



Centro de Investigación en Matemáticas, A.C.

CIMAT

Estimación Máximo Verosímil en Modelos Espaciales

T E S I S

Que para obtener el grado de:

Maestro en ciencias

Con orientación en:

Probabilidad y Estadística

P r e s e n t a:

Erick Alberto Cecilio Ayala

Guajuato, Gto. México. Agosto de 2010

Contenido

1	Introducción	3
	Introducción	3
2	Modelos Espaciales	8
2.1	Un Modelo Espacial para datos discretos	12
2.2	Un Modelo Espacial para datos continuos	13
3	Estimación de Parámetros para Modelos en Látices	15
3.1	Máxima verosimilitud vía MCMC	17
3.2	Estimación vía ABC	19
3.3	Máxima Verosimilitud Aproximada	21
3.4	Un Método Bayesiano para Estimación Espacial	22
4	Modelo Ising	25
4.1	Máxima verosimilitud	25
4.2	Máxima verosimilitud vía MCMC	28
4.3	Estimación vía ABC	33
4.4	Máxima Verosimilitud Aproximada	34

5	Modelo Normal	36
5.1	Máxima verosimilitud	36
5.2	Máxima verosimilitud vía MCMC	38
5.3	Estimación vía ABC	42
5.4	Máxima Verosimilitud Aproximada	43
6	Conclusiones	45
	ANEXOS	47
I	<i>ANEXOS</i>	47
6.1	ANEXO A	48
6.2	ANEXO B	56
6.3	ANEXO C	58
	Bibliografía	60

Agradecimientos

A mi familia, por la confianza depositada en mí.

Al Consejo Nacional de Ciencia y Tecnología por la beca otorgada para la obtención del grado de: Maestro en ciencias con orientación en Probabilidad y Estadística.

Al Consejo de Ciencia y Tecnología del Estado de Guanajuato por la beca otorgada para la finalización de la tesis.

Al Dr. Rogelio Ramos Quiroga por su importante colaboración en la realización de la tesis.

Al Centro de Investigación en Matemáticas A. C.

Capítulo 1

Introducción

La estadística espacial se divide en tres campos de aplicación principales, trabajando con datos en geoestadística, reticulares y patrones puntuales. Estas áreas de acción son clasificadas de acuerdo al tipo de observaciones encontradas. En general, los modelos estadísticos tratan de hacer sentido a un conjunto de datos: Diseño, inferencia, y diagnóstico son consecuencias naturales de la simbiosis de datos-modelo, y juegan un papel importante en Estadística para datos espaciales.

Recientemente ha habido un creciente interés en la modelación de datos espaciales con interacciones entre puntos, los que incluyen procesos Poisson espaciales no homogéneos (Baddeley, 2000), modelos reticulares espaciales (Besag, 1974; Strauss, 1977) y algunos modelos de interacción pareada (Besag, 1974; Ripley, 1977; Diggle et al., 1994). Para una introducción general a la metodología estadística para modelos espaciales, véase Ripley (1981), Diggle (1983), Stoyan *et al.* (1987), Cressie (1993) y Barndorff-Nielsen *et al.* (1999).

La Geoestadística (Matheron, 1962, 1963a, 1963b) surgió a principios de los 60's como una disciplina híbrida de ingeniería minera, geología, matemáticas y estadística. Su fortaleza, más que las aproximaciones clásicas de estimación en reservas mineras, es que reconoce variabilidad espacial en ambas escalas

(grande y pequeña), o en el lenguaje estadístico, modela tendencia espacial y correlación espacial. Los métodos de tendencia en superficies (p. ej. Whitten, 1970) incluyen sólo variación en gran escala, suponiendo errores independientes. Uno de los problemas más importantes en geoestadística es predecir el grado del mineral en una sección minera basado en muestras observadas.

Una retícula de ubicaciones evoca una idea de puntos regularmente espaciados en \mathbb{R}^d , conectados a vecinos cercanos, vecinos cercanos de segundo orden, y así sucesivamente. En este trabajo de tesis se estudiará este tipo de datos. El monitoreo remoto desde satélites ofrece una manera eficiente para la recolección de datos. Por ejemplo, nos ofrece información para ser recolectada rápidamente en patrones de clima, distribución mineral, y superficie de cultivos sin tener que realizar estudios tradicionales largos y tediosos. Los satélites giran alrededor de la Tierra y reciben datos en forma de ondas electromagnéticas reflectantes a un número de frecuencias, incluyendo aquellas que se encuentran en la parte visible del espectro. Por varios muestreos y métodos de integración, la superficie de la Tierra es dividida en pequeños rectángulos (p. ej. 56 m \times 56 m) llamados *pixeles* (elemento pequeño de fotografía). Un escenario agrícola de interés (alrededor de 34,000 Km²) tiene ciertas proporciones dedicadas a trigo, maíz, soya y así sucesivamente, que necesitan de ser estimados. Estos distintos cultivos tienen sus propias propiedades de reflectancia y son registrados como una retícula regular en \mathbb{R}^2 (despreciando la curvatura de la Tierra) y son identificados con el centro de sus pixeles respectivos. Los modelos estadísticos para estos datos necesitan expresar el hecho de que observaciones cercanas (en tiempo o espacio) tienden a ser parecidas. A veces se puede esperar que observaciones cercanas tiendan a ser distintas. La competencia entre plantas por la luz y nutrientes del suelo, pueden conducir a plantas grandes saludables rodeadas por plantas menos resistentes. En contraste a los problemas geoestadísticos, los datos de problemas de retículas pueden ser exhaustivos del fenómeno. Ciertamente, el

muestreo es posible; por ejemplo, si suponemos que tenemos disponible sólo una pequeña ventana del conjunto completo de los datos.

Los patrones puntuales surgieron cuando la variable importante a ser analizada es la ubicación de "eventos". Frecuentemente, la primer pregunta a ser contestada es si el patrón está exhibiendo aleatoriedad espacial completa, aglomeraciones, o regularidad. Por ejemplo, consideramos las ubicaciones de pinos en un bosque natural. ¿Cuál es el significado biológico de la aglomeración de estos árboles? La variable 'diámetro a la altura del pecho', se registra junto con la ubicación del árbol. ¿Son grandes (pequeños) los conjuntos de árboles y cómo interactúan árboles grandes y pequeños? La variable 'tamaño' es llamada comúnmente variable de marca, y el proceso completo es llamado proceso puntual espacial marcado.

El interés primario de los físicos es el comportamiento macroscópico de sistemas de partículas, usualmente modeladas como pertenecientes a una retícula, donde cada partícula tiene un estado asociado continuo o discreto, el cual es especificado estocásticamente a través de una función potencial. Las interacciones en la función potencial pueden estar localizadas espacialmente y todavía resultar una correlación no nula entre los estados de partículas infinitamente lejanos. El ejemplo más famoso es el modelo Ising de dos estados, el cual estudiaremos en esta tesis.

Obviamente, en la práctica, una simulación sólo se puede realizar sobre retículas finitas y, en este contexto, no hay una ventaja general en restringir la atención a arreglos regulares de sitios. Además, para cualquier distribución conjunta (o Gibbs), especificada por una función potencial finita, ahora es bien conocido que hay una definición equivalente a través de características locales del sistema, por lo que se entiende como la distribución condicional de la variable aleatoria (estado) en cada sitio, dados los valores de todos los demás sitios. Esta especificación alternativa se denomina campo aleatorio de

Markov y es referido ampliamente en la literatura estadística como el teorema de Hammersley-Clifford; véase por ejemplo Besag (1974) y, para una perspectiva histórica, Clifford (1990). El hecho de que no hay métodos directos usables para simular distribuciones generales multivariadas sugiere la idea que algoritmos satisfactorios podrían estar basados en las características locales univariadas correspondientes. Por lo tanto, son estos hechos los que sugieren el uso de algoritmos de Metropolis - Hastings, descritos en el artículo de Smith y Roberts (1993). En particular, el muestreo Gibbs, el cual actualiza sucesivamente cada componente de acuerdo a su característica local actual, garantiza la convergencia a la distribución conjunta correspondiente bajo condiciones generales, esencialmente para los cuales la expansión de Brook (Besag, 1974) es válida.

Para realizar inferencia en modelos espaciales (de hecho, para cualquier modelo estadístico), en general se busca hacer estimaciones vía máxima verosimilitud, pero en nuestro caso los datos espaciales, vienen en arreglos matriciales y uno de los problemas que se encuentra regularmente en la función de verosimilitud, es al considerar la constante de normalización. Esta constante de normalización es una integral sobre todas las matrices posibles de datos espaciales del fenómeno en cuestión. Debido a la función de verosimilitud intratable, la estimación por máxima verosimilitud es raramente usada para modelos espaciales, en su lugar el estimador de máxima pseudoverosimilitud para modelos espaciales fue propuesto como una alternativa (Besag, 1977; Baddeley and Turner, 2000). Sin embargo, el estimador de máxima pseudo verosimilitud es ineficiente comparado con el estimador de máxima verosimilitud.

Uno de los objetivos de esta tesis es realizar una comparación computacional entre los métodos de estimación que existen hasta hoy en estadística espacial, y el otro, el hacer notar que cualquier estimación usada en modelos espaciales está sujeta al desarrollo tecnológico de nuestros tiempos.

En el Capítulo 1 se muestra una forma natural para realizar modelaciones espaciales, se enuncian dos teoremas fundamentales que hay que tener en cuenta al realizar la modelación. También se presentan dos modelos (uno discreto y otro continuo), que son los mas usuales en el estudio de datos espaciales, y los cuales serán utilizados en esta tesis. En el Capítulo 2 se presentan los métodos de estimación espacial que serán estudiados y comparados para los dos modelos espaciales mencionados en el Capítulo 1. En los Capítulos 3 y 4 se hacen análisis comparativos basados en simulaciones de los métodos de estimación mencionados en el Capítulo 2, para los modelos de Ising y Normal respectivamente. Finalmente, en el Capítulo 5 ofrecemos las conclusiones de nuestro trabajo.

Todos los cálculos fueron realizados en una computadora personal con procesador Core i7. Para los cálculos y la graficación se utilizó R. El código está disponible escribiendo un correo a erick@cimat.mx.

Capítulo 2

Modelos Espaciales

Hablando en términos computacionales, los campos de Markov son relativamente nuevos en escenarios matemáticos y estadísticos, pero han sido estudiados extensamente durante los últimos 50 años en física estadística, donde son conocidos como conjuntos de Gibbs. La distribución de Gibbs postula que configuraciones de átomos (realizaciones) en una retícula ocurren con frecuencia proporcional a $\exp\{kQ(\mathbf{x})/T\}$, donde k es la constante de Boltzmann y T es la temperatura absoluta. $Q(\mathbf{x})$ describe la energía asociada con configuraciones de \mathbf{x} y, comúnmente, es una suma de contribuciones de energía de varias interacciones entre los átomos. En el caso binario, los conjuntos de Gibbs también son conocidos como modelos de *Ising*. Son importantes en física teórica porque (para dimensiones mayores de 1) exhiben transiciones de fase y por lo tanto explican un fenómeno cooperativo global, como la formación de cristales y magnetismo, en términos de reglas locales (estocásticas) entre átomos individuales.

Los datos pueden ser continuos o discretos, estos pueden ser observaciones espaciales o puntos en el espacio, sus ubicaciones espaciales pueden ser regulares o irregulares, estas ubicaciones pueden ser de un conjunto espacial continuo o discreto.

Sea $y = (y_1, \dots, y_n)$, donde y_i es la observación en la i -ésima localidad. Una

forma natural de modelación es mediante la consideración de relaciones condicionales:

$$P(y_i | y_1, \dots, y_{i-1}, y_{i+1}, \dots, y_n)$$

Dado un conjunto de probabilidades condicionales, la distribución conjunta, $P(y)$, puede obtenerse mediante un teorema de factorización (Brook, (1964)).

Teorema de factorización: Sea $x = (x_1, \dots, x_n)$, otro conjunto de posibles valores para y , entonces:

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

De aquí podemos ver que para especificar la distribución conjunta $P(y)$ tenemos que calcular la constante de normalización:

$$P(x) = \left(\sum_{y_1} \dots \sum_{y_n} \prod_{i=1}^n \frac{P(y_i | y_1, \dots, y_{i-1}, x_{i+1}, x_n)}{P(x_i | y_1, \dots, y_{i-1}, x_{i+1}, x_n)} \right)^{-1},$$

la cual representa un problema computacional. Debemos notar que este teorema supone la existencia de la distribución conjunta y después se dá la factorización condicional antes mencionada. En general, si tenemos densidades condicionales, no necesariamente podemos encontrar la conjunta; por ejemplo, consideremos las siguientes densidades condicionales

$$f_{X|Y}(x|y) = yI(0 < x < y^{-1}) I(y > 0),$$

y

$$f_{Y|X}(y|x) = xI(0 < y < x^{-1}) I(x > 0),$$

puede verse que no existe distribución conjunta con tales distribuciones condicionales (Arnold, Castillo y Sarabia, 1999).

Sin embargo, hay un resultado que toma en consideración este problema, y es el teorema de Hammersley-Clifford. Este teorema es esencial para la construcción de esquemas espaciales válidos a través del enfoque de probabilidades condicionales. Comenzaremos dando una serie de definiciones y resultados para enunciar el teorema.

Condición de Positividad.

Sean $\{i : i = 1, \dots, n\}$ las ubicaciones en una retícula en las cuales una variable discreta Z es observada. Definimos $\varsigma \equiv \{z : \Pr(z) > 0\}$ y $\varsigma_i \equiv \{z_i : \Pr(z_i) > 0\}$, $i = 1, \dots, n$, donde z es una observación de Z y z_i es la observación en la ubicación i de Z . Entonces se dice que hay condición de positividad cuando se satisface $\varsigma = \varsigma_1 \times \dots \times \varsigma_n$. Para una variable continua Z , se aplica la misma definición excepto que $\Pr(\cdot)$ se reemplaza con $f(\cdot)$, que representa las densidades conjunta y marginal.

Definición 1. Un sitio k es definido como un *vecino* del sitio i si la distribución condicional de x_i , dados los demás valores de los sitios depende funcionalmente de x_k , para $k \neq i$. Definimos también

$$N_i \equiv \{k : k \text{ es un vecino de } i\} \tag{2.1}$$

como la vecindad del sitio i .

Definición 2. Una *clique* es definida como el conjunto de los sitios que consiste de un sitio o sitios que son vecinos unos de otros.

Definición 3. Cualquier medida de probabilidad cuyas distribuciones condicionales definen una estructura de vecindad $\{N_i : i = 1, \dots, n\}$ a través de (1.1) se define como un *campo de Markov aleatorio*.

