mortalidad.

# Apéndice A

## Códigos

### A.1. Mapa de mortalidad estandarizada

---

```
### Lectura y preparacion de datos
mama<-read.csv("mama.csv",header=T)
#Numero de municipios
n<-14
#Casos observados y esperados
O<-mama[,1]
E<-mama[,2]
#Tasas (SMR)
mortalidad<-O/E

### Creacion del mapa
library(maptools)
library(RColorBrewer)
suredomex<-readShapeSpatial("Shapes/suredomex.shp")
#Asignacion de tasas
suredomex$color<-array(0,dim=n)
for(i in 1:n){
  for(j in 1:n){
    numerico<-as.integer(suredomex$CVE_MUN[j])
    if(numerico==i) suredomex$color[j]<-mortalidad[i]
  }
}
#Coloracion el mapa
brks<-c(0.01,0.5,1,1.5,2,3)
plot(spplot(suredomex,"color",at=brks,col.regions=(brewer.pal(5,"OrRd"))))
```

---

## A.2. Mapa de mortalidad: BYM

```

### Lectura y preparacion de datos
mama<-read.csv("mama.csv",header=T)
#Numero de municipios
n<-14
#Casos observados y esperados
O<-mama[,1]
E<-mama[,2]
#Vector de vecinos
adj<-c(3,7,10,13,
      7,8,11,
      1,9,10,13,
      6,12,14,
      8,9,10,
      4,
      1,2,8,10,13,
      2,5,7,9,10,12,14,
      3,5,8,10,12,
      1,3,5,7,8,9,
      2,
      4,8,9,14,
      1,3,7,
      4,8,12)
#Vector de numero de vecinos
num<-c(4,3,4,3,3,1,5,7,5,6,1,4,3,3)
#Vector de pesos
weights<-rep(1,52)

### Inferencia via OpenBUGS
library(R2OpenBUGS)
#Modelo
bym.model<-function(){
  for(i in 1:n){
    O[i]~dpois(mu[i])
    mu[i]<-lambda[i]*E[i]
    log(lambda[i])<-u[i]+v[i]
    u[i]~dnorm(0,tau_u)
  }
  v[1:n]~car.normal(adj[],weights[],num[],tau_v)
  tau_u~dgamma(0.5,0.0005)
  tau_v~dgamma(0.5,0.0005)
}
#Datos
bym.data<-list("n","O","E","adj","num","weights")
#Iniciales
bym.inits<-function(){list(u=rep(0,dim=n),v=rep(0,n),tau_u=1,tau_v=1)}
#Parametros de interes

```

```

bym.parameters<-c("lambda")
#Inferencia bajo el modelo BYM
bym<-bugs(data=bym.data,inits=bym.inits,parameters.to.save=bym.parameters,model.
  file=bym.model,n.chains=1,n.burnin=50000,n.iter=100000)

### Creacion del mapa
library(maptools)
library(RColorBrewer)
suredomex<-readShapeSpatial("Shapes/suredomex.shp")
#Asignacion de tasas
mortalidad<-bym$mean$lambda
for(i in 1:n){
  for(j in 1:n){
    numerico<-as.integer(suredomex$CVE_MUN[j])
    if(numerico==i) suredomex$color[j]<-mortalidad[i]
  }
}
#Coloracion el mapa
brks<-c(0,0.5,1,1.25,1.5,2)
plot(spplot(suredomex,"color",at=brks,col.regions=(brewer.pal(6,"OrRd"))))

```

### A.3. Mapa de mortalidad: GR

```

##### Lectura y preparacion de datos
#Numero de municipios
n<-14
#Casos observados y esperados
mama<-read.csv("mama.csv",header=T)
O<-mama$Obs
E<-mama$Esp
#Matriz de pesos
w<-as.matrix(read.csv("pesos.csv",header=F))
#Numero de vecinos
n_i<-c(4,3,4,3,3,1,5,7,5,6,1,4,3,3)
#Matriz de vecinos
v<-as.matrix(read.csv("vecinos.csv",header=F))
#Matriz de constantes de normalizacion
theta<-as.matrix(read.csv("thetas.csv",header=F))

##### Modelo Potts
### Estadistico U
U<-function(z){
  suma<-0

