

**UNIVERSIDAD SAN FRANCISCO DE QUITO USFQ**

**Colegio de Posgrados**

**Inferencia Ecológica sobre las Elecciones Ecuatorianas  
en Febrero del 2013**

**David Alfredo Vivas Paspuel  
Atal Kumar Vivas Paspuel**

**Carlos Jiménez Mosquera, Ph.D.  
Director de Trabajo de Titulación**

Trabajo de titulación de posgrado presentado como requisito  
para la obtención del título de Magister en Matemáticas Aplicadas

Quito, 29 de mayo de 2017

UNIVERSIDAD SAN FRANCISCO DE QUITO USFQ  
COLEGIO DE POSGRADOS

HOJA DE APROBACIÓN DE TRABAJO DE  
TITULACIÓN

Inferencia Ecológica sobre las Elecciones de Ecuador en Febrero  
del 2013

David Alfredo Vivas Paspuel  
Atal Kumar Vivas Paspuel

firmas

Carlos Jiménez Mosquera, Ph.D. ....  
Director del Trabajo de Titulación

Julio Ibarra Fiallo M.Sc. ....  
Miembro del Comité de Trabajo de Titulación

Carlos Jiménez Mosquera, Ph.D. ....  
Miembro del Comité de Trabajo de Titulación  
Director de la Maestría en Matemáticas Aplicadas

César Zambrano, Ph.D. ....  
Decano del Colegio de Ciencias e Ingeniería

Hugo Burgos, Ph.D. ....  
Decano del Colegio de Posgrados

Quito, 29 de Mayo del 2017

© **Derechos de Autor**

Por medio del presente documento certifico que he leído todas las Políticas y Manuales de la Universidad San Francisco de Quito USFQ, incluyendo la Política de Propiedad Intelectual USFQ, y estoy de acuerdo con su contenido, por lo que los derechos de propiedad intelectual del presente trabajo quedan sujetos a lo dispuesto en esas Políticas.

Asimismo, autorizo a la USFQ para que realice la digitalización y publicación de este trabajo en el repositorio virtual, de conformidad a lo dispuesto en el Art. 144 de la Ley Orgánica de Educación Superior.

Firma: .....

Nombre: David Alfredo Vivas Paspuel

Código de estudiante: 00121318

C.I.: 1716664121

Firma: .....

Nombre: Atal Kumar Vivas Paspuel

Código de estudiante: 00121255

C.I.: 1715643977

Fecha: Quito, 29 de Mayo de 2017

## DEDICATORIA

*A Dios y a nuestros Padres*

*A la familia y amigos*

## AGRADECIMIENTOS

*Queremos agradecer a la Universidad San Francisco de Quito por darnos la oportunidad de dar un paso en nuestra carrera profesional por medio de éste Programa de Maestría, a su Director PhD. Carlos Jiménez y a la vez nuestro Director de Tesis por impartir su profundo conocimiento en sus clases y dirigir éste trabajo, a Michel Vargas y su esposa Vanessa Carrión por facilitarnos toda la información necesaria para realizar este trabajo, a Andrea Landázuri por su apoyo con el servidor del GICAS (Grupo de Investigación de Ciencias Aplicadas y Simulación de la USFQ), a Andrea Carrera por el apoyo y la comprensión durante esta etapa, a todos nuestros profesores y amigos: Andrea Moreira, Julio Ibarra, David Hervas, Thomas Reytier, Luca Guzzardi, John Skukalek, Carlos Suárez, Oihane Fernández, Eduardo Alba, a nuestros compañeros de clase con quienes compartimos muchas horas de estudio y por los consejos sobre nuestro trabajo: Gaby, Mayra, Juanito, Bruno, Juan Pablo y Felipe, a nuestros padres Rita y Miguel que siempre nos apoyaron con amor en cualquier proyecto que realizamos, a nuestro amigo Sivananda quien nos dio la claridad en nuestras ideas y el empuje para terminar ésta etapa.*

## Resumen

En la contienda electoral ecuatoriana de febrero 2013 el partido Alianza País gana la Presidencia de la República en primera vuelta con más del 50% de apoyo de los votantes, consolidándose como la primera fuerza política del país y causando revuelo a nivel internacional ya que por tercera vez consecutiva Rafael Correa obtiene dicho cargo. Los resultados oficiales nos muestran la información de apoyo agregada en cada recinto sin distinción de comportamiento individual, pero si dividimos el electorado en clases (sexo, edad, condición económica, etc) nos podemos preguntar: que cantidad de apoyo obtuvo Alianza País por cada clase? Para responder ésta cuestión podemos hacer uso de la inferencia ecológica que nos permite encontrar estadísticamente las proporciones de apoyo de cada clase a cierto partido político en particular. En éste trabajo usamos la inferencia ecológica para tablas de contingencia  $2 \times 2$  y  $R \times C$  por medio de modelos jerárquicos bayesianos Binomial-Beta y Multinomial-Dirichlet que se apoyan en técnicas computacionales MCMC (Markov Chain Monte Carlo) las cuales infieren las proporciones buscadas.