Sin pérdida de generalidad, supongamos que el cero puede ocurrir en cada sitio; esto es, $0 \in \Omega$. Definimos

$$Q(z) \equiv \log \{\Pr(z) / \Pr(0)\}, \quad z \in \Omega. \tag{2.2}$$

Entonces, conocer $Q(\cdot)$ es equivalente a conocer $\Pr(\cdot)$, porque

$$\Pr(z) = \exp(Q(z)) / \sum_{y \in \Omega} \exp(Q(y))$$

en el caso discreto. Una función análoga Q en el caso continuo puede ser definida remplazando $\Pr(\cdot)$ por $f(\cdot)$ y \sum por \int . De aquí en adelante, nos referiremos a Q como la *función potencial*.

Proposición 1. La función potencial Q satisface las siguientes dos propiedades:

1. $\frac{\Pr(z_i/\{z_j : j \neq i\})}{\Pr(0_i/\{z_j : j \neq i\})} = \frac{\Pr(z)}{\Pr(\mathbf{z}_i)} = \exp(Q(z) - Q(\mathbf{z}_i)),$

donde 0_i denota el evento $x_i = 0$ y $\mathbf{z}_i \equiv (z_1, \dots, z_{i-1}, 0, z_{i+1}, \dots, z_n)$.

2. Q puede ser expandida de manera única en Ω como

$$\begin{aligned} Q(z) = & \sum_{1 \leq i \leq n} z_i G_i(z_i) + \sum_{1 \leq i < j \leq n} z_i z_j G_{ij}(z_i, z_j) + \sum_{1 \leq i < j < k \leq n} z_i z_j z_k G_{ijk}(z_i, z_j, z_k) + \dots \\ & + z_1 \cdots z_n G_{1\dots n}(z_1 \cdots z_n), \quad z \in \Omega. \end{aligned} \tag{2.3}$$

donde G_{ij} es una función que nos dá la relación entre las observaciones de las ubicaciones i, j .

Recordemos que $\Pr(z)$ (o $f(z)$) es proporcional a $\exp(Q(z))$. Cómo la verosimilitud es $\Pr(z)$, expresada como una función de los parámetros del proceso aleatorio, es claro que es importante encontrar la constante de proporcionalidad (forma cerrada) como una función de estos parámetros. Esto no siempre es posible. Sin embargo, hay un resultado muy fuerte disponible con respecto a la forma que debe tomar la función Q :

Teorema (Hammersley - Clifford) (*Hammersley y Clifford, 1971*) *Supongamos que Z se distribuye de acuerdo a un campo aleatorio de Markov en Ω que satisface la condición de positividad. Entonces, la función potencial $Q(\cdot)$ dada por (1.3) debe satisfacer la propiedad*

si los sitios i, j, \dots, s no forman una clique, entonces $G_{ij\dots s}(\cdot) \equiv 0$,

donde los cliques están definidos por la estructura de vecindad $\{N_i : i = 1, \dots, n\}$.

La prueba de este resultado se encuentra en el trabajo realizado por Besag (1974), que básicamente considera la expresión $Q(z) - Q(\mathbf{z}_1)$, la desarrolla de acuerdo a sus definiciones correspondientes a (1.3), y supone que un sitio no es vecino del sitio 1; a base de desarrollo algebraico se llega al resultado deseado.

2.1 Un Modelo Espacial para datos discretos

El modelo de Ising fue uno de los primeros en aparecer, es un modelo físico propuesto para estudiar el comportamiento de materiales ferromagnéticos. Se trata de un modelo paradigmático de la Mecánica Estadística pues es un modelo matemático que incorpora transiciones de fase las cuales son observables en el mundo físico.

El modelo clásico de Ising está definido por la distribución conjunta dada por

$$f(\mathbf{x}; \beta) = \exp[-Q(\mathbf{x}; \beta) - \log\{C(\beta)\}]$$

donde $Q(\mathbf{X}; \beta) = -\beta V(\mathbf{x})$ y

$$V(\mathbf{x}) = \sum_{k=1}^{n-1} \sum_{l=1}^n x_{kl} x_{k+1,l} + \sum_{k=1}^n \sum_{l=1}^{n-1} x_{kl} x_{k,l+1} + \sum_{k=1}^n x_{k1} x_{k,n}$$

β representa el grado de interacción entre las variables y $V(\mathbf{x})$ representa la estructura de vecinos la cual, en este caso, corresponde a una configuración de primer orden. El modelo de Ising tiene distribuciones condicionales de acuerdo a un modelo autolístico homogéneo de primer orden en una retícula regular contable y se reducen a expresiones de la forma (ANEXO B)

$$\Pr(x_{i,j} | V_{i,j}) = \frac{\exp(g)}{1 + \exp(g)}, \quad (2.4)$$

donde

$$g \equiv \alpha + \gamma_2 (x_{i-1,j} + x_{i+1,j}) + \gamma_1 (x_{i,j-1} + x_{i,j+1}),$$

$i = 1, \dots, n$ y $j = 1, \dots, m$ donde las x 's toman los valores 0 y 1. En este caso $V_{i,j}$ representa la estructura de vecinos cercanos a la observación $x_{i,j}$.

Es decir, hay nm partículas colocadas en una matriz rectangular. Cada partícula puede apuntar sólo en dos sentidos, arriba o abajo. Cada una de esas orientaciones se llama *espín de la partícula*. El sentido del espín queda determinado mediante la interacción de la partícula con sus vecinas y por fluctuaciones térmicas. El modelo fue presentado originalmente por Ising (1925) bajo la suposición adicional de que $\gamma_1 = \gamma_2$. Pickard (1987) muestra que, para $\gamma_1 = \gamma_2 = \gamma$, existen valores críticos de γ , por debajo de los cuales la correlación de los vecinos mas cercanos es cero, pero arriba de ellos el proceso presenta dependencia de largo rango.

2.2 Un Modelo Espacial para datos continuos

Por mucho, la clase más usada de modelos de campos aleatorios de Markov para datos continuos es la clase de modelos auto-normales. Para datos normales la densidad condicional de $x_{i,j}$ dado $V_{i,j}$ (los vecinos de $x_{i,j}$), puede escribirse como

$$f(x_{i,j}|V_{i,j}) = (2\pi\tau_{i,j}^2)^{-1/2} \exp[-\{x_{i,j} - \theta(V_{i,j})\}^2 / 2\tau_{i,j}^2]$$

donde $\theta(V_{i,j})$ y $\tau_{i,j}^2$ son la media y varianza de $x_{i,j}$ condicionadas a sus vecinos, con $i = 1, \dots, n$, $j = 1, \dots, m$. Bajo la condición de regularidad de "dependencia pareada" entre sitios se tiene que

$$\theta(V_{i,j}) = \mu_{i,j} + \sum_{k,l \neq i,j} c_{i,j}(k,l)(x_{k,l} - \mu_{k,l}), \quad i = 1, \dots, n, \quad j = 1, \dots, m,$$

donde $c_{i,j}(k,l)\tau_{k,l}^2 = c_{k,l}(i,j)\tau_{i,j}^2$, $c_{i,j}(i,j) = 0$, y $c_{i,j}(n,m) = 0$ a menos que haya dependencia pareada entre los sitios (i,j) y (n,m) . De acuerdo a estos resultados, la distribución condicional está dada por

$$X_{i,j}|V_{i,j} \sim N \left(\mu_{i,j} + \sum_{k,l \neq i,j} c_{i,j}(k,l)(x_{k,l} - \mu_{k,l}), \tau_{i,j}^2 \right)$$

Como una consecuencia directa del teorema de factorización, se muestra en ANEXO A que

$$\mathbf{X} \sim \mathbf{N}(\boldsymbol{\mu}, (I - C)^{-1} M), \quad (2.5)$$

donde C es una matriz $nm \times nm$ cuyo (i,j) -ésimo elemento es $c_{i,j}$, $c_{i,j}\tau_j^2 = c_{j,i}\tau_i^2$, y $c_{i,i} = 0$; $M = \text{diag}(\tau_1^2, \dots, \tau_{nm}^2)$, y $\mu_i = E(X_{i,j})$.

En este capítulo vimos la forma condicional de modelar datos espaciales, así como los principales resultados en estadística espacial que se tendrán en cuenta para el estudio de nuestros dos modelos espaciales antes mencionados.

Capítulo 3

Estimación de Parámetros para Modelos en Látices

Por razones teóricas y computacionales, el método de máxima verosimilitud (MV) para estimación de parámetros de un modelo reticular ya no es automáticamente el método de elección. Cuando la independiencia estadística de observaciones no esté garantizada, la logverosimilitud negativa ya no puede ser escrita como una suma de variables aleatorias independientes. En este caso se necesita una forma más general del teorema del límite central (asumiendo dependiencia entre variables), para obtener propiedades asintóticas del estimador de MV. Más aún, no es obvio en estas situaciones más generales cómo obtener un máximo global en la superficie de verosimilitud.

Cuando los datos son indexados por el tiempo, hay un número de resultados disponibles que establecen consistencia, Gaussianidad asintótica, y eficiencia del estimador MV: Roussas (1972), Bhat (1974), Basawa y Prakasa Rao (1980, secciones 2.2, 2.3, 2.4), Hall y Heyde (1980), Heijmans y Magnus (1986a), Moore (1987), y las referencias en ellos, proveen una muestra de resultados disponibles para la estimación MV de parámetros de procesos aleatorios temporales.

En lo que va a seguir, se presentan varios criterios para estimación de pará-

metros de un modelo reticular. Máxima verosimilitud sigue siendo el más popular, aunque a menudo es difícil de probar Gaussianidad asintótica y eficiencia asintótica en el contexto espacial. Se realizarán las estimaciones de los parámetros por tres métodos propuestos en la literatura: Máxima verosimilitud vía MCMC (Gu y Zhu; 2001), Estimación vía ABC (Approximate Bayesian Computation) (Marjoram *et al.*, 2003) y estimación por máxima verosimilitud aproximada (Trabajo de C. Geyer en Barndorff-Nielsen; 1999, pp 79-140).

Supongamos que tenemos un patrón de puntos $\mathbf{X} = \{x_i \in A : i = 1, \dots, n\}$ en una región $A \in R^d$, donde R^d es un espacio Euclidiano d - dimensional. Por *modelo espacial* queremos decir un modelo estadístico para \mathbf{X} con densidad

$$f(\mathbf{x}|\theta) = \exp[-Q(\mathbf{x};\theta) - \log\{C(\theta)\}]$$

donde θ es un vector de parámetros p - dimensional, la función potencial $Q(\cdot; \cdot)$ exhibe la interacción entre componentes de X y la constante de normalización es

$$C(\theta) = \int_{A^n} \exp\{-Q(\mathbf{y};\theta)\} \mu(d\mathbf{y})$$

donde $\mu(d\mathbf{y})$ es la delta de Dirac $\delta_{\mathbf{y}}(d\mathbf{y})$ o $d\mathbf{y}$, en función de si \mathbf{y} toma valores discretos o continuos. Entonces la función de log-verosimilitud de θ para la observación $\mathbf{X} = \mathbf{x}_0$ es

$$l(\theta; \mathbf{x}_0) = \log\{f(\mathbf{x}_0|\theta)\} = -Q(\mathbf{x}_0;\theta) - \log\{C(\theta)\}. \quad (3.1)$$

Por simplicidad de notación, omitiremos \mathbf{x}_0 en $l(\theta; \mathbf{x}_0)$ y $Q(\mathbf{x}_0;\theta)$. Debido a la dificultad para calcular la constante de normalización, el *estimador de máxima verosimilitud* (EMV), es a menudo numéricamente imposible. Sin embargo, para los modelos específicos estudiados en esta tesis, si existen expresiones cerradas para el cálculo del estimador de máxima verosimilitud, en cada modelo se dará la estimación específica.

Besag (1975) propuso el *estimador de máxima pseudo-verosimilitud* (EMPV) el cual maximiza la *pseudo-verosimilitud*,

$$L_p(\theta; \mathbf{x}_0) = \prod_{i=1}^n f_\theta(x_i | \{x_j : j \neq i\})$$

como el producto directo de probabilidades condicionales (o densidades) de cada variable sobre el resto de las variables, o su logaritmo

$$l_p(\theta; \mathbf{x}_0) := \log L_p(\theta; \mathbf{x}_0) = \sum_{i=1}^n \log f_\theta(x_i | \{x_j : j \neq i\}).$$

El estimador de máxima pseudo-verosimilitud (EMPV) para modelos espaciales fué propuesto como una alternativa al EMV debido a la intratabilidad anterior. Besag (1986) y otros han sostenido que el EMPV puede ser más apropiado que el EMV para algunos propósitos prácticos, debido a la simplicidad inherente de la pseudo-verosimilitud. Ha sido matemáticamente demostrado que el EMPV es fuertemente consistente y distribuido asintóticamente normal alrededor del verdadero valor del parámetro para muestras grandes, aunque, por otro lado, también se ha mostrado que en muchas situaciones el EMPV no es eficiente en comparación con el EMV. Ver Jensen y Møller (1991); Comets (1992); Jensen y Künsch (1994); Goulad, Särkkä, y Grabarnik (1996); entre otros. El estudio realizado por Mase (1995) provee evidencia teórica y experimental con respecto a las propiedades del EMPV en varios casos. Geyer (1991) provee algunos resultados de simulación de un modelo Ising.

3.1 Máxima verosimilitud vía MCMC

Puesto que el objetivo es calcular el valor de θ tal que maximize (2.1), para una función suave $l(\theta)$, el método iterativo de Newton-Raphson para encontrar el valor máximo de la función, empezando en un θ_k , es

$$\theta_{k+1} = \theta_k - [\nabla^2 l(\theta)]^{-1} [\nabla l(\theta)]$$

donde l es la función de log-verosimilitud y

$$\begin{aligned}\nabla l(\theta) &= -\nabla Q(\theta) - \nabla [\log \{C(\theta)\}] \\ \nabla^2 l(\theta) &= -\nabla^2 Q(\theta) - \nabla^2 [\log \{C(\theta)\}]\end{aligned}$$

Note que las expresiones anteriores no son computables por la presencia de $C(\theta)$; sin embargo, usando las identidades $E_\theta [\nabla l(\theta; \mathbf{X})] = 0$ y $E_\theta [\nabla^2 l(\theta; \mathbf{X})] = -E_\theta [\nabla l(\theta; \mathbf{X})^{\otimes 2}]$, donde E_θ denota la esperanza con respecto a la densidad de \mathbf{X} , podemos mostrar que

$$\begin{aligned}\nabla [\log \{C(\theta)\}] &= -E_\theta [\nabla Q(\mathbf{X}; \boldsymbol{\theta})] \\ \nabla^2 [\log \{C(\theta)\}] &= -E_\theta [\nabla^2 Q(\mathbf{X}; \boldsymbol{\theta})] + E_\theta [\nabla Q(\mathbf{X}; \boldsymbol{\theta})]^{\otimes 2} - \{E_\theta [\nabla Q(\mathbf{X}; \boldsymbol{\theta})]\}^{\otimes 2},\end{aligned}$$

donde, para un vector \mathbf{a} , $\mathbf{a}^{\otimes 2} = \mathbf{a}\mathbf{a}^T$. En las expresiones anteriores los lados izquierdos (que se necesitan para el Newton-Raphson) dependen de $C(\theta)$, pero podemos aproximar los lados derechos usando simulación. Suponga que podemos simular realizaciones del látice \mathbf{X} , entonces las esperanzas anteriores por propiedades ergódicas, pueden ser aproximadas mediante los promedios correspondientes.