```

```

for(a in 1:(n-1)){
  for(b in (a+1):n){
    if((w[a,b]==1) && (z[a]==z[b])) suma<-suma+1
  }
}
return(suma)
}

### Constante de normalizacion
library(partitions)
library(doParallel)
detectCores()
cl<-makeCluster(8)
registerDoParallel(8)
getDoParWorkers()
constante<-function(k,psi){
  particiones<-restrictedparts(n,k,include.zero=F)
  np<-R(k,n,include.zero=F)
  configuraciones<-foreach(c=1:np,.combine=cbind) %dopar% {
    return(setparts(particiones[,c]))
  }
  casos<-length(configuraciones[1,])
  kernel<-numeric(casos)
  for(d in 1:casos){
    kernel[d]<-exp(psi*U(configuraciones[,d]))
  }
  theta<-log(sum(kernel))
  return(theta)
}

### Medida de probabilidad
potts<-function(k,psi,z){
  return( exp(psi*U(z)-theta[k,10*(psi+0.1)]) )
}

##### Especificacion de probabilidades
### Probabilidad inicial para k
p_k<-c(0.1,0.1,0.7,0.1)
### Constantes
a<-1
b<-sum(E)/sum(O)
c<-0.1
### Distribucion conjunta
joint<-function(lambda,z,psi,k){
  p1<-1
  for(i in 1:n){p1<-p1*dpois(O[i],lambda[z[i]]*E[i])}
  p2<-factorial(k)*prod(dgamma(lambda,shape=a,rate=b))
  p3<-potts(k,psi,z)*p_k[k]
  p<-p1*p2*p3
  return(p)
}

```

```

}
### Probabilidades de transicion
P<-matrix(c(1/2,1/3,0,0,1/2,1/3,1/3,0,0,1/3,1/3,1/2,0,0,1/3,1/2),nrow=4,ncol=4)

##### Inferencia via RJMCMC
### Numero de iteraciones
m<-1000000
### Parametros de interes
k<-array(dim=m)
psi<-array(dim=m)
z<-array(dim=c(n,m))
lambda<-array(dim=c(n,m))
### Iniciales
k[1]<-3
psi[1]<-1
z[,1]<-cbind(3,1,1,1,2,1,1,1,2,2,1,1,2,1)
tasa<-c(0.5,1,1.5)
lambda[,1]<-cbind(tasa[z[,1]])
### Cadenas
for(t in 1:(m-1)){
  move<-switch(k[t],
               sample(c("stay","split"),1),
               sample(c("merge","stay","split"),1),
               sample(c("merge","stay","split"),1),
               sample(c("merge","stay"),1))
  if(move=="split"){
    zp<-cbind(rep(0,n))
    pesos<-array(dim=k[t])
    for(q in 1:k[t]){ pesos[q]<-sum(z[,t]==q) }
    j<-sample(which(pesos!=1),1)
    u<-runif(1)
    tasa_menos<-tasa[j]*(u^c)
    tasa_mas<-tasa[j]*(u^-c)
    p_alloc<-1
    if(k[t]==1){
      for(r in 1:n){
        n_menos<-sum(zp[v[r,1:n_i[r]]]==j)
        n_mas<-sum(zp[v[r,1:n_i[r]]]==j+1)
        menos<-exp(psi[t]*n_menos-tasa_menos*E[r])*(tasa_menos^0[r])
        mas<-exp(psi[t]*n_mas-tasa_mas*E[r])*(tasa_mas^0[r])
        zp[r]<-sample(c(j,j+1),1,prob=c(menos,mas))
        if(zp[r]==j) alloc<-menos/(menos+mas) else alloc<-mas/(menos+mas)
        p_alloc<-p_alloc*alloc
      }
    }
  }
  else{
    if(j==1){
      if(tasa[j+1]<tasa_mas){

