

Enfoque bayesiano en la estimación de área pequeña

Luis Enrique Nieto Barajas

En este trabajo se presenta una clase muy general de modelos bayesianos que pueden ser usados para resolver el problema de estimación en áreas pequeñas. Se discuten de manera breve los detalles de implementación de los modelos y se dan sugerencias de paquetes computacionales. Se presenta a detalle una aplicación en la estimación de indicadores de pobreza multidimensional para los municipios del estado de México, usando información de la ENIGH y del Censo 2010.

Palabras clave: efectos espaciales, inferencia bayesiana, modelos mixtos, normal multivariada, pobreza multidimensional.

Recibido: 25 de enero de 2013

Aceptado: 16 de abril de 2013

In this work we present a general class of Bayesian models to be used to solve the small area estimation problem. We briefly discuss model implementation details and give suggestions of computing packages. We present a detailed application of estimation of multidimensional poverty indicators for the municipalities of the State of Mexico, using information from the ENIGH and Census 2010.

Key words: Bayesian inference, mixed models, multivariate normal, multidimensional poverty, spatial effects.

Introducción

Una forma de conocer una característica (parámetro) que resuma el comportamiento de un grupo de individuos es mediante la realización de una encuesta. Si la población objetivo es muy grande, se puede seleccionar a un grupo menor de individuos a los cuales aplicarles la encuesta y poder, así, tener una idea de la característica de interés. Para minimizar costos y poder conocer con alta precisión el valor del parámetro en la población, la selección de individuos se debe hacer de manera probabilística mediante lo que se conoce como diseño muestral, y el proceso de estimación del parámetro se tiene que hacer con técnicas de estadística inferencial que van acordes con el diseño. Este proceso se realiza mediante estimadores directos basados en el diseño haciendo uso de los pesos muestrales. Existe una literatura muy vasta para seleccionar el mejor diseño muestral de acuerdo con la naturaleza del parámetro a estimar, ya sea un total, una proporción o una media (ver, por ejemplo, Cochran, 1977; Särndal *et al.*, 2003).

Si dividimos a la población en dominios o subpoblaciones, un diseño muestral que originalmente estaba planeado para estimar un parámetro en la población completa es muy posible que no sea útil para estimarlo en cada una de éstos. Quizás, para algunos dominios, se cuente con suficiente información muestral para producir estimaciones directas del parámetro con precisión razonable. A estos dominios se les conoce también como áreas *grandes*. Por otro lado, un área es considerada *pequeña* si no se tiene suficiente información muestral para producir estimaciones directas con precisión razonable (Rao, 2003).

El proceso de estimación en las áreas pequeñas se basa en pedir prestada información de las áreas grandes para las cuales sí fue posible obtener estimadores directos. Éstos son llamados indirectos y, por lo general, hacen uso de modelos estadísticos que ligan los datos de las distintas áreas. Adicionalmente, se hace uso de otras fuentes de información para lograr una mejor precisión en las estimaciones.

Antes de proceder, especificamos la notación de las distribuciones que usaremos: $N(\mu, \sigma^2)$ denota una normal con media μ y varianza σ^2 ; $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ una normal p -variada con vector de medias $\boldsymbol{\mu}$ y matriz de varianzas-covarianzas $\boldsymbol{\Sigma}$; $Be(\alpha, \beta)$ una beta con media $\alpha/(\alpha + \beta)$; $Ga(\alpha, \beta)$, una gamma con media α/β ; e $IGa(\alpha, \beta)$, una gamma invertida con una media $\beta/(\alpha - 1)$.

Inferencia bayesiana

La estadística bayesiana es una forma alternativa a la frecuentista de hacer inferencia sobre los parámetros desconocidos de un modelo. Tiene sus bases en la teoría de decisión, lo cual fundamenta de manera axiomática los procesos inferenciales al ser planteados como problemas de decisión (De Groot, 2004).

Como parte del proceso inferencial, es necesario cuantificar la incertidumbre sobre los parámetros desconocidos del modelo mediante distribuciones de probabilidad. Esta cuantificación puede reflejar las creencias del estadístico, si es que las hay, o ser el reflejo del desconocimiento por completo de los valores posibles del parámetro. Como consecuencia de ésta, tanto las variables aleatorias observables como los parámetros fijos son descritos con distribuciones de probabilidad, lo que simplifica el proceso inferencial.

La cuantificación inicial (antes de observar la muestra) que se hace sobre los parámetros $f(\theta)$ debe ser actualizada mediante la información muestral $\underline{X} = (X_1, X_2, \dots, X_n)$ proveniente del modelo $f(x|\theta)$. Esto se hace con el Teorema de Bayes obteniéndose, así, la cuantificación final o posterior $f(\theta|\underline{x})$ que combina y resume la información inicial y la muestral, es decir, $f(\theta|\underline{x}) = f(\underline{x}|\theta) f(\theta)/f(\underline{x})$.

El Teorema de Bayes tiene una expresión matemática simple, sin embargo, la obtención analítica de la distribución final de los parámetros se puede complicar debido al cálculo de la constante de normalización $f(\underline{x})$. Gracias a los avances computacionales recientes, y en especial a los algorit-

mos de simulación Monte Carlo vía cadenas de Markov (MCMC), es posible obtener características de cualquier distribución final mediante métodos de simulación sin necesidad de calcular la constante de normalización (Chen *et al.*, 2000).

En la actualidad, existen varias rutinas de distribución libre que implementan modelos bayesianos. Muchas de ellas se pueden encontrar, por ejemplo, en el ambiente de trabajo R (R Development Core Team, 2012). Éstas, al ser de distribución libre, proveen una ventaja para el usuario porque pueden ser modificadas de acuerdo con las necesidades específicas. Adicionalmente, es posible realizar inferencias bayesianas en un paquete llamado OpenBUGS (<http://www.openbugs.info/>); sin embargo, se debe ser muy cuidadoso al usar este paquete porque los algoritmos de simulación de la distribución final son elegidos por el mismo paquete mediante *sistemas expertos* y, en caso de ocurrir errores en el proceso, no es trivial su modificación.

Modelos bayesianos

Los primeros en la estimación de área pequeña se propusieron hace más de 30 años. Algunos de los trabajos más relevantes son los de Fay y Herriot (1979), quienes propusieron el primer modelo lineal mixto; Ghosh *et al.*, (1998), ellos usaron los lineales generalizados y, más recientemente, You y Rao (2002), quienes utilizan modelos bayesianos jerárquicos.

Considera una población dividida en n dominios o áreas y sea θ_i el parámetro de interés en el dominio i para $i = 1, \dots, n$. Sea θ_i^D el estimador directo basado en información específica del área i , haciendo uso de los factores de expansión del diseño muestral o incluso de modelos. Nuestro objetivo es producir estimadores indirectos θ_i^I del área i , mediante el uso de un modelo.