Primero introducimos dos pasos básicos en la aproximación estocástica. Tengamos en mente que es un procedimiento cíclico y que θ_k es el estimador actual de $\hat{\theta}$, \mathbf{h}_k es el estimador actual de $E_{\hat{\theta}} [\nabla Q(\mathbf{X}; \hat{\boldsymbol{\theta}})]$ y Γ_k es el estimador actual de

$$-E_{\hat{\theta}} [\nabla^2 Q(\mathbf{X}; \hat{\boldsymbol{\theta}})] + E_{\hat{\theta}} [\nabla Q(\mathbf{X}; \hat{\boldsymbol{\theta}})]^{\otimes 2}.$$

en cada ciclo.

También asumimos que, para cada θ , hay una densidad de probabilidad de transición de Markov $\Pi_\theta(\cdot, \cdot)$ tal que la cadena regida por esta probabilidad de transición es aperiódica e irreducible con distribución estacionaria $f(\mathbf{x}|\theta)$.

Paso 1: a la k –ésima iteración, ponemos $\mathbf{X}_{k,0} = \mathbf{X}_{k-1,m}$. Para $i = 1, \dots, m$, generamos $\mathbf{X}_{k,i}$ de la densidad de probabilidad de transición $\Pi_{\theta_{k-1}}(\mathbf{X}_{k,i-1}, \cdot)$.

Paso 2: actualizamos θ_{k-1} a θ_k , \mathbf{h}_{k-1} a \mathbf{h}_k y Γ_{k-1} a Γ_k por

$$\begin{aligned}\mathbf{h}_k &= \overline{H}(\theta_{k-1}; \mathbf{X}_k), \\ \Gamma_k &= \overline{I}(\theta_{k-1}; \mathbf{X}_k), \\ \theta_k &= \theta_{k-1} + \{\Gamma_{k-1} - \mathbf{h}_{k-1}^{\otimes 2}\}^{-1} \{-\nabla Q(\theta_{k-1}) + \overline{H}(\theta_{k-1}; \mathbf{X}_k)\},\end{aligned}$$

donde $\mathbf{X}_k = (\mathbf{X}_{k,1}, \dots, \mathbf{X}_{k,m})$,

$$\overline{H}(\theta_{k-1}; \mathbf{X}_k) = \frac{1}{m} \sum_{i=1}^m \nabla Q(\mathbf{X}_{k,i}; \theta_{k-1})$$

y

$$\overline{I}(\theta_{k-1}; \mathbf{X}_k) = -\frac{1}{m} \sum_{i=1}^m \nabla^2 Q(\mathbf{X}_{k,i}; \theta_{k-1}) + \frac{1}{m} \sum_{i=1}^m \{\nabla Q(\mathbf{X}_{k,i}; \theta_{k-1})\}^{\otimes 2}.$$

Este método, con algunas modificaciones, fué propuesto por Gu y Zhu en 2001.

3.2 Estimación vía ABC

Uno de los problemas básicos de la estadística bayesiana es el cálculo de las distribuciones posteriores. En este caso tenemos datos \mathbf{X} generados de un modelo M determinados por los parámetros θ y la densidad apriori denotada por $\pi(\theta)$. La distribución posterior de interés es $f(\theta|\mathbf{X})$, la cual está dada por

$$f(\theta|\mathbf{X}) = L(\mathbf{X}|\theta) \pi(\theta) / C(\theta)$$

donde $C(\theta) = \int L(\mathbf{X}|\theta) \pi(\theta) d\theta$ es la constante normalizante.

En particular tenemos el caso en el cual la verosimilitud $L(\mathbf{X}|\theta)$ no puede ser calculada explícitamente, en este caso se utiliza un algoritmo de MCMC basado en los siguientes pasos:

1. Estar en θ y proponer cambiar a θ' de acuerdo a un kernel de transición $q(\theta \rightarrow \theta')$.
2. Generar X' usando el modelo M con parámetros θ' .
3. Si $\rho(S, S') \leq \varepsilon$, seguir al paso 4, de otro modo permanecer con θ y regresar a 1, donde $S = (S_1, \dots, S_p)$ es un conjunto de estadísticas suficientes para θ y ρ una métrica apropiada.
4. Calcular

$$h = h(\theta, \theta') = \min \left(1, \frac{\pi(\theta') q(\theta' \rightarrow \theta)}{\pi(\theta) q(\theta \rightarrow \theta')} \right)$$

5. Aceptar θ' con probabilidad h , de otra forma quedarse con θ , después regresar al paso 1.

Para nuestro caso, usaremos el siguiente procedimiento consecuencia del anterior, cuando consideramos a $\pi(\theta) = 1$. Esto es,

1. Consideramos nuestra matriz de datos de observaciones \mathbf{x} , de la cual tenemos un estadístico suficiente $S_0 = S(\mathbf{x})$.
2. Proponemos varios parámetros $\theta^{(1)}, \theta^{(2)}, \dots, \theta^{(n)}$
3. De los estimadores antes dados, calculamos $S_1 = S(\mathbf{x}^{(1)})$, $S_2 = S(\mathbf{x}^{(2)})$, \dots , $S_n = S(\mathbf{x}^{(n)})$.
4. Y por último, aceptamos los parámetros que cumplan con la condición $d(S_0, S_i) < \varepsilon$, para un $\varepsilon > 0$ dado.

Tomamos el estimador del parámetro como la mediana de los parámetros aceptados. Este enfoque está basado en el trabajo de Marjoram et al. (2003).

3.3 Máxima Verosimilitud Aproximada

Como hemos comentado uno de los problemas que se presentan en la estimación de modelos en estadística espacial es que la constante de normalización es difícil de calcular computacionalmente; para evitar el cálculo de esta constante en el trabajo de C. Geyer (Barndorff-Nielsen et al.(1999)), se presenta un método para aproximar la verosimilitud. Consideremos la log-verosimilitud para una observación dada por (2.1). Resulta más conveniente usar la razón de log-verosimilitud contra un punto fijo $\psi \in \Theta$

$$l(\theta) = \log \frac{b_\theta(\mathbf{x}_0)}{b_\psi(\mathbf{x}_0)} - \log \frac{C(\theta)}{C(\psi)}$$

donde $b_\theta(\mathbf{x}) = \exp[-Q(\mathbf{x};\theta)]$. El primer término involucra las densidades no-normalizadas y se conoce en forma cerrada, el segundo involucra la función normalizante, la cual no se conoce. Pero, si $b_\theta(x) = 0$ cuando $b_\psi(x) = 0$, entonces

$$\begin{aligned} \frac{C(\theta)}{C(\psi)} &= \frac{1}{C(\psi)} \int b_\theta(x) \mu(dx) \\ &= \int \frac{b_\theta(x) b_\psi(x)}{b_\psi(x) C(\psi)} \mu(dx) \\ &= \int \frac{b_\theta(x)}{b_\psi(x)} f_\psi(x) \mu(dx) \\ &= E_\psi \frac{b_\theta(X)}{b_\psi(X)}. \end{aligned}$$

Esto permite el cálculo de la log-verosimilitud por MCMC. Sean $\mathbf{X}_1, \mathbf{X}_2, \dots$ simulaciones de P_ψ . Entonces

$$l(\theta) = \log \frac{b_\theta(\mathbf{x}_0)}{b_\psi(\mathbf{x}_0)} - \log \left(E_\psi \frac{b_\theta(X)}{b_\psi(X)} \right) \quad (3.2)$$

es aproximado por

$$l_n(\theta) = \log \frac{b_\theta(\mathbf{x}_0)}{b_\psi(\mathbf{x}_0)} - \log \left(\frac{1}{n} \sum_{i=1}^n \frac{b_\theta(X_i)}{b_\psi(X_i)} \right)$$

Maximizando esta última expresión tenemos una aproximación de Monte Carlo $\hat{\theta}_n$ al EMV $\hat{\theta}$, el cual maximiza (2.2).

Los tres métodos antes mencionados para estimación de parámetros en modelos espaciales son los que se estudiarán en este trabajo. En particular, el método de máxima verosimilitud para los dos modelos que se trabajaran es implementable en forma cerrada.

3.4 Un Método Bayesiano para Estimación Espacial

Recordemos que nuestro trabajo está enfocado en estimaciones donde se considera la función de verosimilitud, aún cuando hay métodos que evitan el uso de esta función. Sin embargo, hay métodos bayesianos que estudian las estimaciones espaciales y sus complicaciones como el que mencionaremos a continuación.

Comenzaremos recordando el algoritmo de Metropolis-Hastings para generar observaciones de una variable aleatoria X utilizando las salidas de una cadena de Markov.

1. Generar $y_t \sim q(y|x^t)$ donde $q(\cdot)$ es una distribución instrumental.
2. Tomar

$$x^{t+1} = \begin{cases} y_t & \text{con probabilidad } \rho(x^t, y_t) \\ x^t & \text{con probabilidad } 1 - \rho(x^t, y_t) \end{cases}$$

donde

$$\rho(x, y) = \min \left\{ \frac{f(y) q(x|y)}{f(x) q(y|x)}, 1 \right\}$$

3. Repetir los pasos 1 y 2, n veces.

En nuestro caso consideraremos la siguiente distribución condicional,

$$\pi(y|\theta) = \frac{q_\theta(y)}{Z_\theta}$$

y una distribución a priori para θ , $\pi(\theta)$, donde, $q_\theta(\cdot)$ es una función que involucra la función potencial mencionada anteriormente y Z_θ es la constante de normalización.

La distribución posterior está dada por

$$\pi(\theta|y) \propto \pi(\theta) \pi(y|\theta)$$

Si queremos muestrear de la posterior usando Metropolis-Hastings entonces tenemos que calcular

$$\begin{aligned} r &= \frac{\pi(\theta'|y) q(\theta|\theta')}{\pi(\theta|y) q(\theta'|\theta)} = \frac{\pi(\theta') \pi(y|\theta') q(\theta|\theta')/f(y)}{\pi(\theta) \pi(y|\theta) q(\theta'|\theta)/f(y)} \\ &= \frac{\pi(\theta') q_{\theta'}(y) q(\theta|\theta')/Z_{\theta'}}{\pi(\theta) q_\theta(y) q(\theta'|\theta)/Z_\theta} \\ &= \frac{\pi(\theta') q_{\theta'}(y) q(\theta|\theta')}{\pi(\theta) q_\theta(y) q(\theta'|\theta)} \cdot \frac{Z_\theta}{Z_{\theta'}} = r(\theta'|\theta) \end{aligned}$$

Un esquema estándar para simular de la posterior presenta los mismos problemas que máxima verosimilitud, esto es, las constantes de normalización no pueden evitarse en forma fácil. Sin embargo, Møller et al. (2006) propusieron una forma en la que se elige una distribución instrumental y una variable auxiliar adecuadamente, con las cuales nuestro cociente para el Metropolis-Hastings, queda como (ANEXO C)

$$r = \frac{f(x'|\theta', y) \pi(\theta') q_{\theta'}(y) q(\theta|\theta', x') q_\theta(x)}{f(x|\theta, y) \pi(\theta) q_\theta(y) q(\theta'|\theta, x) q_{\theta'}(x)},$$

es importante notar que esta expresión ya no depende de las constantes de normalización Z_θ ni $Z_{\theta'}$.

Entonces, Metropolis-Hastings acepta (θ', x') con probabilidad

$$\min\{r, 1\},$$

y si se rechaza (θ', x') entonces nos quedamos en (θ, x) y el algoritmo sigue.

Capítulo 4

Modelo Ising

Pickard (1987) dió una descripción detallada del modelo de Ising. En nuestro estudio de simulación, el modelo de Ising está definido en un conjunto de retículas cuadradas $M \times M$ en el plano: $\{X_{i,j} : i, j = 1, \dots, M\}$ con $M = 64$. Asumimos fronteras periódicas para la retícula cuadrada, equivalente a un toro; esto es, $X_{0,j} = X_{M,j}$, $X_{M+1,j} = X_{1,j}$, $X_{i,0} = X_{i,M}$, y $X_{i,M+1} = X_{i,1}$ para toda i y j . Las condiciones anteriores se asumieron para todos los métodos de estimación, con excepción para el de MV, ya que en éste se considerarán condiciones de frontera de forma cilíndrica en lugar de la del toro. Para simular el proceso, usamos el muestreador de Gibbs.

4.1 Máxima verosimilitud

En este caso estudiaremos el método propuesto por Pettitt & Friel (2003), para calcular la constante de normalización para el modelo, en el cual consideraremos el modelo en forma cilíndrica, matrices de dimensión $n \times n$, y la siguiente distribución autologística sin normalizar,

$$q(\mathbf{x}|\beta) = \exp\{\beta[V_f(\mathbf{x}) + V_c(\mathbf{x})]\},$$

con

$$V_f(\mathbf{x}) = \sum_{k=1}^{n-1} \sum_{l=1}^n x_{kl} x_{k+1,l} + \sum_{k=1}^n \sum_{l=1}^{n-1} x_{kl} x_{k,l+1}$$

$$V_c(\mathbf{x}) = \sum_{k=1}^n x_{k1} x_{k,n}.$$

El primer término de $V_f(\mathbf{x})$ es la suma del producto de los vecinos cercanos dentro de las columnas, mientras que el segundo término es la suma análoga dentro de los renglones y $V_c(\mathbf{x})$ denota la suma de los productos de vecinos entre la primer y la última columna (arreglo cilíndrico).

Sea $\mathbf{x} = \{x_{ij}; i = 1, \dots, m; j = 1, \dots, n\}$ un arreglo K -valuado con x_{ij} tomando K valores discretos denotados por $B = \{b_1, b_2, \dots, b_K\}$. Definimos un m -vector por $\mathbf{x}_j = x_{1j}, \dots, x_{mj}$, $j = 1, \dots, n$. Sea $A = \{a_1, \dots, a_N\}$ el conjunto de todos los valores posibles de x_j con $N = K^m$. Definimos x_0 igual a x_n , imponiendo la condición de cilindro. Pettitt & Friel (2003), proponen el siguiente teorema para calcular la constante de normalización.

Teorema *Supongamos que la distribución sin normalizar $q(\mathbf{x}|\beta)$ satisface una factorización*

$$q(\mathbf{x}|\beta) = \prod_{j=1}^n h(\mathbf{x}_j, \mathbf{x}_{j-1}),$$

para una función real positiva dada $h(\cdot, \cdot)$ definida en el conjunto $A \times A$. Entonces la constante de normalización para $q(\mathbf{x}|\beta)$ está dada por $\text{tr}(Q^n)$ donde Q es una matriz $N \times N$ con su k -ésimo renglón (Q_{k1}, \dots, Q_{kN}) definido por

$$\{h(x_1 = a_1; x_0 = a_k); h(x_1 = a_2; x_0 = a_k), \dots, h(x_1 = a_N; x_0 = a_k)\}$$

para $k = 1, \dots, N$.