```

```

k[t+1]=k[t]
psi[t+1]=psi[t]
z[,t+1]=z[,t]
lambda[,t+1]=lambda[,t]
next
}
else{
for(r in which(z[,t]==j)){
n_menos<-sum(zp[v[r,1:n_i[r]]]==j)
n_mas<-sum(zp[v[r,1:n_i[r]]]==j+1)
menos<-exp(psi[t]*n_menos-tasa_menos*E[r])*(tasa_menos^0[r])
mas<-exp(psi[t]*n_mas-tasa_mas*E[r])*(tasa_mas^0[r])
zp[r]<-sample(c(j,j+1),1,prob=c(menos,mas))
if(zp[r]==j) alloc<-menos/(menos+mas) else alloc<-mas/(menos+mas)
p_alloc<-p_alloc*alloc
}
for(r in which(z[,t]>j)){ zp[r]<-z[r,t]+1 }
}
} else
if(j==k[t]){
if(tasa_menos<tasa[j-1]){
k[t+1]=k[t]
psi[t+1]=psi[t]
z[,t+1]=z[,t]
lambda[,t+1]=lambda[,t]
next
}
} else{
for(r in which(z[,t]<j)){ zp[r]<-z[r,t] }
for(r in which(z[,t]==j)){
n_menos<-sum(zp[v[r,1:n_i[r]]]==j)
n_mas<-sum(zp[v[r,1:n_i[r]]]==j+1)
menos<-exp(psi[t]*n_menos-tasa_menos*E[r])*(tasa_menos^0[r])
mas<-exp(psi[t]*n_mas-tasa_mas*E[r])*(tasa_mas^0[r])
zp[r]<-sample(c(j,j+1),1,prob=c(menos,mas))
if(zp[r]==j) alloc<-menos/(menos+mas) else alloc<-mas/(menos+mas)
p_alloc<-p_alloc*alloc
}
}
}
else{
if( (tasa_menos<tasa[j-1]) | (tasa[j+1]<tasa_mas) ){
k[t+1]=k[t]
psi[t+1]=psi[t]
z[,t+1]=z[,t]
lambda[,t+1]=lambda[,t]
next
}
}

```

```

else{
  for(r in which(z[,t]<j)){ zp[r]<-z[r,t] }
  for(r in which(z[,t]==j)){
    n_menos<-sum(zp[v[r,1:n_i[r]]]==j)
    n_mas<-sum(zp[v[r,1:n_i[r]]]==j+1)
    menos<-exp(psi[t]*n_menos-tasa_menos*E[r])*(tasa_menos^0[r])
    mas<-exp(psi[t]*n_mas-tasa_mas*E[r])*(tasa_mas^0[r])
    zp[r]<-sample(c(j,j+1),1,prob=c(menos,mas))
    if(zp[r]==j) alloc<-menos/(menos+mas) else alloc<-mas/(menos+mas)
    p_alloc<-p_alloc*alloc
  }
  for(r in which(z[,t]>j)){ zp[r]<-z[r,t]+1 }
}
}
}
tasap<-sort(c(tasa[-j],tasa_menos,tasa_mas))
r1<-joint(tasap,zp,psi[t],k[t]+1)/joint(tasa,z[,t],psi[t],k[t])
r2<-P[k[t]+1,k[t]]/(P[k[t],k[t]+1]*p_alloc)
r3<-(2*c*tasa[j])/u
R<-r1*r2*r3
alpha_z<-min(1,R)
u_z<-runif(1)
if(u_z<=alpha_z){
  k[t+1]=k[t]+1
  psi[t+1]=psi[t]
  z[,t+1]=zp
  tasa<-tasap
  lambda[,t+1]=rbind(tasap[z[,t+1]])
}
else{
  k[t+1]=k[t]
  psi[t+1]=psi[t]
  z[,t+1]=z[,t]
  lambda[,t+1]=lambda[,t]
}
} else
if(move=="merge"){
  j<-sample(c(1:(k[t]-1)),1)
  zp<-cbind(rep(0,n))
  tasa_j<-sqrt(tasa[j]*tasa[j+1])
  u_j<-(tasa[j]/tasa[j+1])^(1/(2*c))
  for(r in which(z[,t]<=j)){ zp[r]<-z[r,t] }
  for(r in which(z[,t]>j)){ zp[r]<-z[r,t]-1 }
  tasap<-sort(c(tasa[-c(j,j+1)],tasa_j))
  r1<-joint(tasap,zp,psi[t],k[t]-1)/joint(tasa,z[,t],psi[t],k[t])
  r2<-P[k[t]-1,k[t]]/(P[k[t],k[t]-1])
  r3<-u_j/(2*c*tasa_j)
R<-r1*r2*r3