Sea $Y_i = \theta_i^D$ la estimación directa del indicador o parámetro de interés en el dominio i , la cual estará disponible sólo para las áreas grandes y no así para

las pequeñas. Sea $X_i = (X_{1i}, \dots, X_{pi})$ un vector de dimensión p de variables explicativas del dominio i , las cuales pueden incluir un intercepto. Por lo general, estas variables explicativas estarán disponibles en todos los dominios.

La mayoría de los modelos usados para la estimación de área pequeña pertenecen a la clase general de modelos lineales generalizados mixtos. Un modelo se considera mixto si contiene efectos fijos, derivados, por lo común, de variables explicativas, y efectos aleatorios, que representan otras fuentes de incertidumbre no capturadas por las variables explicativas. Para ilustrar, consideremos el caso normal y supongamos, además, que los estimadores directos Y_{it} se conocen a lo largo de una ventana de tiempo $t = 1, \dots, T$. Entonces, tenemos:

$$Y_{it} | \theta_{it}, \sigma_{it}^2 \sim N(\theta_{it}, \sigma_{it}^2), i = 1, \dots, n. \quad (1)$$

Es decir, estamos suponiendo que los estimadores directos Y_{it} se encuentran alrededor de los parámetros reales θ_{it} con cierto error medido a través de la varianza σ_{it}^2 . El parámetro θ_{it} se expresará de manera mixta como:

$$\theta_{it} = \beta' x_i + v_i + u_i + z_i \quad (2)$$

donde:

- β determina la importancia de las variables explicativas x_i en los efectos fijos.
- v_i es un efecto aleatorio que captura la heterogeneidad específica del dominio i , por lo general se considera $v_i \sim N(0, \sigma_v^2)$.
- u_i es un efecto aleatorio espacial que hace que los dominios vecinos compartan información. Existen dos alternativas comunes para especificar distribuciones espaciales. Los modelos condicionalmente autorregresivos, $\mathbf{u} \sim CAR(\rho, \sigma_u^2)$, o los simultáneamente autorregresivos, $\mathbf{u} \sim SAR(\rho, \sigma_u^2)$, con $\mathbf{u}' = (u_1, u_2, \dots, u_n)$. Ambos están definidos por distribuciones normales multivariadas con media cero, parámetro de asociación ρ , pero con

distintas especificaciones para la matriz de varianzas y covarianzas. Para más detalles ver Banerjee *et al.* (2004).

- z_t es un efecto aleatorio temporal que captura las dependencias a lo largo del tiempo. Una forma común de especificar este efecto es mediante un modelo lineal dinámico, $z_t = \alpha z_{t-1} + \epsilon_t$ con $\epsilon_t \sim N(0, \sigma_\epsilon^2)$ (West y Harrison, 1997).
- σ_{it}^2 es la varianza del error de medición.

Cuando se conoce el estimador directo Y_{it} , σ_{it}^2 es el error de estimación determinado por el diseño muestral. Alternativamente, σ_{it}^2 puede ser especificado como $\sigma_{it}^2 = c_i \sigma^2$, con c_i un factor de escala que determina la importancia de la observación Y_{it} en el proceso de estimación de θ_{it} . La inclusión de los efectos aleatorios temporales en el modelo permite compartir información entre estimaciones del mismo parámetro en momentos del tiempo previos para el mismo dominio. En caso de que los estimadores directos Y_{it} tengan un soporte no continuo y no acotado y no sea válido el supuesto de normalidad, es posible buscar una transformación adecuada que los lleve a cumplir con los supuestos del modelo. De manera alterna, se puede recurrir a modelos lineales generalizados (McCullagh y Nelder, 1989), donde se cuenta con cualquier miembro de la familia exponencial para definir, por ejemplo, regresiones logísticas, Poisson, gamma, etc., de tal manera que el rango de valores de los estimadores directos Y_{it} y los supuestos del modelo sean compatibles.

El modelo descrito por las ecuaciones (1) y (2) sirve para representar el comportamiento aleatorio de las cantidades observables (al menos para algunos dominios). Como se mencionó en el apartado *Inferencia bayesiana*, el proceso inferencial bayesiano requiere que se especifique el conocimiento inicial que se tiene sobre los parámetros desconocidos, que en este caso son $\varphi = (\beta, \sigma^2, \sigma_v^2, \rho, \sigma_w^2, \alpha, \sigma_\epsilon^2)$. Esto se hace a través de familias paramétricas conocidas y, en algunos casos, para simplificar los cálculos posteriores se usan distribuciones iniciales llamadas (condicionalmente) conjugadas. Éstas tienen la ventaja de que la distribución (condicio-

nal) final pertenece a la misma familia que la inicial. Un ejemplo de distribuciones iniciales es:

$$\beta \sim N_p(\mathbf{b}_0, \mathbf{B}_0), \rho \sim Be(r_0, r_1), \alpha \sim Ga(a_0, a_1)$$

$$\text{y } \sigma^2, \sigma_v^2, \sigma_w^2, \sigma_\epsilon^2 \sim IGa(s_0, s_1).$$

En caso de que no se cuente con información inicial válida, los hiperparámetros $\mathbf{b}_0, \mathbf{B}_0, r_0, r_1, a_0, a_1, \sigma^2, \sigma_v^2, \sigma_w^2, \sigma_\epsilon^2$, son especificados de tal manera que reflejen una alta incertidumbre. Esto se logra haciendo que la varianza inicial sea grande. Este tipo de distribuciones iniciales son llamadas no informativas, y su fundamento es el de no influenciar la información contenida en los datos en el proceso inferencial.

De manera inicial, por lo general, todos los parámetros del modelo se consideran independientes. La interacción entre ellos dentro de la especificación (1)-(2) hace que los parámetros lleguen a tener cierta dependencia en la distribución final.

Cuando las distribuciones iniciales se eligen de familias (condicionalmente) conjugadas, las finales condicionales son fáciles de obtener. Bajo ciertas condiciones de regularidad, éstas caracterizan el comportamiento posterior conjunto de todo el vector de parámetros φ ; sin embargo, para hacer inferencias sobre el comportamiento posterior de φ , se requerirán de métodos de simulación MCMC y en específico del muestreador de Gibbs (ver, por ejemplo, Chen *et al.*, 2000), el cual se basa en las distribuciones finales condicionales.

Aplicación a indicadores de pobreza multidimensional

Definiciones y contextos de ley

El Consejo Nacional de Evaluación de la Política de Desarrollo Social (CONEVAL), como parte de su misión de evaluar las políticas de desarrollo social, establece los lineamientos y criterios para la definición, identificación y medición de la pobreza en México.