Para estudiar este resultado, simulamos una matriz 10×10 con un *burning* de 5000 y con $\beta = 0.3$. La elección del valor de este parámetro se discutirá al final de la sección. En la Figura 1 se muestran tres gráficas, todas muestran los

resultados de cada iteración del burning; la primera muestra las medias de cada matriz simulada en cada ciclo; la segunda muestra las correlaciones que hay en los datos de cada matriz y la tercera (último renglón) nos muestra las medias acumuladas en cada iteración del procedimiento; como se puede observar, las dos primeras gráficas no nos dicen mucho de la estabilidad de las matrices simuladas, pero la última nos muestra que desde la iteración número 1000 se ha llegado a una matriz simulada estable.

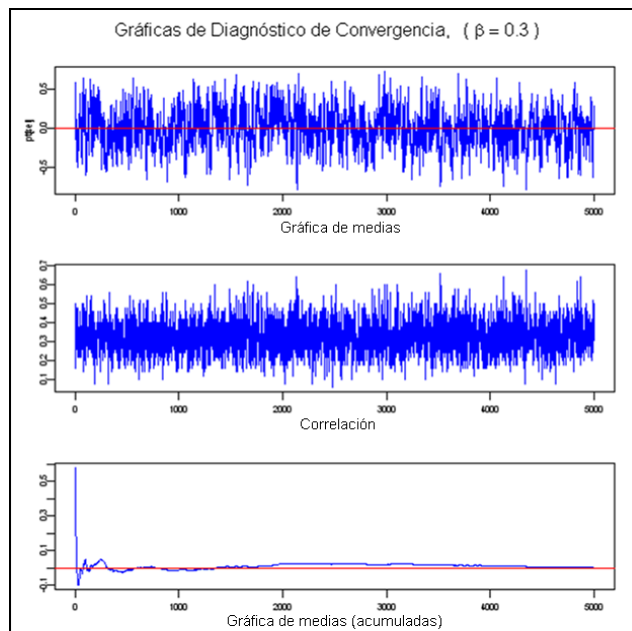


Figura 1

Una vez que hemos decidido que el muestreador de Gibbs ha llegado a una región estable nos quedamos con una realización de la matriz de datos. Para estos datos estimamos el valor de β usando MV, aprovechando que en este caso tenemos la constante de normalización, es importante notar que el método de Pettitt y Friel es implementable solamente para látices de dimensión pequeña.

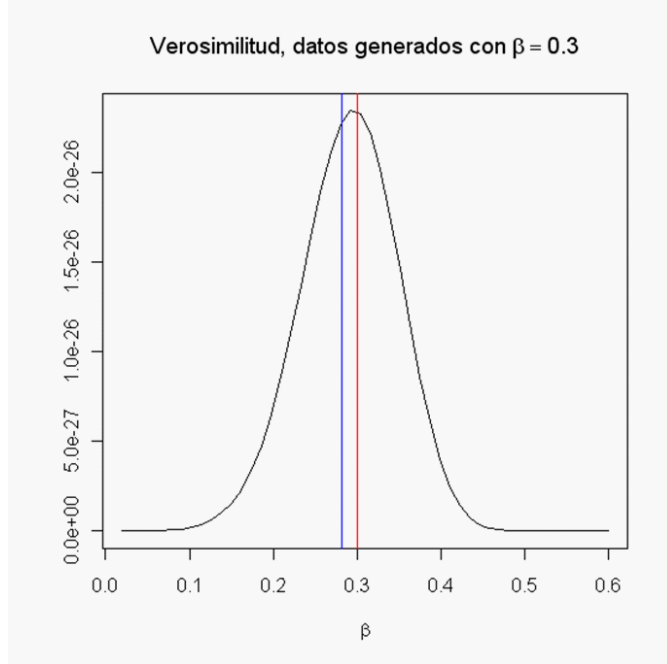


Figura 2

En la Figura 2 la línea azul corresponde a la estimación vía Pseudo-verosimilitud ($\tilde{\beta} = 0.2821$) y la roja a la estimación por Máxima Verosimilitud ($\hat{\beta} = 0.2961$).

4.2 Máxima verosimilitud vía MCMC

De ahora en adelante consideraremos el modelo:

$$g \equiv \gamma [(x_{i-1,j} + x_{i+1,j}) + (x_{i,j-1} + x_{i,j+1})], \quad (4.1)$$

para la expresión (1.4). El estadístico que cuenta los puntos vecinos en la retícula se define como

$$V = V(\mathbf{X}) = \sum_{nn} x_{i,j} x_{u,v}$$

donde nn significa que la sumatoria es sobre todos los pares $x_{i,j}$ y $x_{u,v}$ tales que los dos puntos son vecinos mas cercanos. La diferencia de este modelo

al modelo utilizado para la estimación máxima verosimilitud, es que en este caso tomaremos los datos en forma de toro y la dimensión de nuestra matriz de datos puede ser mayor a la anterior. La función potencial es $Q(\mathbf{X}; \boldsymbol{\theta}) = -\boldsymbol{\theta}V(\mathbf{X})$ y el factor normalizante es obtenido sumando sobre todas las posibles configuraciones en \mathbf{X} ,

$$C(\boldsymbol{\theta}) = \sum_{\mathbf{X}} \exp\{\boldsymbol{\theta}V(\mathbf{X})\}$$

Para comparar los algoritmos propuestos en el capítulo 2 para la estimación espacial, simulamos un conjunto de datos de 64×64 donde $x_{ij} = -1, 1$ con $i, j = 1, \dots, 64$ y con un sólo parámetro de asociación entre vecinos $\gamma = 0.3$ entonces nuestro modelo se verá como (3.1). Los datos fueron simulados vía *Gibbs sampling*. Cada sitio (i, j) fué seleccionado en orden lexicográfico. El estado inicial del proceso fué tomado aleatoriamente tal que $x_{i,j}$ es independiente $\forall i, j$ con valores ± 1 con probabilidad igual. El algoritmo de Gibbs fue repetido al menos 300×64^2 veces (300 pasos Monte Carlo). Entonces,

$$M_T(\theta) = \sum_{n=1}^T \left(\frac{\sum_i \sum_j x_{i,j}}{4096} \right) / T$$

fué utilizada (promedio acumulado) para asegurar la convergencia Metropolis, donde $T \geq 300$. La Figura 3 muestra las gráficas de las medias, correlaciones y medias acumuladas de las matrices en cada iteración,

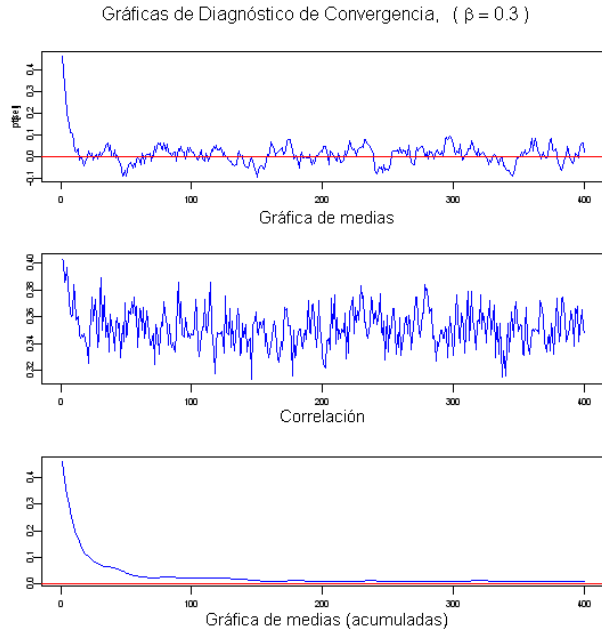


Figura 3

Para obtener el estimador de máxima verosimilitud vía MCMC y proceder con el método de Newton Raphson, se dió como punto inicial del parámetro el estimador de máxima pseudo-verosimilitud; para esta estimación se consideran la probabilidades condicionales, esto es

$$Z_{i,j} | V_{i,j} \sim Ber(p)$$

donde,

$$p = \frac{\exp(g)}{1 + \exp(g)},$$

con

$$g \equiv \gamma [(x_{i-1,j} + x_{i+1,j}) + (x_{i,j-1} + x_{i,j+1})].$$

Luego la función de pseudo-verosimilitud a maximizar es

$$L(\gamma) = \prod_{i=1}^{64} \prod_{j=1}^{64} \frac{[\exp(\gamma [(x_{i-1,j} + x_{i+1,j}) + (x_{i,j-1} + x_{i,j+1})])]^{x_{i,j}}}{1 + \exp(\gamma [(x_{i-1,j} + x_{i+1,j}) + (x_{i,j-1} + x_{i,j+1})])}$$

o maximizar

$$l(\gamma) = \sum_{i=1}^{64} \sum_{j=1}^{64} (x_{i,j}) (\gamma [(x_{i-1,j} + x_{i+1,j}) + (x_{i,j-1} + x_{i,j+1})]) - \sum_{i=1}^{64} \sum_{j=1}^{64} \log [1 + \exp(\gamma [(x_{i-1,j} + x_{i+1,j}) + (x_{i,j-1} + x_{i,j+1})])]$$

y el estimador de máxima pseudo-verosimilitud fué $\hat{\gamma} = 0.2836$. En la Figura 4 se muestra la función de pseudoverosimilitud y el estimador.

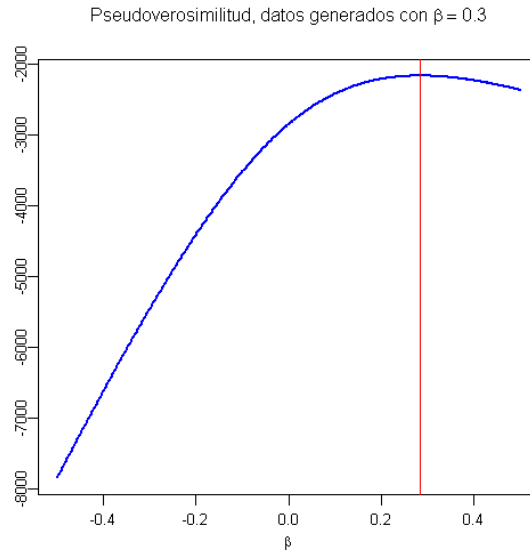


Figura 4

En el trabajo realizado por Gu y Zhu (2001), estudian el método por MCMC junto con el algoritmo de Newton-Raphson para el estimador deseado, además comparan su método con el estimador de MPV. En este mismo trabajo, se tiene como resultado que su método es apropiado cuando se tiene un buen

punto de inicio, ya que sin éste el algoritmo es muy lento; en nuestro caso los estimadores de θ para cada iteración estarán dados por la expresión,

$$\theta_k = \theta_{k-1} - \frac{\{\frac{1}{m} \sum_{i=1}^m \nabla Q(\mathbf{X}_{k,i}; \theta_{k-1}) - \nabla Q(\theta_{k-1})\}}{\frac{1}{m} \sum_{i=1}^m \{\nabla Q(\mathbf{X}_{k,i}; \theta_{k-1})\}^2 - [\frac{1}{m} \sum_{i=1}^m \nabla Q(\mathbf{X}_{k,i}; \theta_{k-1})]^2},$$

donde

$$\begin{aligned} \nabla Q(\theta_{k-1}) &= V(\mathbf{X}) = \sum_{nn} x_{i,j} x_{u,v} \\ \nabla Q(\mathbf{X}_{k,i}; \theta_{k-1}) &= \sum_{nn} x'_{i,j} x'_{u,v} \end{aligned}$$

y $\mathbf{X}_{k,i}$ es una matriz de datos simulados con el parámetro θ_{k-1} para $i = 1, 2, \dots, m$, tomando a $m = 1100$ y como punto inicial el estimador de MPV; después de 200 iteraciones y 5 horas con 45 minutos, llegamos al estimador $\hat{\gamma}_{200} = 0.2882$. En la siguiente figura mostramos las medias acumuladas de los 200 estimadores en cada iteración; como se puede ver el estimador por este método se acerca al valor real del parámetro (Figura 5).

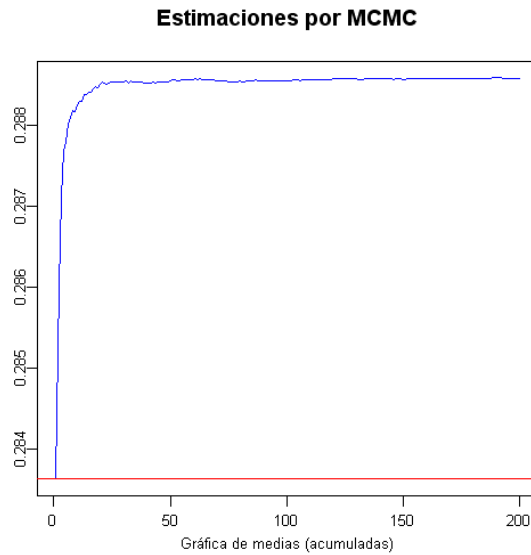


Figura 5

4.3 Estimación vía ABC

En el trabajo de Marjoram (2003) se estudian distintos algoritmos basados en MCMC para la estimación de parámetros, sin embargo, no se realizan comparaciones donde intervengan aproximaciones a la verosimilitud.

En el trabajo de Pickard (1987) se trabaja la interacción entre vecinos cercanos para retículas y puede ser descrita por el parámetro

$$S = N^{-1} \sum_{nn} corr(x_i, x_j)$$

donde N es el número de pares de vecinos cercanos en la retícula.

De acuerdo al procedimiento descrito en la sección 2.2, tomamos al operador $\rho(S, S')$ como la distancia entre las dos correlaciones S y S' ; donde S' es la estadística suficiente para la matriz de datos simulada X' usando el parámetro propuesto θ' , y tomamos como la distancia mínima aceptable $\varepsilon = 0.1$.

De igual manera tomamos como punto fijo el EMPV $\hat{\gamma} = 0.2836$. En este caso sólo se aceptaron 450 de 1000 parámetros en el algoritmo, con una duración de 2 horas con 22 minutos, dando como resultado $\hat{\gamma}_{450} = 0.2567$ y el estimador por medias acumuladas es $\hat{\gamma}_{med} = 0.2778$. En la siguiente figura mostramos las medias acumuladas de los 450 estimadores en cada iteración.