```

```

alpha_z<-min(1,R)
u_z<-runif(1)
if(u_z<=alpha_z){
  k[t+1]=k[t]-1
  psi[t+1]=psi[t]
  z[,t+1]=zp
  tasa<-tasap
  lambda[,t+1]=rbind(tasap[z[,t+1]])
}
else{
  k[t+1]=k[t]
  psi[t+1]=psi[t]
  z[,t+1]=z[,t]
  lambda[,t+1]=lambda[,t]
}
} else
if(move=="stay"){
  #Parametro psi
  if(psi[t]<0.1){ psip=0.1 } else
  if(psi[t]>1.9){ psip=1.9 } else{ psip<-sample(c(psi[t]-0.1,psi[t]+0.1),1) }
  alpha_psi<-min(1,potts(k[t],psip,z[,t])/potts(k[t],psi[t],z[,t]))
  u_psi<-runif(1)
  if(u_psi<=alpha_psi){
    psi[t+1]=psip
  }
  else{
    psi[t+1]=psi[t]
  }
  #Parametro lambda
  tasap<-sort(tasa*exp(rnorm(k[t])))
  mh<-array(dim=k[t])
  for(s in 1:k[t]){
    mh[s]<-((tasap[s]/tasa[s])^(a+sum(0[which(z[,t]==s)])))*exp(-(tasap[s]-tasa
[s])*(b+sum(E[which(z[,t]==s)])))
  }
  alpha_lambda<-min(1,prod(mh))
  u_lambda<-runif(1)
  if(u_lambda<=alpha_lambda){
    tasa<-tasap
    lambda[,t+1]<-rbind(tasap[z[,t]])
  }
  else{
    lambda[,t+1]<-rbind(tasa[z[,t]])
  }
  #Parametro z
  for(i in 1:n){
    n_ij<-array(dim=k[t])
    p_zi<-array(dim=k[t])

```

```
    for(j in 1:k[t]){
      n_ij[j] <- sum(z[which(w[i,]==1),t]==j)
      p_zi[j] <- exp(-tasa[j]*E[i])*(tasa[j]^0[i])*exp(psi[t+1]*n_ij[j])
    }
    z[i,t+1] <- sample(c(1:k[t]),1,prob=p_zi)
  }
  k[t+1]=k[t]
}
}

##### Creacion del mapa
library(maptools)
library(RColorBrewer)
suredomex <- readShapeSpatial("Shapes/suredomex.shp")
#Asignacion de tasas
mortalidad <- rowMeans(lambda[,100001:m])
for(i in 1:n){
  for(j in 1:n){
    numerico <- as.integer(suredomex$CVE_MUN[j])
    if(numerico==i) suredomex$color[j] <- mortalidad[i]
  }
}
#Coloracion el mapa
brks <- c(0,0.3,0.4,0.5,1)
plot(spplot(suredomex,"color",at=brks,col.regions=(brewer.pal(5,"OrRd"))))
```

---



# Bibliografía

- Bernardinelli, L., Clayton, D., y Montomoli, C. (1995). Bayesian estimates of disease maps: How important are priors? *Statistics in Medicine*, 14(21-22):2411–2431.
- Besag, J. (1974). Spatial Interaction and the statistical analysis of lattice systems (with discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 36(2):192–236.
- Besag, J., York, J., y Mollié, A. (1991). Bayesian image restoration, with two applications in spatial statistics (with discussion). *Annals of the Institute of Mathematical Statistics*, 43(1):1–59.
- Best, N., Richardson, S., y Thomson, A. (2005). A comparison of Bayesian spatial models for disease mapping. *Statistical Methods in Medical Research*, 14:35–59.
- Clayton, D. y Kaldor, J. (1987). Empirical Bayes estimates of age-standardized relative risks for use in disease mapping. *Biometrics*, 43:671–681.
- Cressie, N. (1993). *Statistics for Spatial Data*. Wiley.
- Denison, D. G. T. y Holmes, C. C. (2001). Bayesian partitioning for estimating disease risk. *Biometrics*, 57(1):143–149.
- Elliott, P., Wakefield, J. C., Best, N. G., y Briggs, D. J. (2000). *Spatial Epidemiology: Methods and Applications*. Oxford University Press.
- Fay, R. E. y Herriot, R. A. (1979). Estimates of income for small places: An application of James-Stein procedures to census data. *Journal of the American Statistical Association*, 74:269–277.