La *Ley General de Desarrollo Social*, promulgada en el 2004, asienta que la pobreza se debe medir desde distintas dimensiones. Con la ayuda de especialistas científicos, el CONEVAL (2009) presenta una metodología para la medición multidimensional de la pobreza en México basada en un indicador bidimensional: el bienestar económico por un lado y las carencias en derechos por otro. El primero se mide en términos del ingreso corriente per cápita. Por otro lado, para la medición de las carencias en derechos se creó un índice de privación social basado en seis características que son: rezago educativo promedio en el hogar, acceso a los servicios de salud, acceso a la seguridad social, calidad y espacios de la vivienda, acceso a los servicios básicos de la vivienda y acceso a alimentación. Finalmente, con base en ciertos umbrales para las dos dimensiones, se clasifica a las personas en una de cuatro posibilidades: pobres, vulnerables por carencias sociales, vulnerables por ingresos y no pobres.

El CONEVAL fue requerido por ley a producir indicadores de pobreza multidimensional de manera bianual para todos los estados de la República Mexicana y quinquenal para los municipios. Para tal efecto, se diseñó el Módulo de Condiciones Socioeconómicas de la Encuesta Nacional de Ingresos y Gastos de los hogares (MCS-ENIGH). Este módulo se levantó, por primera vez, en el 2008 y, posteriormente, en el 2010. El diseño muestral de la Encuesta tiene el propósito de producir estimadores de los indicadores de pobreza a nivel estatal; sin embargo, es de interés también producirlos a nivel municipal. Nos encontramos aquí en un problema de estimación de área pequeña debido a que muchos municipios no tendrán suficiente representación en la muestra como para obtener estimaciones con alta precisión. De hecho, muchos municipios tendrán tamaño de muestra cero.

Modelo bayesiano multivariado

Además de la clasificación de la población en alguna de las cuatro categorías de pobreza, es de interés para el CONEVAL medir las incidencias en cada

una de las siete variables que definen la pobreza multidimensional a nivel municipal.

Sean θ_{ij}^p el estimador directo de θ_{ij} , la proporción de personas con la carencia j en el municipio i , para $i = 1, \dots, n$ y $j = 1, \dots, 7$. Cada uno de estos siete indicadores corresponden (en orden) a: proporción de personas con rezago educativo, proporción de personas sin acceso a servicios de salud, proporción de personas sin acceso a seguridad social, proporción de personas sin calidad y espacio de vivienda, proporción de personas sin servicios básicos de vivienda, proporción de personas sin acceso a alimentación, e ingreso corriente per cápita.

La idea es definir un modelo similar al descrito en (1), pero que reconozca, por un lado, la naturaleza multivariada de la medición de la pobreza y, por el otro, el espacio de los parámetros. Como nuestros parámetros de interés toman valores en una escala (0, 1), exceptuando el ingreso corriente per cápita, proponemos una transformación logística, para los correspondientes estimadores directos, de la siguiente manera:

$$Y_{ij} = \log\left(\frac{\theta_{ij}^p}{1-\theta_{ij}^p}\right), \text{ para } j = 1, \dots, 6,$$

$$\text{y } Y_{ij} = \log\left(\frac{\theta_{ij}^p}{\tau-\theta_{ij}^p}\right), \text{ para } j = 7,$$

donde τ es una cota superior para los ingresos per cápita estimados.

Sea $Y_i' = (Y_{i1}, Y_{i2}, \dots, Y_{im})$ el vector de estimadores directos transformados para los $m = 7$ indicadores de pobreza en el municipio i . Sea $\vartheta_i' = (\vartheta_{i1}, \vartheta_{i2}, \dots, \vartheta_{im})$ un vector de parámetros y sea $X_i' = (X_{i1}, X_{i2}, \dots, X_{ip})$ un vector de dimensión p de variables explicativas proveniente del censo para el municipio i . Entonces, proponemos un modelo multivariado de la forma

$$Y_i|\vartheta_i, \Sigma \sim N_m(\vartheta_i, c_i \Sigma), i = 1, \dots, n,$$

donde:

- Σ es una matriz de varianzas y covarianzas de dimensión $m \times m$.

- c_i es un factor de escala para la varianza de Y_i . Valores pequeños (grandes) de c_i indican mayor (menor) confianza en la observación Y_i .

El parámetro ϑ_i se expresa de manera mixta como:

$$\vartheta_i = \mathbf{B}\mathbf{x}_i + v_i\mathbf{1} + u_i\mathbf{1}$$

donde:

- \mathbf{B} es una matriz de coeficientes de regresión de dimensión $m \times p$.
- $\mathbf{1}$ es un vector de unos de dimensión $m \times 1$.
- $\mathbf{v}' = (v_1, \dots, v_n) \sim N_n(\mathbf{0}, \sigma_v^2\mathbf{I})$ determinan los efectos aleatorios específicos para cada municipio, con \mathbf{I} la matriz identidad de dimensión $n \times n$.
- $\mathbf{u}' = (u_1, \dots, u_n) \sim CAR(\rho, \sigma_u^2) \equiv N_n(\mathbf{0}, \sigma_u^2(\mathbf{D}_w - \rho\mathbf{W})^{-1})$, son los efectos aleatorios espaciales, uno para cada municipio, donde $\rho \in (0, 1)$ es un parámetro que determina el grado de dependencia espacial, valores cercanos a 1 producen mayor dependencia espacial, pero si $\rho = 1$ el modelo *CAR* se vuelve impropio; σ_u^2 es un parámetro no negativo; $\mathbf{W} = (w_{ij})$ es la matriz de vecindades con $w_{ij} = 1$ si los municipios i y j son vecinos y $w_{ij} = 0$ en otro caso; $\mathbf{D}_w = \text{diag}(w_{1+}, \dots, w_{n+})$ con $w_{i+} = \sum_{j=1}^n w_{ij}$.

Resulta matemáticamente conveniente reescribir el modelo para todas las Y_{ij} a través de una distribución normal matricial (ver, por ejemplo, Rowe, 2003), es decir,

$$Y|\mathbf{B}, \Sigma \sim N_{n \times m}(\mathbf{X}\mathbf{B}' + \mathbf{v}'\mathbf{1} + \mathbf{u}'\mathbf{1}, \mathbf{C} \otimes \Sigma),$$

con $\mathbf{Y}' = (\mathbf{Y}'_1, \dots, \mathbf{Y}'_n)$, una matriz de dimensión $n \times m$, $\mathbf{X}' = (\mathbf{X}'_1, \dots, \mathbf{X}'_n)$, matriz de dimensión $n \times p$, $\mathbf{C} = \text{diag}(c_1, \dots, c_n)$ y \otimes el producto de Kronecker.