Estimaciones por MCMC sin verosimilitudes

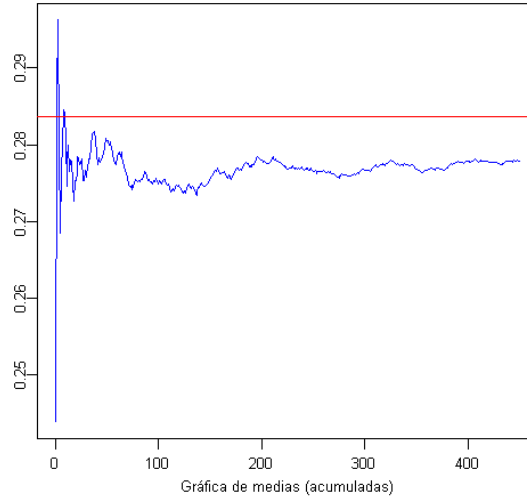


Figura 6

El método ABC tal y como lo hemos implementado intenta aproximar la media de la distribución posterior partiendo de una a priori uniforme.

4.4 Máxima Verosimilitud Aproximada

El método de C. Geyer (1999) aproxima a la log-verosimilitud mediante

$$l_n(\theta) = \log \frac{b_\theta(\mathbf{x}_0)}{b_\psi(\mathbf{x}_0)} - \log \left(\frac{1}{n} \sum_{i=1}^n \frac{b_\theta(X_i)}{b_\psi(X_i)} \right)$$

donde $b_\theta(\mathbf{x}) = \exp[-Q(\mathbf{x}; \theta)]$. Para el modelo Ising tenemos que $Q(\mathbf{x}; \theta) = -\theta V(X)$, y en este caso la función a maximizar está dada por la siguiente expresión

$$l_n(\theta) = \theta V(X) - \log \left\{ \frac{1}{n} \exp[(\theta - \theta_i) V(X_i)] \right\}$$

donde $V(X)$ es la suma de vecinos de la retícula, θ_i es un punto fijo y $V(X_i)$ es la suma de vecinos de la retícula simulada con θ_i . En este caso, realizamos esta optimización en forma recursiva; tomando $n = 1500$ y el punto fijo de

inicio el EMPV. Con 200 iteraciones duró 9 horas con 10 minutos y tenemos el siguiente resultado

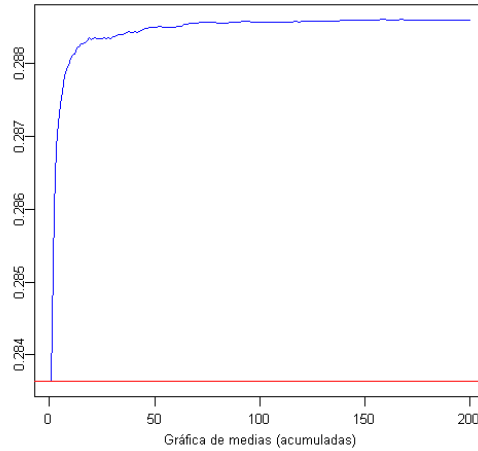


Figura 7

el método de búsqueda empleado, arroja un máximo en $\hat{\gamma}_{200} = 0.2883$ y el estimador por medias acumuladas es $\hat{\gamma}_{med} = 0.2886$.

Para este modelo Ising en particular, hay transiciones de fase cerca de $\gamma = \pm.44$ (Pickard 1987), entonces el valor del parámetro $\gamma = 0.3$ representa una interacción fuerte. En el trabajo realizado por Huang y Ogata (1999) estudian la variabilidad de los estimadores de máxima verosimilitud, máxima pseudoverosimilitud y máxima verosimilitud aproximada; ésta última es simplemente la que realizamos en el presente trabajo la cual nombramos como máxima verosimilitud vía MCMC. En ese mismo trabajo (Huang, Ogata;1999), trabajan con matrices de datos simulados con valores del parámetro $\gamma = .1, .2, .3, \text{ y } .4$ de dimensión 64×64 , y se observa que para los valores $.3$ y $.4$ del parámetro, los estimadores de máxima verosimilitud aproximada tienen menos variabilidad que los de máxima pseudoverosimilitud. Esta es una de las razones por las cuales trabajamos con $\gamma = 0.3$, en nuestra tesis.

Capítulo 5

Modelo Normal

5.1 Máxima verosimilitud

En lo que sigue se supondrá la homoscedasticidad de los datos, esto es considerando $M = \tau^2 I$ en

$$\mathbf{X} \sim \mathbf{N}(\boldsymbol{\mu}, (I - C)^{-1} M),$$

además consideramos $\boldsymbol{\mu} = X\boldsymbol{\beta}$, el cual modela la variación de los datos a gran escala con un número pequeño de parámetros que aparecen como combinaciones lineales de variables X independientes (Estas variables pueden ser funciones de la localización espacial).

De la expresión anterior y las suposiciones anteriores, la log-verosimilitud negativa resulta

$$\begin{aligned} L(\boldsymbol{\beta}, \tau^2, C) &= (n/2) \log(2\pi\tau^2) - (1/2) \log(|I - C|) \\ &\quad + (1/2) (\mathbf{x} - X\boldsymbol{\beta})' (I - C) (\mathbf{x} - X\boldsymbol{\beta}) / \tau^2 \end{aligned} \tag{5.1}$$

donde $I - C$ es una matriz simétrica y positiva definida. Esto puede ser minimizado en etapas: Para $I - C$ fijo,

$$\begin{aligned}\hat{\boldsymbol{\beta}} &= (X'(I - C)X)^{-1}X'(I - C)\mathbf{x}, \\ \hat{\tau}^2 &= (\mathbf{x} - X\hat{\boldsymbol{\beta}})'(I - C)(\mathbf{x} - X\hat{\boldsymbol{\beta}})/n\end{aligned}\tag{4.2}$$

son los estimadores MV de $\boldsymbol{\beta}$ y τ^2 . Sustituyendo en (4.1), el EMV de C puede ser obtenido minimizando la log-verosimilitud perfil negativa,

$$\begin{aligned}L^*(C) &= (n/2)\log(2\pi + 1) - (1/2)\log(|I - C|) \\ &\quad + (n/2)\log[\mathbf{x}'(I - C)\{I - X(X'(I - C)X)^{-1}X'(I - C)\}\mathbf{x}],\end{aligned}\tag{4.3}$$

con respecto a las incógnitas c_{ij} s.

Como estamos considerando el modelo auto-normal de media cero y la homoscedasticidad de los datos entonces,

$$\begin{aligned}\theta(V_{i,j}) &= \sum_{k,l \neq i,j} c_{i,j}(k,l)x_{k,l}, \\ \text{var}(V_{i,j}) &= \tau^2,\end{aligned}$$

donde cada renglón de la matriz C tiene la mayor parte de las entradas ceros excepto por las entradas $c_{i,j}$, que corresponden a los vecinos. Luego de (4.2) tenemos que,

$$\begin{aligned}\hat{\mu} &= \sum_{i=1}^n \sum_{j=1}^m x_{i,j}/nm = \bar{x}, \\ \hat{\tau}^2 &= (\mathbf{x} - \bar{x}\mathbf{1})'(I - C)(\mathbf{x} - \bar{x}\mathbf{1})/nm,\end{aligned}\tag{4.4}$$

donde $\mathbf{1} \equiv (1, \dots, 1)'$. Se puede mostrar (Moran, 1973a; Besag and Moran, 1975) que C tiene eigenvalores

$$\left\{ \sum_i \sum_j c_{i,j} \cos(\varpi_p i + \eta_q j) : p = 1, \dots, n; q = 1, \dots, m \right\},$$

donde $\varpi_p \equiv 2\pi p/n$, $p = 1, \dots, n$, y $\eta_q \equiv 2\pi q/m$, $q = 1, \dots, m$. Entonces,

$$|I - C| = \prod_{p=1}^n \prod_{q=1}^m \left\{ 1 - \sum_i \sum_j c_{i,j} \cos(\varpi_p i + \eta_q j) \right\}.$$

Definimos

$$\begin{aligned} \bar{C}(h_1, h_2) &\equiv \sum_{i=1}^{n-h_1} \sum_{j=1}^{m-h_2} (x_{i,j} - \bar{x})(x_{i+h_1, j+h_2} - \bar{x}) / nm, \\ h_1 &= 0, 1, \dots, n-1, \quad h_2 = 0, 1, \dots, m-1, \end{aligned}$$

el cual es el covariograma empírico. Entonces, de (4.4)

$$\hat{\tau}^2 = \bar{C}(0, 0) - \sum_i \sum_j c_{i,j} \bar{C}(i, j),$$

porque $\bar{C}(-h_1, -h_2) = \bar{C}(h_1, h_2)$. Sustituyendo esto en (4.1) tenemos la log-verosimilitud perfil negativa

$$\begin{aligned} &(nm/2) (\log(2\pi) + 1) + (nm/2) \log \left\{ \bar{C}(0, 0) - \sum_i \sum_j c_{i,j} \bar{C}(i, j) \right\} \\ &- (1/2) \sum_{i=1}^n \sum_{j=1}^m \log \left\{ 1 - \sum_i \sum_j c_{i,j} \cos(\varpi_p i + \eta_q j) \right\}, \end{aligned}$$

la cual debe ser minimizada con respecto a las incógnitas $c_{i,j}$ s.

5.2 Máxima verosimilitud vía MCMC

En el estudio de simulación, consideramos el siguiente modelo autonormal (Besag 1974)

$$C^{-1}(\beta, \sigma) \times \exp \left\{ -1/2 \tau^{-2} \sum_{i=1}^M \sum_{j=1}^N X_{i,j} (X_{i,j} - \beta V_{i,j}) \right\}$$

en una reticula $M \times N$ con $M = N = 64$, donde $V_{i,j}$ es la sumatoria de los vecinos de primer orden de $X_{i,j}$. En nuestra simulación, $\beta = 0.1$, mientras que el valor

de τ^2 se fija en 2 sin pérdida de generalidad. En el artículo de Huang y Ogata (1999) se observó que la estacionariedad del modelo autonormal requiere la restricción $|\beta| < 1/4$, valores mayores representan dependencia relativamente fuerte sobre los vecinos cercanos.

La función potencial en este caso es

$$U(X; \beta, \tau) = -1/2 \tau^{-2} \sum_{i=1}^M \sum_{j=1}^N X_{i,j} (X_{i,j} - \beta V_{i,j})$$

y para estimar el gradiente y Hessiano de la log-verosimilitud, calculamos las derivadas de primer y segundo orden de la función potencial con respecto a β y τ . Nuevamente usamos el *Gibbs sampling* para generar el proceso. Esto es, cada sitio (i, j) fué seleccionado en orden lexicográfico para reemplazar $X_{i,j}$ por $\beta V_{i,j} + \epsilon_{i,j}$, donde $\epsilon_{i,j}$ es generado independientemente con distribución $N(0, \tau^2)$. Empezamos fijando $X_{i,j} = 0$ para todos los sitios (i, j) y después iteramos el proceso de Gibbs sampling 600 pasos de Monte Carlo para alcanzar el equilibrio. Para este modelo, el EMPV es simplemente el estimador de mínimos cuadrados que minimiza $\sum_{i,j} (x_{i,j} - \beta V_{i,j})^2$.

Las siguientes gráficas se obtuvieron durante la simulación,

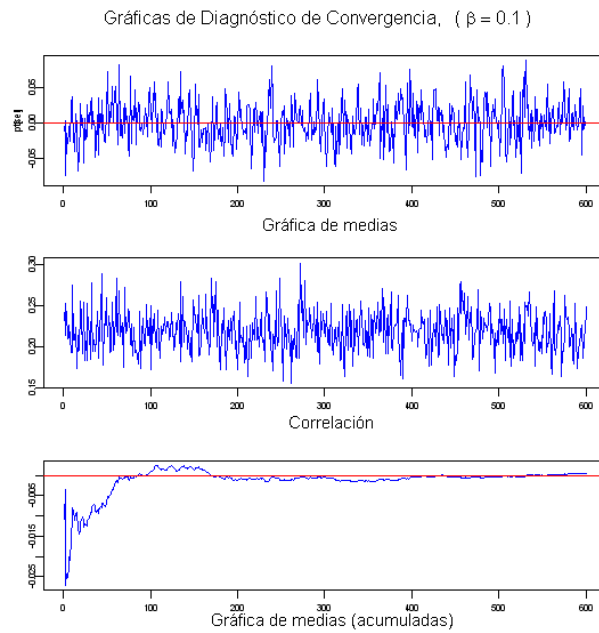


Figura 8

y el estimador de máxima verosimilitud en este caso es 0.1012, mientras que el estimador de máxima pseudo-verosimilitud fué $\hat{\beta} = 0.1056$.

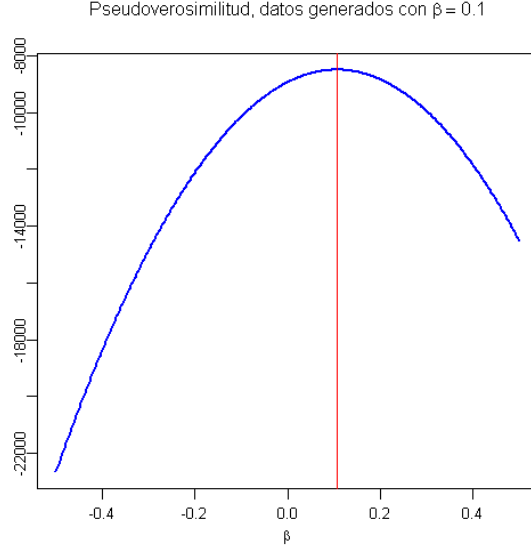


Figura 9

En nuestro caso los estimadores de θ para cada iteración estarán dados por la expresión,

$$\theta_k = \theta_{k-1} - \frac{\{\frac{1}{m} \sum_{i=1}^m \nabla Q(\mathbf{X}_{k,i}; \theta_{k-1}) - \nabla Q(\theta_{k-1})\}}{\frac{1}{m} \sum_{i=1}^m \{\nabla Q(\mathbf{X}_{k,i}; \theta_{k-1})\}^2 - [\frac{1}{m} \sum_{i=1}^m \nabla Q(\mathbf{X}_{k,i}; \theta_{k-1})]^2},$$

donde

$$\begin{aligned} \nabla Q(\theta_{k-1}) &= V(\mathbf{X}) = \sum_{nn} x_{i,j} x_{u,v} / 2\hat{\tau}^2 \\ \nabla Q(\mathbf{X}_{k,i}; \theta_{k-1}) &= \sum_{nn} x'_{i,j} x'_{u,v} / 2\hat{\tau}_{k-1}^2 \end{aligned}$$

y $\mathbf{X}_{k,i}$ es una matriz de datos simulados con los parámetros θ_{k-1} y $\hat{\tau}_{k-1}^2$ para $i = 1, 2, \dots, m$, tomando a $m = 1100$ y como punto inicial el estimador de MPV; teniendo el estimador MPV como punto inicial del método por MCMC, después de 200 iteraciones llegamos al estimador $\hat{\beta}_{200} = 0.1026$. La gráfica de las medias acumuladas de los estimadores en cada iteración se muestra a continuación.