- Fernández, C. y Green, P. J. (2002). Modelling spatially correlated data via mixtures: a Bayesian approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(4):805–826.
- Gelfand, A. E. y Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85(410):389–409.
- Geman, S. y Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(6):721–741.
- Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, 82(4):711–732.
- Green, P. J. y Richardson, S. (2002). Hidden Markov models and disease mapping. *Journal of the American Statistical Association*, 97:1055–1070.
- Gutiérrez-Peña, E. (1997). Métodos Computacionales en la Inferencia Bayesiana. En *Serie Monografías*, volumen 6-15. IIMAS, UNAM.
- Gutiérrez-Peña, E. (1998). Análisis Bayesiano de Modelos Jerárquicos Lineales. En *Serie Monografías*, volumen 7-16. IIMAS, UNAM.
- Hammersley, J. y Clifford, P. (1971). Markov fields on finite graphs and lattices. No publicada.
- Hastings, W. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1):97–109.
- Knorr-Held, L. y Becker, N. (2000). Bayesian modelling of spatial heterogeneity in disease maps with application to German cancer mortality data. *Journal of the German Statistical Society*, 84(2):121–140.
- Knorr-Held, L. y Raßer, G. (2000). Bayesian detection of clusters and discontinuities in disease maps. *Biometrics*, 56(1):13–21.

- Lawson, A. B. (2013). *Bayesian Disease Mapping: Hierarchical Modeling in Spatial Epidemiology*. Chapman & Hall/CRC Press, 2.<sup>a</sup> edición.
- López-Díaz, M. A., Aguirre-Pérez, R., y Mendéz-Villalobos, P. (2010). Estimación Bayesiana de tasas de mortalidad en regiones pequeñas con dependencia espacial. En *Aportaciones y Aplicaciones de la Probabilidad y la Estadística*, págs. 44–59. Benemérita Universidad Autónoma de Puebla.
- López-Moreno, S., Garrido-Latorre, F., y Hernández-Avila, M. (2000). Desarrollo histórico de la epidemiología: su formación como disciplina científica. *Salud Pública Méx*, 42(2):133–143.
- Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A., y Teller, E. (1953). Equation of state calculations by fast computing machines. *Journal of Chemical Physics*, 21(6):1087–1092.
- Meyn, S. y Tweedie, R. (2009). *Markov Chains and Stochastic Stability*. Cambridge University Press, 2.<sup>a</sup> edición.
- Mollié, A. (1996). Bayesian mapping of disease. En *Markov Chain Monte Carlo in Practice*, Capítulo 20, págs. 359–379. Chapman & Hall/CRC Press.
- Moreno-Altamirano, A., López-Moreno, S., y Corcho-Berdugo, A. (2000). Principales medidas en epidemiología. *Salud Pública Méx*, 42(4):337–348.
- Nieto-Barajas, L. E. (2008). A Markov gamma random field for modelling disease mapping data. *Statistical Modelling*, 8(1):97–114.
- R Core Team (2015). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org>.
- Richardson, S. y Green, P. J. (1997). On Bayesian analysis of mixtures with an unknown number of components (with discussion). *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 59(4):731–792.
- Schoenbach, V. J. (1999). Standardization of rates and ratios. <http://www.epidemiolog.net/evolving/Standardization.pdf>[Revisado 10/2014].

- Silva-Aycaguer, L., Benavides-Rodríguez, A., y Vidal-Rodeiro, C. (2003). Análisis espacial de la mortalidad en áreas geográficas pequeñas. El enfoque bayesiano. *Rev Cubana Salud Pública*, 29(4):314–322.
- Snow, J. (1955). *On the Mode of Communication of Cholera*. London:John Churchill, 2.<sup>a</sup> edición.
- The BUGS project (2014). *BUGS (Bayesian inference Using Gibbs Sampling)*. <http://www.openbugs.net>.
- Torres-Sánchez, L. E., Rojas-Martínez, R., Escamilla-Núñez, C., De la Vara-Salazar, E., y Lazcano-Ponce, E. (2014). Tendencias en la mortalidad por cáncer en México de 1980 a 2011. *Salud Pública de México*, 56(5):473–491.
- Waller, L. y Carlin, B. (2010). Disease Mapping. En *Handbook of Spatial Statistics*, Capítulo 14, págs. 217–243. Chapman & Hall/CRC Press.