El modelo bayesiano se completa al especificar las distribuciones iniciales para los parámetros desconocidos, que en este caso son \mathbf{B} , Σ , σ_v^2 , y σ_u^2 . Tanto el parámetro ρ como los c_i son considerados fijos. La distribución inicial para el par (\mathbf{B}, Σ) es una

Normal-Whishart invertida, i.e., $\mathbf{B}|\Sigma \sim N_{n \times p}(\mathbf{B}_0, \mathbf{D} \otimes \Sigma)$, y $\Sigma \sim IW(\mathbf{Q}_0, m, \nu)$. Las distribuciones iniciales para las varianzas de los efectos aleatorios específicos y espaciales son distribuciones gamma invertidas independientes, i.e., $\sigma_v^2 \sim IGa(s_0, s_1)$ y $\sigma_u^2 \sim IGa(s_0, s_1)$.

Con esto especificamos por completo el modelo multivariado. Todo el conocimiento sobre las cantidades desconocidas queda representado mediante la distribución final $f(\mathbf{B}, \Sigma, \mathbf{v}, \mathbf{u}, \sigma_v^2, \sigma_u^2 | \mathbf{Y}, \mathbf{X})$ que se obtiene a través del Teorema de Bayes. Notamos que los efectos aleatorios \mathbf{v} y \mathbf{u} son tratados, en el enfoque bayesiano, como si fueran parámetros desconocidos, por lo cual son parte de la distribución final.

Ésta es difícil de manejar debido a la presencia de los efectos aleatorios. Para poder hacer inferencia con ella y tomar decisiones, se recurrirá a los métodos de simulación MCMC. Con el fin de implementar el muestreador de Gibbs, es necesario tener las distribuciones condicionales completas, las cuales se encuentran descritas en el *Apéndice: muestreador de Gibbs*.

Consideraciones en la implementación

Existen dos importantes para la implementación del modelo: una es el tratamiento de las estimaciones directas no disponibles para los municipios *pequeños* y la otra, la definición de vecindades necesarias para los efectos espaciales.

La distribución final de los parámetros desconocidos para nuestro modelo multivariado supone que se cuenta con la información Y_{ij} de todos los municipios; sin embargo, debido a la naturaleza del problema de áreas pequeñas, para muchos de éstos no se contará con estimaciones directas del parámetro de interés. Para tener una idea de la magnitud del problema, con los datos provenientes del MCS-ENIGH 2010, de los 212 municipios del estado de Veracruz de Ignacio de la Llave, sólo para 28% de ellos se tienen estimaciones directas de los indicadores de pobreza multidimensional.

La solución a este problema de datos faltantes se obtiene de manera natural, en el enfoque bayesiano, con el uso de la distribución predictiva final, la cual caracteriza el comportamiento de la variable no observada mediante el uso de la información de las variables que sí se observaron. La distribución predictiva final se obtiene con un proceso de marginalización de los parámetros desconocidos usando la distribución final de los parámetros de la siguiente manera:

$$f(\mathbf{y}_i | \mathbf{x}_i, \text{datos}) = \int \dots \int N(\mathbf{y}_i | \mathbf{B}\mathbf{x}_i + \mathbf{v}_i \mathbf{1} + \mathbf{u}_i \mathbf{1}, c_i \Sigma) f(\mathbf{B}, \Sigma, \mathbf{v}, \mathbf{u}, | \mathbf{Y}, \mathbf{X}) d\mathbf{B} d\Sigma d\mathbf{v} d\mathbf{u}_i.$$

Las integrales múltiples de la expresión anterior parecieran presentar un reto en el proceso inferencial, sin embargo, ellas nunca se resuelven ya que la inferencia se realiza mediante simulación. Sólo basta con incluir una etapa más en el muestreador de Gibbs descrito en el *Apéndice: muestreador de Gibbs*. Supongamos que $(\mathbf{B}^{(k)}, \Sigma^{(k)}, \mathbf{v}_i^{(k)}, \mathbf{u}_i^{(k)})$ es el estado de la cadena en la iteración (k) . Entonces, dados estos valores, es posible generar un valor de la distribución predictiva final simulando un valor $\mathbf{Y}_i^{(k)}$ de la distribución $N(\mathbf{B}^{(k)}\mathbf{x}_i + \mathbf{v}_i^{(k)}\mathbf{1} + \mathbf{u}_i^{(k)}\mathbf{1}, c_i \Sigma^{(k)})$. Al terminar el simulador de Gibbs se contará con una muestra de valores predichos para las variables Y_{ij} de los municipios pequeños.

Otro aspecto importante en la implementación de nuestro modelo es la definición de municipios vecinos para la componente espacial. Pareciera trivial, al ver la localización de los municipios de un estado en un mapa, identificar aquellos que son vecinos; sin embargo, ésta sería una tarea artesanal complicada. La realidad es que no existe disponible en ninguna base de datos la estructura de vecindades (colindancias) de los municipios en el país. Hay, no obstante, información georreferenciada (longitud, latitud y altitud) de la localización geográfica del centro de la cabecera municipal.

La información de localización de los centros de los municipios es útil, pero no suficiente para recuperar la estructura de vecindades real existente en un estado, pero esto nos da la libertad

de definir estructuras de vecindades alternativas usando criterios de cercanía geográfica entre los centros y otros criterios. Después de proponer varias alternativas, se llegó a la conclusión de que la más adecuada correspondía a la definición de vecinos como aquellos que satisfacían las siguientes propiedades:

- $w_{ij} = 1$ si el municipio i es uno de los dos municipios más cercanos al municipio j , o si el municipio j es uno de los dos vecinos más cercanos al municipio i , o
- $w_{ij} = 1$ si el municipio j se encuentra en la posición del tercer al quinto vecino más cercano al municipio i , o viceversa, y si además los municipios i y j son del mismo tipo rural o urbano, o
- $w_{ij} = 1$ si el municipio j se encuentra en la posición del tercer al quinto vecino más cercano al municipio i , o viceversa, y si además la distancia entre estos municipios es menor a 1 000 (en las unidades de medición de las coordenadas geográficas), y
- $w_{ij} = 0$ en otro caso, incluyendo el caso en que $i = j$.

Resultados

Para ilustrar consideremos al estado de México, el cual está dividido en 125 municipios. El MCS-ENIGH 2010 produjo estimaciones de los indicadores de pobreza a nivel estatal con alta precisión. Al hacer una introspección, se observó que únicamente 58 municipios (46%) estaban representados en la muestra y para ellos sí fue posible producir estimaciones de los indicadores de pobreza; sin embargo, para el resto de éstos, es decir 67 (54%), será necesario estimar sus indicadores mediante nuestro modelo.