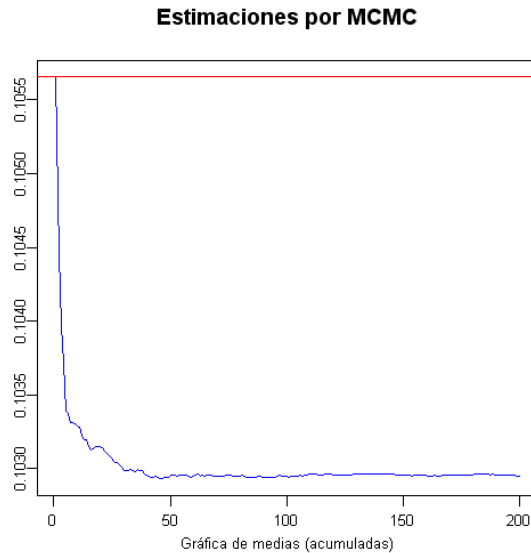


Figura 10

El tiempo de duración de la computadora para este método fué de 4 horas 37 minutos y 12 segundos.

5.3 Estimación vía ABC

Puesto que la estadística suficiente que representa la interacción entre vecinos cercanos mencionada en el trabajo de Pickard (1987) es general para cualquier retícula; estudiamos este método con las consideraciones de la sección 3.3. De igual manera tomamos como punto fijo el EMPV $\hat{\beta} = 0.1056$. El algoritmo ABC requiere de la simulación de las matrices aleatorias, X ; para ello, de igual forma que en la sección anterior usamos el muestreador de Gibbs con recorrido secuencial. En este caso sólo se aceptaron 673 parámetros de 1000 en el algoritmo, con duración de 1 hora con 40 minutos y dando como resultado $\hat{\beta}_{673} = 0.114$. Aunque tomando las medias acumuladas se estabilizan en 0.1001 aproximadamente, el comportamiento de las medias acumuladas de

los estimadores se muestra en la siguiente gráfica.

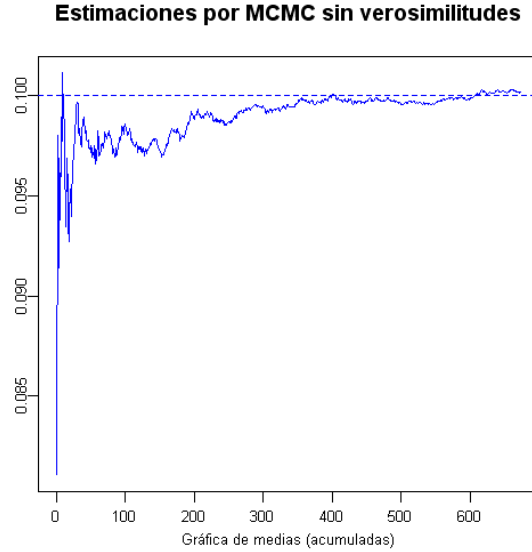


Figura 11

5.4 Máxima Verosimilitud Aproximada

De igual manera que en la sección 3.4 la función a maximizar está dada por la siguiente expresión

$$l_n(\beta) = \frac{\beta V(X)}{2\tau^2} - \log \left\{ \frac{1}{n} \exp \left[\left(\frac{\beta}{\tau^2} - \frac{\beta_i}{\hat{\tau}_{i-1}^2} \right) V(X_i) / 2 \right] \right\}$$

donde $V(X)$ es la suma de vecinos de la retícula, β_i es un punto fijo y $V(X_i)$ es la suma de vecinos de la retícula simulada con β_i y $\hat{\tau}_{i-1}^2$. Tomando el mismo punto fijo de inicio el EMPV. Con 200 iteraciones tenemos el siguiente resultado

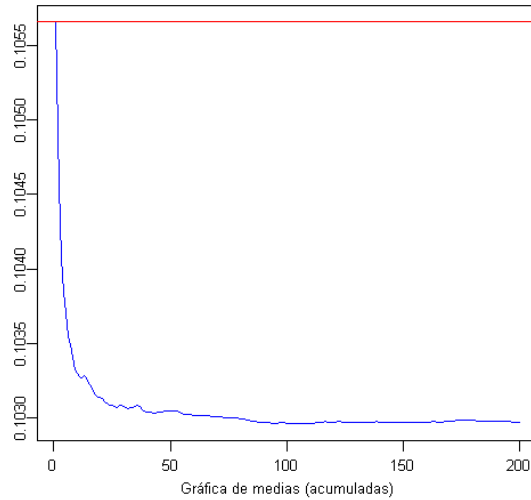


Figura 12

el estimador encontrado es $\hat{\beta} = 0.1029$. El tiempo que tomó este procedimiento fué de 6 horas con 12 minutos.

En el trabajo realizado por Huang y Ogata (1999) estudian la variabilidad de los estimadores de máxima verosimilitud, máxima pseudoverosimilitud y máxima verosimilitud aproximada; trabajan con matrices de datos simulados con valores del parámetro $\beta = .098, .147, .196$, y $.245$, $\sigma = 1$ de dimensión 64×64 , y se observa que para los valores cercanos a $.1$ del parámetro, los estimadores de máxima verosimilitud aproximada tienen una variabilidad similar a los de máxima pseudoverosimilitud.

Capítulo 6

Conclusiones

Nuestro objeto de estudio en este trabajo se enmarca dentro del área de modelación espacial, la cual comprende tres grandes subáreas: Geoestadística, patrones puntuales y datos en látrices, siendo esta última subárea la de nuestro interés. Estudiamos técnicas de estimación cuando se tiene la función de verosimilitud espacial. El tener la capacidad de estimar la verosimilitud espacial tiene el potencial de poder utilizar toda la teoría de verosimilitud, p. ej., cociente de verosimilitudes para la comparación de modelos, simulación y predicción, mientras que técnicas estándar en geoestadística, como Kriging, tienen roles muy específicos como en este caso predicción. Por otro lado, ese potencial creemos que todavía está lejos de ser alcanzable, pues las técnicas computacionales requieren ser más eficientes para que sean prácticas. Se mostró que realizar una estimación en un modelo espacial de un sólo parámetro y para matrices relativamente pequeñas es complicado hablando en términos de tiempo que se toma cada método computacional.

Particularizando a los modelos estudiados en este trabajo observamos que el método de Máxima Verosimilitud Aproximada para ambos modelos (Ising y Normal) es el que se tomaba mas tiempo en realizar los cálculos; pero, hay que notar que éste método sólo se utiliza en la literatura tomando un punto fijo para aproximar a la expresión (2.2), en nuestro trabajo se realizó

en forma recursiva para propósitos de exactitud de los estimadores. Esta mejor precisión fue consistente en ambos modelos (Ising y Normal), pues sus medias acumuladas se acercaban más al valor real. En este caso se sacrificó tiempo para la estimación por esta consistencia.

En general las estimaciones para el modelo de Ising tomaron más tiempo que las estimaciones del modelo Normal correspondientes; notando que el tamaño de la matriz afecta el tiempo de estimación, es por eso que en análisis de imágenes, donde se estudian matrices de pixeles mayores de 64×64 , raramente enfrentan problemas de estimación y sus problemáticas típicamente son diferentes, por ej. filtrado de ruido.

Uno de los puntos importantes que se deben notar en este trabajo, es que existen expresiones para calcular los estimadores de máxima verosimilitud para los modelos estudiados, pero que estas están sujetas a varias consideraciones que hacen que los modelos tengan poca flexibilidad a cambios de condiciones iniciales.

Esto da pie a considerar estudios de modelos discretos donde los datos tengan más de dos posibles resultados o un criterio de vecinos cercanos de segundo orden, es decir, modelos más generales que el modelo Ising estudiado en este trabajo. También sería necesario pensar en una simulación a gran escala para comparar los métodos de estimación; al hacer esta simulación, también sería necesario migrar a otro lenguaje de programación como C o C++, y paralelización de procesadores. Nuestro trabajo lo realizamos en R, porque es con el que estamos familiarizados y porque se aprovecharon comandos propios de R para el estudio y simulación de datos.

Part I

ANEXOS

6.1 ANEXO A

Modelo Gaussiano

Sin pérdida de generalidad definimos la siguiente densidad condicional para datos normales

$$f(x_i|x_j ; j \neq i) = (2\pi\tau_i^2)^{-1/2} \exp \left[- (x_i - \theta_i)^2 / 2\tau_i^2 \right]$$

donde θ_i y τ_i^2 son su media y varianza condicional respectivamente. Bajo la condición de regularidad de "dependencia pareada" entre sitios

$$\theta_i = \mu_i + \sum_{j=i} c_{i,j}(x_j - \mu_j),$$

donde $c_{i,j}\tau_j^2 = c_{ji}\tau_i^2$, $c_{ii} = 0$, y $c_{ik} = 0$ a menos que haya dependencia pareada entre los sitios i y k .

Bajo esta especificación, consideremos la forma de la distribución conjunta.

Para ello recordemos el teorema de factorización (Cressie p. 412)

Sean $z = (z_1, \dots, z_n)$ y $y = (y_1, \dots, y_n)$ entonces

$$\begin{aligned} f(z) &= f(z_n|z_1, \dots, z_{n-1}) f(z_1, \dots, z_{n-1}) \\ &= f(z_n|z_1, \dots, z_{n-1}) \frac{f(z_1, \dots, z_{n-1}, y_n)}{f(y_n|z_1, \dots, z_{n-1})} \\ &= \frac{f(z_n|z_1, \dots, z_{n-1})}{f(y_n|z_1, \dots, z_{n-1})} f(z_1, \dots, z_{n-1}, y_n) \\ &= \underbrace{\frac{f(z_n|z_1, \dots, z_{n-1})}{f(y_n|z_1, \dots, z_{n-1})}}_{\square_1} f(z_{n-1}|z_1, \dots, z_{n-2}, y_n) f(z_1, \dots, z_{n-2}, y_n) \end{aligned}$$

$$\begin{aligned}
f(z) &= \square_1 f(z_{n-1}|z_1, \dots, z_{n-2}, y_n) \frac{f(z_1, \dots, z_{n-2}, y_{n-1}, y_n)}{f(y_{n-1}|z_1, \dots, z_{n-2}, y_n)} \\
&= \square_1 \underbrace{\frac{f(z_{n-1}|z_1, \dots, z_{n-2}, y_n)}{f(y_{n-1}|z_1, \dots, z_{n-2}, y_n)}}_{\square_2} f(z_1, \dots, z_{n-2}, y_{n-1}, y_n) \\
&= \square_1 \square_2 f(z_1, \dots, z_{n-2}, y_{n-1}, y_n) \\
&= \square_1 \square_2 f(z_{n-2}|z_1, \dots, z_{n-3}, y_{n-1}, y_n) f(z_1, \dots, z_{n-3}, y_{n-1}, y_n) \\
&= \square_1 \square_2 \underbrace{\frac{f(z_{n-2}|z_1, \dots, z_{n-3}, y_{n-1}, y_n)}{f(y_{n-2}|z_1, \dots, z_{n-3}, y_{n-1}, y_n)}}_{\square_3} f(z_1, \dots, z_{n-3}, y_{n-2}, y_{n-1}, y_n) \\
&\quad \dots \text{ etc.} \\
f(z) &= \square_1 \square_2 \dots \square_{n-1} \cdot f(z_1, y_2, \dots, y_n)
\end{aligned}$$

donde

$$\square_{n-1} = \frac{f(z_{n-(n-2)}|z_1, y_3, \dots, y_n)}{f(y_{n-(n-2)}|z_1, y_3, \dots, y_n)}$$

finalmente

$$\begin{aligned}
f(z_1, y_2, \dots, y_n) &= f(z_1|y_2, \dots, y_n) f(y_2, \dots, y_n) \\
&= f(z_1|y_2, \dots, y_n) \frac{f(y_1, y_2, \dots, y_n)}{f(y_1|y_2, \dots, y_n)} \\
&= \underbrace{\frac{f(z_1|y_2, \dots, y_n)}{f(y_1|y_2, \dots, y_n)}}_{\square_n} f(y_1, y_2, \dots, y_n)
\end{aligned}$$

Por lo tanto

$$f(z) = \square_1 \square_2 \dots \square_{n-1} \square_n f(y)$$

Reescribiendo "en reversa"

$$f(z) = \prod_{i=1}^n \frac{f(z_i|z_1, \dots, z_{i-1}, y_{i+1}, \dots, y_n)}{f(y_i|z_1, \dots, z_{i-1}, y_{i+1}, \dots, y_n)} f(y)$$

Y ahora, usemos el modelo Gaussiano con la especificación condicional antes mencionada (con $y \equiv (\mu_1, \dots, \mu_n)$)

$$f(z_i | z_1, \dots, z_{i-1}, \mu_{i+1}, \dots, \mu_n) = (2\pi\tau_i^2)^{-1/2} \exp[-(z_i - \theta_i)^2 / 2\tau_i^2]$$

en este caso tenemos que

$$\theta_i = \mu_i + \sum_{j=1}^n c_{i,j}(z_j - \mu_j) = \mu_i + \sum_{j=1}^{i-1} c_{i,j}(z_j - \mu_j)$$

por lo que

$$f(z_i | z_1, \dots, z_{i-1}, \mu_{i+1}, \dots, \mu_n) = (2\pi\tau_i^2)^{-1/2} \exp\left[-\left(z_i - \mu_i - \sum_{j=1}^{i-1} c_{i,j}(z_j - \mu_j)\right)^2 / 2\tau_i^2\right]$$

similarmente

$$\begin{aligned} f(\mu_i | z_1, \dots, z_{i-1}, \mu_{i+1}, \dots, \mu_n) &= (2\pi\tau_i^2)^{-1/2} \exp[-(\mu_i - \theta_i)^2 / 2\tau_i^2] \\ &= (2\pi\tau_i^2)^{-1/2} \exp\left[-\left(\sum_{j=1}^{i-1} c_{i,j}(z_j - \mu_j)\right)^2 / 2\tau_i^2\right] \end{aligned}$$

entonces

$$\begin{aligned} \log\left(\frac{f(z)}{f(\mu)}\right) &= -\sum_{i=1}^n \left(z_i - \mu_i - \sum_{j=1}^{i-1} c_{i,j}(z_j - \mu_j)\right)^2 / 2\tau_i^2 + \sum_{i=1}^n \left(\sum_{j=1}^{i-1} c_{i,j}(z_j - \mu_j)\right)^2 / 2\tau_i^2 \\ &= -\sum_{i=1}^n (z_i - \mu_i)^2 / 2\tau_i^2 + \sum_{i=1}^n \left\{ (z_i - \mu_i) \sum_{j=1}^{i-1} c_{i,j}(z_j - \mu_j) \right\} / \tau_i^2 \\ &\quad - \sum_{i=1}^n \left(\sum_{j=1}^{i-1} c_{i,j}(z_j - \mu_j)\right)^2 / 2\tau_i^2 + \sum_{i=1}^n \left(\sum_{j=1}^{i-1} c_{i,j}(z_j - \mu_j)\right)^2 / 2\tau_i^2 \\ &= -\frac{1}{2} \sum_{i=1}^n (z_i - \mu_i)^2 / \tau_i^2 + \sum_{i=1}^n \left\{ \sum_{j=1}^{i-1} (z_i - \mu_i) c_{i,j}(z_j - \mu_j) \right\} / \tau_i^2 \end{aligned}$$

Sean $M = \text{diag}(\tau_1^2, \dots, \tau_n^2)$, $z = (z_1, \dots, z_n)$ y $\mu \equiv (\mu_1, \dots, \mu_n)$,

$$\begin{aligned}\log\left(\frac{f(z)}{f(\mu)}\right) &= -\frac{1}{2}(z-\mu)^T M^{-1}(z-\mu) + \frac{1}{2}(z-\mu)^T M^{-1}C(z-\mu) \\ &= -\frac{1}{2}(z-\mu)^T M^{-1}(I-C)(z-\mu)\end{aligned}$$

donde C es una matriz $nm \times nm$ cuyo (i, j) -ésimo elemento es $c_{i,j}$, $c_{i,j}\tau_j^2 = c_{j,i}\tau_i^2$, y $c_{i,i} = 0$; entonces,

$$f(z) = f(\mu) \exp\left[-\frac{1}{2}(z-\mu)^T M^{-1}(I-C)(z-\mu)\right]$$

de aquí que $f(\mu) = \frac{1}{(2\pi)^{n/2} |(I-C)^{-1}M|^{1/2}}$. Por lo tanto,

$$z \sim \mathbf{N}(\boldsymbol{\mu}, (I-C)^{-1}M).$$

NOTA: Este resultado nos dice que la conjunta de z es normal con media μ y varianza $(I-C)^{-1}M$. Sin embargo, si hubieramos numerado los sitios en otro orden, potencialmente, hubieramos otra forma para $(I-C)^{-1}M$. Para que la especificación condicional sea válida, todas las $n!$ formas de numerar los sitios deben conducir a la misma distribución. Este requerimiento es fuerte.