La información de las variables explicativas, disponible para todos los 125 municipios, fue obtenida del Censo 2010. Se tomaron un total de $p = 68$ variables que consideramos importantes para explicar las condiciones de pobreza de los municipios. Estas variables incluyen características de la

vivienda, condiciones sociodemográficas de los miembros de la familia, acceso a servicios públicos y seguridad social, etcétera. Vale la pena resaltar la inclusión de la variable altitud como una condición puramente territorial que podría explicar algunas condiciones de pobreza.

Se implementó el modelo multivariado usando las siguientes especificaciones. Como no se contaba con los errores estándar de las estimaciones, se tomaron dos valores $c_i = 1$ y $c_i = 100000/M_i$ para comparar, donde M_i es el número de personas en el municipio i . $\mathbf{B}_0 = \mathbf{0}$ matriz de ceros de dimensión $m \times p$ y $\mathbf{D} = \text{diag}(d, \dots, d)$ de dimensión $p \times p$ con $d \in \{1, 1000\}$ para comparar. Esto implica dos escenarios de poca ($d = 1$) o mucha ($d = 1000$) incertidumbre en la especificación inicial de \mathbf{B} . $\mathbf{Q}_0 = \mathbf{I}$, matriz identidad de dimensión $m \times m$, y $v = 2m + 3$, los cuales especifican una distribución inicial Wishart invertida para Σ con media dada por la matriz identidad y varianza grande. Para el parámetro de asociación en el modelo CAR de los efectos espaciales, se consideraron dos opciones $\rho \in \{0.80, 0.95\}$. Ambos casos implican un modelo propio, pero con moderada y alta asociación espacial. Entre más cercano a 1 sea ρ se permite más intercambio de información entre municipios vecinos. Finalmente, las distribuciones iniciales gammas invertidas para las varianzas σ_v^2 y σ_u^2 se especificaron con $(s_0, s_1) = (2, 1)$ que implican media 1 y varianza grande.

Se implementó un muestreador de Gibbs con 5 mil iteraciones y un periodo de calentamiento de 500, quedando un total de 4 500 para producir las estimaciones. La convergencia de la cadena se determinó de manera informal visualizando las gráficas de promedios ergódicos de algunos parámetros y se concluyó que las 5 mil iteraciones son suficientes. Para comparar entre las distintas especificaciones iniciales se usó una medida que resume variabilidad y sesgo en las predicciones llamada *medida L* (Ibrahim y Laud, 1994). Si n^* es el número de áreas grandes en la muestra, la *medida L* se define como:

$$L(\pi) = \sum_{i=1}^{n^*} \sum_{j=1}^m \text{Var}(Y_{ij}^F | \mathbf{y}) + \pi \sum_{i=1}^{n^*} \sum_{j=1}^m \{E(Y_{ij}^F | \mathbf{y}) - y_{ij}\}^2,$$

donde Y_{ij}^F es el valor predicho de la observación Y_{ij} y $\pi \in [0, 1]$ es un ponderador que establece un compromiso entre la varianza y el sesgo. Valores pequeños de esta medida indican un mejor modelo. La tabla 1 resume las medidas L obtenidas con las distintas especificaciones iniciales. Como se puede observar, el modelo es poco sensible a los valores de ρ , pero muy sensible a los de c_i y d . El mejor modelo se obtiene con $c_i = 1$, $d = 1000$ y $\rho = 0.95$. Usaremos esta especificación inicial para obtener los estimadores.

Tabla 1

Medidas L para distintos valores $\pi = 0, 0.5, 1$ y para distintas especificaciones iniciales del modelo. $c_i = 100000$ y M_i es el número de personas en el municipio i

c_i	d	ρ	$L(0)$	$L(0.5)$	$L(1)$
1	1000	0.95	172.33	206.61	240.85
1	1000	0.80	172.58	206.71	240.89
1	1	0.95	351.22	473.93	596.64
1	1	0.80	352.37	475.53	598.68
c/M_i	1000	0.95	284.74	355.29	425.84
c/M_i	1000	0.80	287.07	357.18	427.30
c/M_i	1	0.95	752.75	913.29	1073.82
c/M_i	1	0.80	754.95	915.47	1075.98

Para tener una idea de la precisión en la estimación, se compararon las estimaciones obtenidas del modelo con las estimaciones directas, para aquellos municipios en la muestra de la ENIGH. La figura 1 muestra gráficas de las distribuciones predictivas finales para los siete indicadores de pobreza en el municipio 15008 (Amatepec). Las líneas verticales rojas punteadas corresponden a intervalos de credibilidad a 95%, mientras que la roja sólida indica la mediana como estimador puntual del indicador. La línea vertical negra corresponde con el valor de la estimación directa. Se observa que las estimaciones directas y las obtenidas con el modelo son muy parecidas, lo cual indica que el modelo se está ajustando de manera adecuada a los datos. Una ventaja del modelo, además de producir estimaciones de los indicadores de pobreza para todos los municipios pequeños que no están en la muestra, es que para los municipios en la muestra que por alguna razón no se pudo estimar directamente alguno de sus indicadores, también es po-

sible hacerlo con el modelo. Éste es el caso del indicador de carencias en condiciones de vivienda (*ic_cv*), el cual tiene un valor observado de cero; en cambio, el modelo estima un valor positivo.

Por razones de espacio, las estimaciones obtenidas con el modelo para los siete indicadores de pobreza en todos los 125 municipios del estado de México se presentan de manera gráfica en las figuras de la 2 a la 5. Aquí sólo comentamos con detalle el caso del municipio 15052 (Malinalco), el cual no se encuentra en la muestra de la ENIGH. Usando las medianas de las distribuciones finales como estimadores puntuales se obtiene que la proporción de personas con rezago educativo es de 28%; la de sin acceso a servicios de salud, de 30%; la de sin acceso a seguridad social, de 68%; la de sin calidad y espacio de vivienda, de 7%; la de sin servicios básicos de vivienda, de 0.4%; y la de sin acceso a alimentación, de 21%; así como, finalmente, el ingreso corriente per cápita, de 1.049 millones de pesos.