Anteriormente en el modelo Gaussiano condicionalmente especificado asumimos que $\theta_i = \mu_i + \sum_{j=1}^n c_{i,j}(z_j - \mu_j)$. Ahora vemos que este supuesto es en realidad una condición necesaria para el modelo normal (de hecho, para toda la familia exponencial) cuando se tienen dependencia sólo por parejas. De acuerdo al teorema de Besag (p. 420 de Cressie) tenemos el siguiente modelo,

$$f(z_i|z_j; j \neq i) = (2\pi\tau_i^2)^{-1/2} \exp[-(z_i - \theta_i)^2 / 2\tau_i^2]$$

donde $\theta_i = \theta_i(z_j; j \neq i)$, (aquí lo particularizamos al caso Gaussiano), luego,

$$f(z_i|z_j; j \neq i) = (2\pi\tau_i^2)^{-1/2} \exp[-(z_i^2 - 2\theta_i z_i + \theta_i^2) / 2\tau_i^2] \quad (\text{A1})$$

Ahora, por el Teorema de Hammersley-Clifford y bajo dependencias por parejas, se tiene que

$$f(z) = \frac{1}{C} e^{Q(z)}$$

donde $Q(z) = \sum_{i=1}^n z_i G_i(z_i) + \sum \sum_{i < j} z_i z_j G_{ij}(z_i, z_j)$.

NOTA: El supuesto de dependencia por parejas no implica que los cliques tengan que tener 2 o menos elementos. El Teorema de Hammersley-Clifford dice que los G_{ijjk} pueden ser distintos de cero si z_i, z_j y z_k son vecinos, NO dice que G_{ijjk} tenga que ser diferente de cero.

De A1 se tiene que,

$$\log f(z_i | z_j ; j \neq i) = - (z_i^2 - 2\theta_i z_i + \theta_i^2) / 2\tau_i^2 - \frac{1}{2} \log(2\pi\tau_i^2),$$

en general $\tau_i^2 = \tau_i^2(z_j ; j \neq i)$ pero, estamos considerando el caso de varianza homogénea, luego desarrollando y reagrupando la expresión anterior tenemos que,

$$\begin{aligned} \log f(z_i | z_j ; j \neq i) &= \underbrace{\frac{\theta_i}{\tau_i^2}}_{A_i(z_j ; j \neq i)} \underbrace{z_i}_{B_i(z_i)} - \underbrace{\frac{1}{2\tau_i^2} z_i^2 - \frac{1}{2\tau_i^2} \theta_i^2 - \frac{1}{2} \log(2\pi\tau_i^2)}_{C_i(z_i)} + D_i(z_j ; j \neq i) \end{aligned} \quad (A2)$$

Por otro lado definamos $z_{(i)} = (z_1, z_2, \dots, z_{i-1}, 0, z_{i+1}, \dots, z_n)$, luego,

$$f(z_i | z_j ; j \neq i) = \frac{f(z)}{f(z_j ; j \neq i)} = f(z) \cdot \frac{f(z_i = 0 | z_j ; j \neq i)}{f(z_{(i)})}$$

eso es porque $f(z_j ; j \neq i) = \frac{f(z_{(i)})}{f(z_i = 0 | z_j ; j \neq i)}$, entonces:

$$\frac{f(z)}{f(z_{(i)})} = \frac{f(z_i | z_j ; j \neq i)}{f(z_i = 0 | z_j ; j \neq i)},$$

por lo que,

$$Q(z) - Q(z_{(i)}) = \log f(z_i | z_j ; j \neq i) - \log f(z_i = 0 | z_j ; j \neq i)$$

sustituyendo A2 y simplificando,

$$\begin{aligned}
Q(z) - Q(z_{(i)}) &= A_i(z_j ; j \neq i) B_i(z_i) + C_i(z_i) + D_i(z_j ; j \neq i) & (A3) \\
&\quad - A_i(z_j ; j \neq i) B_i(z_i = 0) - C_i(z_i = 0) - D_i(z_j ; j \neq i) \\
&= A_i(z_j ; j \neq i) [B_i(z_i) - B_i(z_i = 0)] + C_i(z_i) - C_i(z_i = 0)
\end{aligned}$$

Sin embargo, tenemos que,

$$\begin{aligned}
Q(z) - Q(z_{(i)}) &= \sum_{i=1}^n z_i G_i(z_i) + \sum_{i < j} z_i z_j G_{ij}(z_i, z_j) & (A4) \\
&\quad - \sum_{\substack{i \\ i \neq i}} z_i G_i(z_i) - \sum_{\substack{i < j \\ i \neq i}} z_i z_j G_{ij}(z_i, z_j) \\
&= z_i G_i(z_i) + z_i \sum_{j \neq i} z_j G_{ij}(z_i, z_j)
\end{aligned}$$

así, los lados derechos de A3 y A4 son iguales; en particular podemos evaluar en $z_1 = 0, z_2 = 0, \dots, z_{i-1} = 0, z_{i+1} = 0, \dots, z_n = 0$, en este caso:

$$z_i G_i(z_i) = A_i(z_j = 0 ; j \neq i) [B_i(z_i) - B_i(z_i = 0)] + C_i(z_i) - C_i(z_i = 0) \quad (A5)$$

Ahora, consideramos dos sitios que sean vecinos, digamos 1 y 2, veamos la relación que hay entre A3 y A4 con $z_3 = 0, \dots, z_n = 0$,

$$\underbrace{A_1(z_j ; j \neq 1)}_{A_1(z_2, z_3=0, \dots, z_n=0)} [B_1(z_1) - B_1(z_1 = 0)] + C_1(z_1) - C_1(z_1 = 0) = z_1 G_1(z_1) + z_1 z_2 G_{12}(z_1, z_2) \quad (*)$$

y también

$$A_2(z_1, z_3 = 0, \dots, z_n = 0) [B_2(z_2) - B_2(z_2 = 0)] + C_2(z_2) - C_2(z_2 = 0) = z_2 G_2(z_2) + z_1 z_2 G_{12}(z_1, z_2) \quad (**)$$

sustituyendo A5 en (*) obtenemos,

$$\begin{aligned}
& A_1(z_2, z_3 = 0, \dots, z_n = 0) [B_1(z_1) - B_1(z_1 = 0)] + \underline{C_1(z_1)} - \underline{\underline{C_1(z_1 = 0)}} \\
= & A_1(z_j = 0 ; j \neq 1) [B_1(z_1) - B_1(z_1 = 0)] + \underline{C_1(z_1)} - \underline{\underline{C_1(z_1 = 0)}} + z_1 z_2 G_{12}(z_1, z_2)
\end{aligned}$$

por lo tanto,

$$z_1 z_2 G_{12}(z_1, z_2) = [A_1(z_2, z_3 = 0, \dots, z_n = 0) - A_1(z_2 = 0, z_3 = 0, \dots, z_n = 0)] [B_1(z_1) - B_1(z_1 = 0)]. \quad (\#)$$

Similarmente, de (***) se obtiene

$$z_1 z_2 G_{12}(z_1, z_2) = [A_2(z_1, z_3 = 0, \dots, z_n = 0) - A_2(z_1 = 0, z_3 = 0, \dots, z_n = 0)] [B_2(z_2) - B_2(z_2 = 0)]. \quad (\#\#)$$

Las expresiones (#) y (\#\#) son iguales, entonces,

$$\begin{aligned}
& \frac{A_1(z_2, z_3 = 0, \dots, z_n = 0) - A_1(z_2 = 0, z_3 = 0, \dots, z_n = 0)}{B_2(z_2) - B_2(z_2 = 0)} \\
= & \frac{A_2(z_1, z_3 = 0, \dots, z_n = 0) - A_2(z_1 = 0, z_3 = 0, \dots, z_n = 0)}{B_1(z_1) - B_1(z_1 = 0)}
\end{aligned}$$

el lado izquierdo es función sólo de z_2 y el lado derecho sólo de z_1 , y son iguales, entonces deben ser iguales a una constante, digamos θ_{12} , luego

$$z_1 z_2 G_{12}(z_1, z_2) = \theta_{12} [B_1(z_1) - B_1(z_1 = 0)] [B_2(z_2) - B_2(z_2 = 0)],$$

similarmente, para cualesquier dos sitios i y j que sean vecinos cumplen que

$$z_i z_j G_{ij}(z_i, z_j) = \theta_{ij} [B_i(z_i) - B_i(z_i = 0)] [B_j(z_j) - B_j(z_j = 0)]$$

además $\theta_{ij} = \theta_{ji}$.

Por lo tanto de A4 tenemos,

$$\begin{aligned}
Q(z) - Q(z_{(i)}) &= z_i G_i(z_i) + \sum_{j \neq i} z_i z_j G_{ij}(z_i, z_j) \\
&= \underbrace{A_i(z_j = 0 ; j \neq i)}_{\alpha_i} [B_i(z_i) - B_i(z_i = 0)] + C_i(z_i) - C_i(z_i = 0) \\
&\quad + \sum_{j \neq i} \theta_{ij} [B_i(z_i) - B_i(z_i = 0)] [B_j(z_j) - B_j(z_j = 0)] \\
&= \alpha_i B_i(z_i) - \alpha_i B_i(z_i = 0) + C_i(z_i) - C_i(z_i = 0) \\
&\quad + \sum_{j \neq i} \theta_{ij} \left[\begin{array}{c} B_i(z_i) B_j(z_j) - B_i(z_i) B_j(z_j = 0) \\ -B_i(z_i = 0) B_j(z_j) + B_i(z_i = 0) B_j(z_j = 0) \end{array} \right] \\
&= \underline{\alpha_i B_i(z_i)} - \alpha_i B_i(z_i = 0) + C_i(z_i) - C_i(z_i = 0) \\
&\quad + \left[\sum_{j \neq i} \theta_{ij} B_j(z_j) \right] B_i(z_i) - \left[\sum_{j \neq i} \theta_{ij} B_j(z_j = 0) \right] B_i(z_i) \\
&\quad - \left[\sum_{j \neq i} \theta_{ij} B_j(z_j) \right] B_i(z_i = 0) + \left[\sum_{j \neq i} \theta_{ij} B_j(z_j = 0) \right] B_i(z_i = 0)
\end{aligned}$$

reorganizando lo anterior poniendo primero los términos subrayados,

$$\begin{aligned}
Q(z) - Q(z_{(i)}) &= \left[\alpha_i - \sum_{j \neq i} \theta_{ij} B_j(z_j = 0) + \sum_{j \neq i} \theta_{ij} B_j(z_j) \right] B_i(z_i) \\
&\quad + \text{"términos que no contienen } B_i(z_i) \text{"}.
\end{aligned}$$

En la expresión de $Q(z) - Q(z_{(i)})$ en A3, el coeficiente de $B_i(z_i)$ es $A_i(z_j ; j \neq i)$ y, por otro lado, en la igualdad anterior, tenemos otra expresión equivalente para $Q(z) - Q(z_{(i)})$ y el coeficiente para $B_i(z_i)$ es

$$\underbrace{\alpha_i - \sum_{j \neq i} \theta_{ij} B_j(z_j = 0)}_{\mu_i} + \sum_{j \neq i} \theta_{ij} B_j(z_j).$$

Por lo tanto, $A_i(z_j ; j \neq i) = \mu_i + \sum_{j \neq i} \theta_{ij} B_j(z_j)$.

Volviendo a nuestro modelo condicional,

$$f(z_i | z_j ; j \neq i) = (2\pi\tau_i^2)^{-1/2} \exp \left[- (z_i - \theta_i)^2 / 2\tau_i^2 \right],$$

entonces necesariamente $\theta_i = \mu_i + \sum_{j \neq i} \theta_{ij} B_j(z_j)$, esto es

$$\theta_i = \mu_i + \sum_{j \neq i} \theta_{ij} z_j,$$

que es la media condicional que corresponde al modelo autonormal usual.

6.2 ANEXO B

Modelo Ising

Sin pérdida de generalidad consideramos el siguiente modelo espacial para \mathbf{X} la cual es una matriz de dimensión $n \times n$

$$f(\mathbf{x}; \beta) = \exp[-Q(\mathbf{x}; \beta) - \log\{C(\beta)\}]$$

donde β es el parámetro de interés, la función potencial $Q(\cdot; \cdot)$ exhibe la interacción entre componentes de \mathbf{X} y en este caso es $Q(\mathbf{X}; \beta) = -\beta V(\mathbf{x})$ y el factor normalizante es obtenido sumando sobre todas las posibles configuraciones en \mathbf{X} ,

$$C(\beta) = \sum_{\mathbf{y} \in A} \exp\{-Q(\mathbf{y}; \beta)\}$$

donde A es el conjunto de todas las posibles matrices de dimensión $n \times n$.