Figura 1

Distribuciones predictivas para los siete indicadores de pobreza en el municipio 15008 (Amatepec)

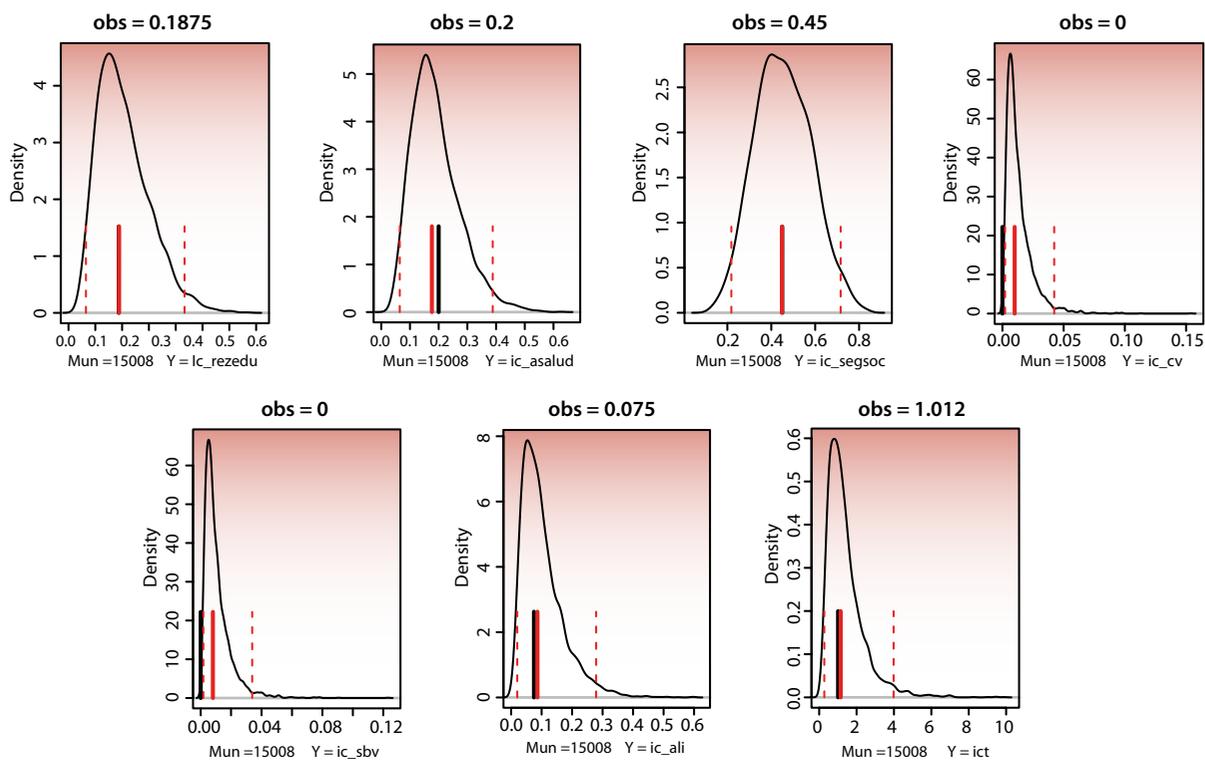


Figura 2

Estimaciones bayesianas del indicador de rezago educativo (izquierda) e indicador de carencia en acceso a servicios de salud (derecha)

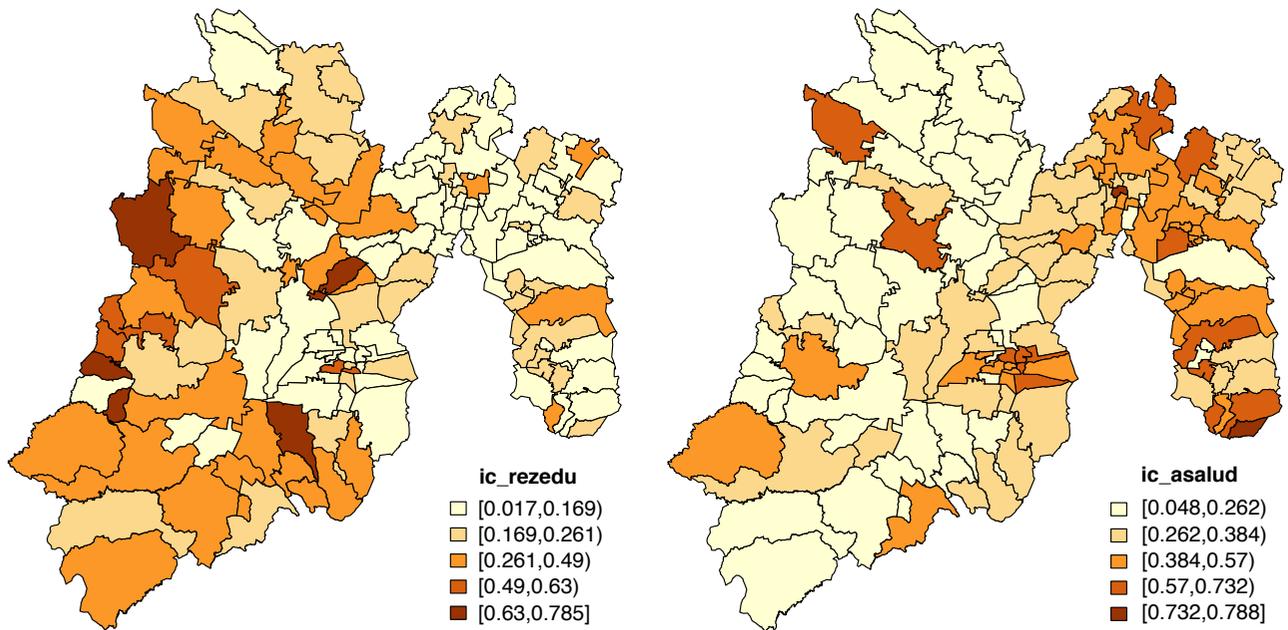


Figura 3

Estimaciones bayesianas del indicador de carencia en acceso a seguridad social (izquierda) e indicador de carencia en calidad y espacio de vivienda (derecha)