La densidad condicional de $x_{i,j}$ dado $V(\mathbf{x})$ (suma de los vecinos de la matriz \mathbf{x}), puede ser escrita como

$$\begin{aligned} f(x_{i,j} | V(\mathbf{x}); \beta) &= \frac{f(x_{i,j}, V(\mathbf{x}); \beta)}{f(V(\mathbf{x}); \beta)} \\ &= \frac{f(x_{i,j}, V(\mathbf{x}); \beta)}{\sum_{x_{i,j} \in \{-1,1\}} f(x_{i,j}, V(\mathbf{x}); \beta)} \end{aligned}$$

En este caso consideraremos:

$$V(\mathbf{x}) = \sum_{k=1}^{n-1} \sum_{l=1}^n x_{kl} x_{k+1,l} + \sum_{k=1}^n \sum_{l=1}^{n-1} x_{kl} x_{k,l+1} + \sum_{k=1}^n x_{k1} x_{k,n}$$

donde el primer término es la suma del producto de los vecinos cercanos dentro de las columnas, el segundo término es la suma análoga dentro de los renglones y el tercer término denota la suma de los productos de vecinos entre la primer y la última columna (arreglo cilíndrico). Entonces nuestro modelo condicional anterior quedaría de la siguiente forma:

$$f(x_{i,j}|V(\mathbf{x});\beta) = \frac{\frac{1}{C(\beta)} \exp\{-\beta V(\mathbf{x})\}}{\frac{1}{C(\beta)} \exp\{-\beta V(\mathbf{x};x_{i,j} = -1)\} + \frac{1}{C(\beta)} \exp\{-\beta V(\mathbf{x};x_{i,j} = 1)\}} \quad (5.2.1)$$

Vamos a considerar el siguiente hecho para la observación x_{ij} , sin pérdida de generalidad supondremos que esta observación no se encuentra en las fronteras:

$$\begin{aligned} V(\mathbf{x}) &= \sum_{k=1}^{n-1} \sum_{l=1}^n x_{kl}x_{k+1,l} + \sum_{k=1}^n \sum_{l=1}^{n-1} x_{kl}x_{k,l+1} + \underbrace{\sum_{k=1}^n x_{k1}x_{k,n}}_{\square} \\ &= x_{ij}x_{i+1,j} + x_{i-1,j}x_{ij} + \sum_{\substack{k=1 \\ k \neq i, i-1}}^{n-1} \sum_{\substack{l=1 \\ l \neq j}}^n x_{kl}x_{k+1,l} + \\ &\quad x_{ij}x_{i,j+1} + x_{i,j-1}x_{ij} + \sum_{\substack{k=1 \\ k \neq i}}^n \sum_{\substack{l=1 \\ l \neq j, j-1}}^{n-1} x_{kl}x_{k,l+1} + \square \end{aligned}$$

además:

$$\begin{aligned} V(\mathbf{x};x_{i,j} = -1) &= -x_{i+1,j} - x_{i-1,j} + \sum_{\substack{k=1 \\ k \neq i, i-1}}^{n-1} \sum_{\substack{l=1 \\ l \neq j}}^n x_{kl}x_{k+1,l} + \\ &\quad -x_{i,j+1} - x_{i,j-1} + \sum_{\substack{k=1 \\ k \neq i}}^n \sum_{\substack{l=1 \\ l \neq j, j-1}}^{n-1} x_{kl}x_{k,l+1} + \square \\ &= -V_{i,j} + \sum_{\substack{k=1 \\ k \neq i, i-1}}^{n-1} \sum_{\substack{l=1 \\ l \neq j}}^n x_{kl}x_{k+1,l} + \sum_{\substack{k=1 \\ k \neq i}}^n \sum_{\substack{l=1 \\ l \neq j, j-1}}^{n-1} x_{kl}x_{k,l+1} + \square \end{aligned}$$

donde $V_{i,j}$ es la suma los vecinos de $x_{i,j}$. De forma análoga tenemos:

$$\begin{aligned}
V(\mathbf{x}; x_{i,j} = 1) &= x_{i+1,j} + x_{i-1,j} + \sum_{\substack{k=1 \\ k \neq i, i-1}}^{n-1} \sum_{\substack{l=1 \\ l \neq j}}^n x_{kl} x_{k+1,l} + \\
&\quad + x_{i,j+1} + x_{i,j-1} + \sum_{\substack{k=1 \\ k \neq i}}^n \sum_{\substack{l=1 \\ l \neq j, j-1}}^{n-1} x_{kl} x_{k,l+1} + \square \\
&= V_{i,j} + \sum_{\substack{k=1 \\ k \neq i, i-1}}^{n-1} \sum_{\substack{l=1 \\ l \neq j}}^n x_{kl} x_{k+1,l} + \sum_{\substack{k=1 \\ k \neq i}}^n \sum_{\substack{l=1 \\ l \neq j, j-1}}^{n-1} x_{kl} x_{k,l+1} + \square
\end{aligned}$$

Sustituyendo estas igualdades en (5.2.1) y simplificando tenemos:

$$\begin{aligned}
f(x_{i,j} | V(\mathbf{x}); \beta) &= \frac{\exp\{-\beta x_{i,j} V_{i,j}\} \exp\left\{-\beta \left[\sum_{\substack{k=1 \\ k \neq i, i-1}}^{n-1} \sum_{\substack{l=1 \\ l \neq j}}^n x_{kl} x_{k+1,l} + \sum_{\substack{k=1 \\ k \neq i}}^n \sum_{\substack{l=1 \\ l \neq j, j-1}}^{n-1} x_{kl} x_{k,l+1} + \square \right] \right\}}{[\exp\{\beta V_{i,j}\} + \exp\{-\beta V_{i,j}\}] \exp\{-\beta \blacksquare\}} \\
&= \frac{\exp\{-\beta x_{i,j} V_{i,j}\}}{\exp\{\beta V_{i,j}\} + \exp\{-\beta V_{i,j}\}}
\end{aligned}$$

6.3 ANEXO C

Cociente Metropolis-Hastings

Sea x una variable auxiliar (definida en el mismo espacio que y), con distribución condicional $f(x|\theta, y)$, ahora la posterior es

$$\begin{aligned}
\pi(\theta, x|y) &\propto \pi(\theta, x, y) = f(x|\theta, y) \cdot \pi(\theta, y) \\
&\propto f(x|\theta, y) \pi(\theta) \pi(y|\theta) \\
&= f(x|\theta, y) \pi(\theta) \frac{q_\theta(y)}{Z_\theta},
\end{aligned}$$

por un lado, si marginalizamos sobre x obtenemos $\pi(\theta|y)$, pero, por otro lado, la constante Z_θ todavía se encuentra en la expresión.

Si (θ, x) son los valores actuales siendo generados por el Metropolis-Hastings, donde se propone primero θ' con densidad $q(\theta'|\theta, x)$ y x' con densidad $q(x'|\theta', \theta, x)$ (la elección de q y de $f(x|\theta, y)$ son arbitrarias bajo ciertas consideraciones). Eligiendo la nuestra función instrumental para x' de la siguiente manera,

$$q(x'|\theta', \theta, x) = \pi(x'|\theta') = \frac{q_{\theta'}(x')}{Z_{\theta'}}$$

entonces el cociente Metropolis-Hastings es

$$\begin{aligned} r &= r(\theta', x'|\theta, x) = \frac{\pi(\theta', x'|y)}{\pi(\theta, x|y)} \frac{q(\theta, x|\theta', x')}{q(\theta', x'|\theta, x)} \\ &= \frac{f(x'|\theta', y)}{f(x|\theta, y)} \frac{\pi(\theta')}{\pi(\theta)} \frac{\frac{q_{\theta'}(y)}{Z_{\theta'}}}{\frac{q_{\theta}(y)}{Z_{\theta}}} \frac{q(\theta, x|\theta', x')}{q(\theta', x'|\theta, x)} \\ &= \frac{f(x'|\theta', y)}{f(x|\theta, y)} \frac{\pi(\theta')}{\pi(\theta)} \frac{\frac{q_{\theta'}(y)}{Z_{\theta'}}}{\frac{q_{\theta}(y)}{Z_{\theta}}} \frac{q(\theta|\theta', x')}{q(x'|\theta, x)} \frac{q(x|\theta, \theta', x')}{q(x'|\theta', \theta, x)} \\ &= \frac{f(x'|\theta', y)}{f(x|\theta, y)} \frac{\pi(\theta')}{\pi(\theta)} \cdot \frac{\frac{q_{\theta'}(y)}{Z_{\theta'}}}{\frac{q_{\theta}(y)}{Z_{\theta}}} \cdot \frac{q(\theta|\theta', x')}{q(x'|\theta, x)} \cdot \frac{\frac{q_{\theta}(x)}{Z_{\theta}}}{\frac{q_{\theta'}(x)}{Z_{\theta'}}}, \end{aligned}$$

el cual se simplifica,

$$r = \frac{f(x'|\theta', y)}{f(x|\theta, y)} \frac{\pi(\theta')}{\pi(\theta)} \cdot \frac{q_{\theta'}(y)}{q_{\theta}(y)} \cdot \frac{q(\theta|\theta', x')}{q(x'|\theta, x)} \cdot \frac{q_{\theta}(x)}{q_{\theta'}(x)},$$

que es la expresión que buscábamos.

Bibliografía

- [1] Arnold, B.C., Castillo, E. y Sarabia, J.M. (1999), *Conditional Specification of Statistical Models*. New York: Springer-Verlag, p. 9.
- [2] Baddeley, A. J. and Turner, R. (2000). "Practical maximum pseudo-likelihood for spatial point patterns (with discussion)". *Aust. New Z. J. Statist.*, **42**, 283-322.
- [3] Barndorff-Nielsen, O. E., Kendall, W. S. and van Lieshout, M. C. (1999). *Stochastic Geometry: Likelihood and Computation*. London: Chapman and Hall.
- [4] Besag, J. E. (1974). "Spatial interaction and the statistical analysis of lattice systems (with discussion)". *Journal of the Royal Statistical Society Series B*, **36**, 2, 192-236.
- [5] Besag, J. E. (1977). "Efficiency of pseudo likelihood estimators for simple Gaussian fields". *Biometrika*, **64**, 616-618.
- [6] Besag, J. E. (1986). "On the Statistical Analysis of Dirty Pictures (with discussion)". *Journal of the Royal Statistical Society Series B*, **48**, 2, 259-302.
- [7] Besag, J. E. and Moran, P. A. P (1975). "On the estimation and testing of spatial interaction in Gaussian lattice processes". *Biometrika*, **62**, 555-562.

- [8] Clifford, P. (1990) *Markov random fields in statistics. In Disorder in Physical Systems* (ed. G. Grimmet and D. J. Welsh). Oxford: Clarendon
- [9] Comets, F. (1992). "On Consistency of a Class of Estimators for Exponential Families of Markov Random Fields on the Lattice". *The Annals of Statistics*, **20**, 455-468
- [10] Cressie, N. A. C. (1993) *Statistics for Spatial Data*. New York: Wiley
- [11] Diggle, P. J. (1983). *Statistical Analysis of Spatial Point Patterns*. New York: Academic Press
- [12] Diggle, P.J., Fiksel, T., Grabarnik, P., Ogata, Y., Stoyan, D. and Tanemura, M. (1994). "On parameter estimation for pairwise interaction point processes". *Int. Statist. Rev.*, **62**, 99-117.
- [13] Geyer, C. J. (1991). "Markov Chain Monte Carlo Maximum Likelihood", in *Computing Science and Statistics: 23rd Symposium on the Interface*, ed E. M. Keramidas, Fairfax Station, VA: Interface Foundation of North America, pp. 156-163.
- [14] Goulard, M., Särkkä, A., y Grabarnik, P. (1996). "Parameter Estimation for Marked Gibbs Point Processes Through the Maximum Pseudolikelihood Method", *Scandinavian Journal of Statistics*, **23**, 365-379.
- [15] Ming Gao Gu and Hong-Tu Zhu (2001). "Maximum likelihood estimation for spatial models by Markov chain Monte Carlo stochastic approximation". *Journal of the Royal Statistical Society Series B*, **63**, Part 2, pp. 339-355.
- [16] Hammersley, J. M. y Clifford, P. (1971). "Markov fields on finite graphs and lattices", Unpublished manuscript, Oxford University.
- [17] Jensen, J. L., y Künsch, H. R. (1994). "On Asymptotic Normality of Pseudo-likelihood Estimates for Pairwise Interaction Processes", *Annals of the Institute of Statistical Mathematics*, **46**, 475-486.

- [18] Jensen, J. L., y Møller, J. (1991). "Pseudolikelihood for Exponential Family Models of Spatial Point Processes", *The Annals of Applied Probability*, **1**, 445-461
- [19] Paul Marjoram, John Molitor, Vincent Plagnol, and Simon Tavaré (2003). "Markov chain Monte Carlo without likelihoods". *PNAS*. Vol. 100 No. 26, 15324–15328
- [20] Mase, S. (1995). "Consistency of the Maximum Pseudo-likelihood Estimator of Continuous State Space Gibbsian Processes", *The Annals of Applied Probability*, **5**, 603-612
- [21] Matheron, G. (1962). *Traite de Geostatistique Appliquee, Tome I: Memoires du Bureau de Recherches Geologiques et Minieres, No. 14*. Editions Technip, Paris.
- [22] Matheron, G. (1963a). *Traite de Geostatistique Appliquee, Tome II: Le Krigeage. Memoires du Bureau de Recherches Geologiques et Minieres, No. 24*. Editions Bureau de Recherche Geologiques et Minieres, Paris.
- [23] Matheron, G. (1963a). "Principles of geostatistics", *Economic Geology*, **58**, 1246-1266.
- [24] Moran, P. A. P. (1973a). "A Gaussian Markovian process on a square lattice". *Journal of Applied Probability*, **10**, 54-62.
- [25] Møller, J., Pettitt, A.N., Reeves, R. y Berthelsen, K.K. (2006) "An efficient Markov chain Monte Carlo method for distributions with intractable normalising constants". *Biometrika*, **93**, 2, 451-458.
- [26] Pettitt, A.N. and Friel, N. (2003). "Efficient calculation of the normalizing constant of the autologistic and related models on the cylinder and lattice". *Journal of the Royal Statistical Society Series B*, **65**, Part 1, pp. 235-246.

- [27] Pickard, D. K. (1987). "Inference for Discrete Markov Fields: The simplest Nontrivial Case". *Journal of the American Statistical Association*, Vol. 82, No. 397.
- [28] Ripley, B.D. (1977). "Modelling spatial patterns (with discussion)". *J. R. Statist. Soc. B*, **39**, 172-212.
- [29] Ripley, B.D. (1981). *Spatial Statistics*. New York: Wiley.
- [30] Smith, A. F. M. and Roberts, G. O. (1993). "Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods". *J. R. Statist. Soc. B*, **55**, 3-23.
- [31] Stoyan, D., Kendall, W. S. and Mecke, J. (1987). *Stochastic Geometry and Its Applications*. New York: Wiley.
- [32] Strauss, D.J. (1975). "A model for clustering". *Biometrika*, **62**, 467-475.
- [33] Witten, E. H. T. (1970). "Orthogonal polynomial trend surfaces for irregularly spaced data". *Journal of the International Association for Mathematical Geology*, **2**, 141-152.