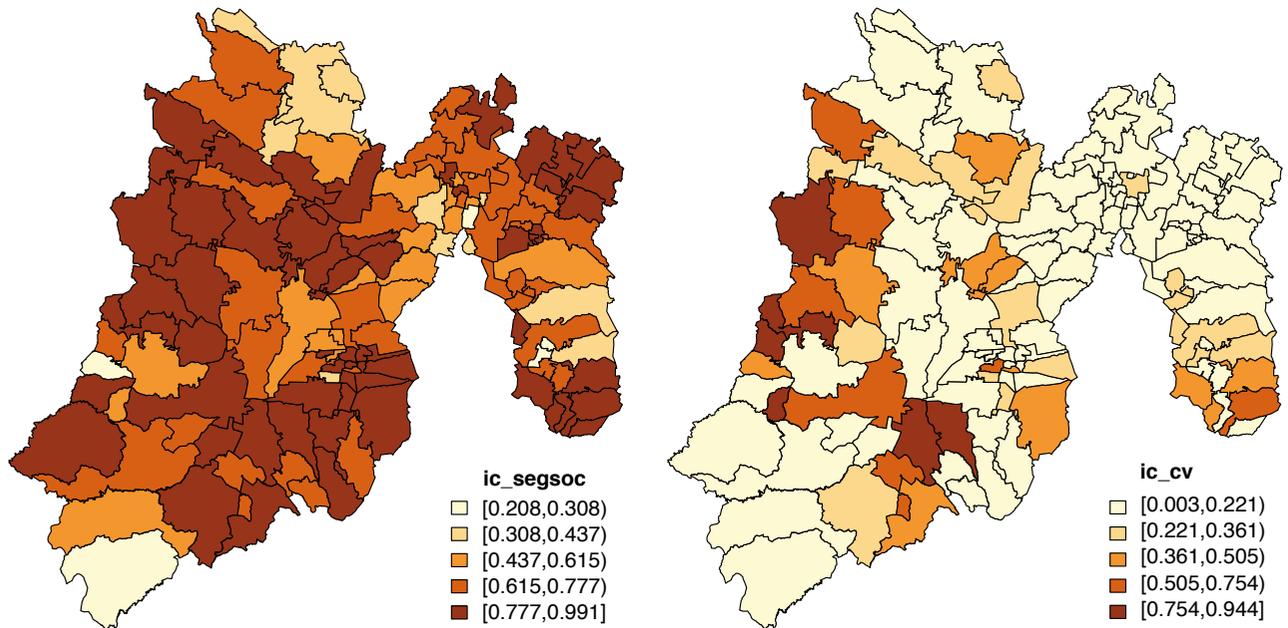


Figura 4

Estimaciones bayesianas del indicador de carencia en servicios básicos de vivienda (izquierda) e indicador de carencia en acceso a alimentación (derecha)

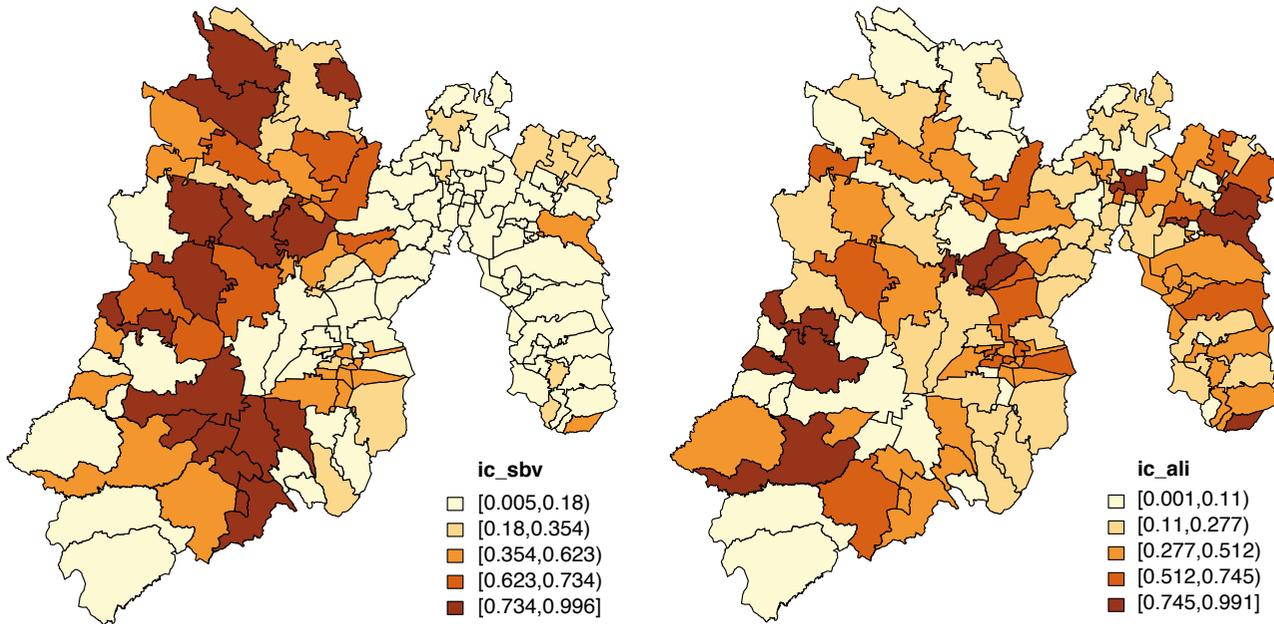
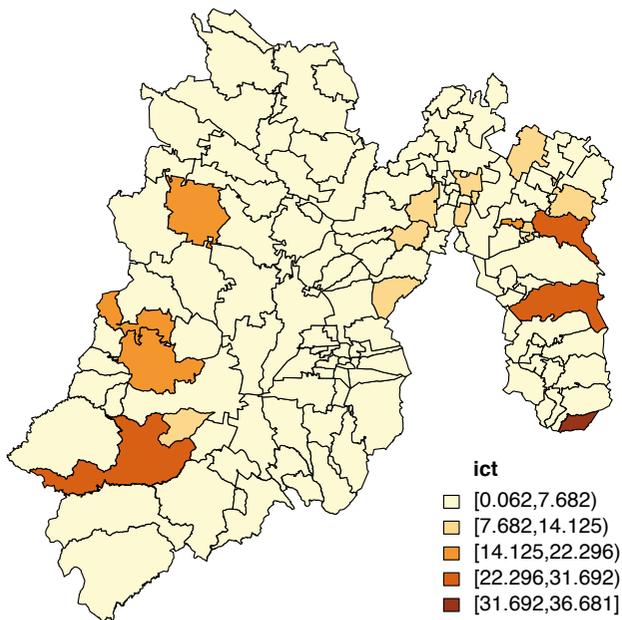


Figura 5

Estimaciones bayesianas del indicador de ingreso corriente total per cápita promedio (en millones)



Conclusiones

Los modelos bayesianos en la estimación de área pequeña se han usado con éxito en diversos países

y en distintas aplicaciones, principalmente económicas. Su ventaja ha sido evidente en la aplicación al tema de pobreza desarrollado en este artículo. El tratamiento multivariado de los indicadores y la inclusión de los efectos espaciales hicieron que se explotara al máximo la información proveniente de las pocas áreas grandes disponibles y producir, así, buenos estimadores para las áreas pequeñas. La complicación incurrida en términos de la implementación del modelo al considerar el caso multivariado no fue mayor debido al uso de distribuciones iniciales (condicionalmente) conjugadas.

De acuerdo con el conocimiento del autor, ésta es la primera aplicación de la metodología estadística en áreas pequeñas al caso multivariado, lo que hace de este trabajo una contribución importante. Los retos en la especificación de las vecindades a través de distancias entre los centros de los municipios y mediante el uso de características de tipo rural/urbano se pueden considerar como otra contribución importante en la implementación. Las rutinas para la implementación del modelo de pobreza se hicieron en el lenguaje R y, junto con los datos utilizados, están disponibles a solicitud del interesado.

Apéndice: muestreador de Gibbs

Sean $(\mathbf{B}^{(0)}, \Sigma^{(0)}, \mathbf{v}^{(0)}, \mathbf{u}^{(0)}, \sigma_v^{2(0)}, \sigma_u^{2(0)})$ valores iniciales de la cadena, posiblemente arbitrarios pero dentro del soporte de las respectivas distribuciones. El muestreador de Gibbs consiste en una serie de simulaciones de manera iterativa. En general, los valores de la cadena para la iteración $(k + 1)$ se obtienen a partir de los valores de la cadena en la iteración (k) . En cada paso se simula de la siguiente distribución condicional utilizando la información de los parámetros simulados más recientes. Las distribuciones condicionales necesarias para implementar el muestreador de Gibbs son:

a. Distribución final condicional para (\mathbf{B}, Σ) ,

$$f(\mathbf{B}, \Sigma | \text{resto}) = N_{m \times p}(\mathbf{B} | \mathbf{B}_1, (\mathbf{D}^{-1} + \mathbf{X}\mathbf{C}\mathbf{X}) \otimes \Sigma) IW(\Sigma | \mathbf{Q}_1, m, v + n)$$

donde:

$$\mathbf{B}_1 = \{(\mathbf{Y} - \mathbf{v}\mathbf{1}' - \mathbf{u}\mathbf{1}')' \mathbf{C}^{-1} \mathbf{X} + \mathbf{B}_0 \mathbf{D}^{-1}\} (\mathbf{D}^{-1} + \mathbf{X}' \mathbf{C}^{-1} \mathbf{X})^{-1},$$

$$\mathbf{Q}_1 = \mathbf{Q}_0 + (\mathbf{Y} - \mathbf{v}\mathbf{1}' - \mathbf{u}\mathbf{1}')' \mathbf{C}^{-1} (\mathbf{Y} - \mathbf{v}\mathbf{1}' - \mathbf{u}\mathbf{1}') + \mathbf{B}_0 \mathbf{C}^{-1} \mathbf{B}_0' - \mathbf{B}_1 \{(\mathbf{Y} - \mathbf{v}\mathbf{1}' - \mathbf{u}\mathbf{1}')' \mathbf{C}^{-1} \mathbf{X} + \mathbf{B}_0 \mathbf{D}^{-1}\}.$$

b. Distribución final condicional para \mathbf{v} ,

$$f(\mathbf{v} | \text{resto}) = N_n(\mathbf{v} | \boldsymbol{\mu}_v, (m \mathbf{C}^{-1} + \mathbf{D}_{\sigma_v^2}^{-1})^{-1})$$

donde:

$$\boldsymbol{\mu}_v = (m \mathbf{C}^{-1} + \mathbf{D}_{\sigma_v^2}^{-1})^{-1} \mathbf{C}^{-1} (\mathbf{Y}\mathbf{1} - \mathbf{X}\mathbf{B}\mathbf{1}' - m\mathbf{u}), \mathbf{y}$$

$$\mathbf{D}_{\sigma_v^2} = \text{diag}(\sigma_v^2, \dots, \sigma_v^2) \text{ de dimensión } n \times n.$$

c. Distribución final condicional para \mathbf{u} ,

$$f(\mathbf{u} | \text{resto}) = N_n(\mathbf{u} | \boldsymbol{\mu}_u, \{m \mathbf{C}^{-1} + \mathbf{D}_{\sigma_u^2}^{-1} (\mathbf{D}_w - \rho \mathbf{W})\}^{-1})$$

donde:

$$\boldsymbol{\mu}_u = \{m \mathbf{C}^{-1} + \mathbf{D}_{\sigma_u^2}^{-1} (\mathbf{D}_w - \rho \mathbf{W})\}^{-1} \mathbf{C}^{-1} (\mathbf{Y}\mathbf{1} - \mathbf{X}\mathbf{B}\mathbf{1}' - m\mathbf{v}),$$

$$\mathbf{y} \mathbf{D}_{\sigma_u^2} = \text{diag}(\sigma_u^2, \dots, \sigma_u^2) \text{ de dimensión } n \times n.$$

d. Distribución final condicional para σ_v^2 ,

$$f(\sigma_v^2 | \text{datos}) = IGa\left(\sigma_v^2 | s_0 + \frac{n}{2}, s_1 + \frac{1}{2} \mathbf{v}' \mathbf{v}\right)$$

e. Distribución final condicional para σ_u^2 ,

$$f(\sigma_u^2 | \text{datos}) = IGa\left(\sigma_u^2 | s_0 + \frac{n}{2}, s_1 + \frac{1}{2} \mathbf{u}' (\mathbf{D}_w - \rho \mathbf{W}) \mathbf{u}\right).$$

Referencias

- Banerjee, S., B. P. Carlin y A. E. Gelfand. *Hierarchical modeling and analysis for spatial data*. Boca Raton, Chapman & Hall, 2004.
- Chen, M.-H., Q.-M. Shao e J. G. Ibrahim. *Monte Carlo methods in Bayesian computation*. New York, Springer, 2000.
- CONEVAL. *Metodología para la medición multidimensional de la pobreza en México*. México, DF, 2009.
- Cochran, W. G. *Sampling techniques*. Chichester, Wiley, 1977.
- DeGroot, M. H. *Optimal statistical decisions*. New Jersey, Wiley, 2004.
- Fay R. E. y R. A. Herriot. "Estimates of income for small places: an application of james-stein procedures to census data", in: *Journal of the American Statistical Association*. 85, 398-409, 1979.
- Ghosh M., K. Natarajan, T. W. F. Stroud y B. P. Carlin. "Generalized linear models for small-area estimation", in: *Journal of the American Statistical Association*. 93, 273-282, 1998.
- Ibrahim, J. G., y P. W. Laud. "A predictive approach to the analysis of designed experiments", in: *Journal of the American Statistical Association*. 89, 309-319, 1994.
- McCullagh, P. y J. A. Nelder. *Generalized linear models*. Boca Raton, Chapman & Hall, 1989.
- Rao, J. N. K. *Small area estimation*. New Jersey, Wiley, 2003.
- R Development Core Team. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria, 2012. ISBN 3-900051-07-0, URL <http://www.r-project.org/>
- Rowe, D. B. *Multivariate Bayesian statistics*. New York, Chapman & Hall, 2003.
- Särndal, C.-E. B. Swensson y J. Wretman. *Model assisted survey sampling*. New York, Springer, 2003.
- West, M. y J. Harrison. *Bayesian forecasting and dynamic models*. New York, Springer, 1997.
- You Y. y J. N. K. Rao. "Small area estimation using unmatched sampling and linking models", in: *Canadian Journal of Statistics*. 30, 3-15, 2002.