Palabras clave: Inferencia ecológica, modelos jerárquicos, estadística bayesiana, métodos MCMC, Binomial-beta, Multinomial-Dirichlet.

## Abstract

In the Ecuadorian electoral contest of February 2013, the Alianza País party won the presidency of the republic in the first round with more than 50% voter support, consolidating itself as the country's first political force and causing international upheaval because third consecutive time Rafael Correa obtains this position. The official results show us the aggregated support information in each room without distinction of individual behavior, but if we divide the electorate into classes (sex, age, economic status, etc.) we can ask what amount of support did Alianza País get for each class? To answer this question we can make use of the ecological inference that allows us to find statistically the proportions of support of each class to certain political party in particular. In this paper we use the ecological inference for 2x2 and RxC contingency tables by Bayesian hierarchical models Binomial-Beta and Multinomial-Dirichlet that rely on MCMC (Markov Chain Monte Carlo) computational techniques which infer the desired proportions.

Keywords: Ecological inference, hierarchical models, Bayesian statistics, MCMC methods, Binomial-beta, Multinomial-Dirichlet.

# TABLA DE CONTENIDO

<b>Resumen</b>	<b>6</b>
<b>Abstract</b>	<b>7</b>
<b>1. Introducción</b>	<b>10</b>
<b>2. Preliminares</b>	<b>11</b>
2.1. La Inferencia Ecológica . . . . .	11
2.1.1. Inferencia Ecológica: Tablas 2x2. . . . .	13
2.1.2. Inferencia Ecológica: tablas RxC. . . . .	19
2.2. Inferencia Bayesiana . . . . .	20
2.2.1. Modelos Jerárquicos Bayesianos . . . . .	22
2.3. Computación Bayesiana y el lenguaje de programación Stan . . . . .	22
2.3.1. Métodos Monte Carlo . . . . .	23
2.3.2. Métodos de muestreo Markov Chain Monte Carlo (MCMC) . . . . .	24
2.3.3. Muestreo Hamiltoniano Monte Carlo (HMC) . . . . .	28
2.4. Método con Regresión Multinomial Penalizada. . . . .	29
<b>3. Datos y métodos aplicados</b>	<b>32</b>
3.1. Datos . . . . .	32
3.2. Metodología . . . . .	33
3.2.1. El modelo Bayesiano . . . . .	33
3.2.2. Algoritmo HMC . . . . .	34
3.2.3. Modelo Jerárquico Bayesiano para tablas 2x2 sin covariable . . . . .	34
3.2.4. Modelo Jerárquico Bayesiano para tablas 2x2 con el uso de covariables . . . . .	36
3.2.5. Modelo Jerárquico Bayesiano para tablas de tamaño RxC . . . . .	38
3.3. Resultados . . . . .	40
3.3.1. Resultados para tablas de contingencia 2x2 sin covariable . . . . .	40
3.3.2. Resultados para tablas de contingencia 2x2 con covariable . . . . .	43
3.3.3. Resultados para tablas de contingencia de tamaño RxC . . . . .	48
3.4. Evaluación de las simulación . . . . .	50
3.5. Resultados del segundo método . . . . .	55
<b>4. Conclusiones</b>	<b>57</b>



	9
<b>Apéndice 1</b>	<b>60</b>
<b>Apéndice 2</b>	<b>63</b>
<b>Apéndice 3</b>	<b>72</b>
<b>Apéndice 4</b>	<b>74</b>
<b>REFERENCIAS</b>	<b>75</b>

## 1. Introducción

La contienda electoral ecuatoriana que se produjo en febrero de 2013 ratificó la popularidad y el apoyo que los diferentes sectores sociales dieron a Rafael Correa. Alianza País (creado el 19 de Febrero del 2006) presentó su binomio Correa - Glass a los comicios teniendo como punto de mira la consolidación de un proyecto político que venía gestándose desde la redacción de la nueva Constitución. Antes de las elecciones generales de ese año, el Presidente Correa logró obtener la Presidencia en dos ocasiones consecutivas (2006 y 2009)[1]. Es innegable el hecho que desde la primera victoria electoral de Rafael Correa en 2006, Ecuador ha experimentado numerosas transformaciones sociales, económicas, educativas y culturales [2]. Además, ésta década de *Revolución Ciudadana* tuvo el apoyo de un gran número de simpatizantes; así como también, la gran oposición de otros.

En los procesos electorales suelen presentarse resultados previos a las elecciones a través del análisis de *intención de voto* que suele ser la respuesta a la pregunta: "Suponiendo que mañana se llevan a cabo las elecciones...", también se tienen los análisis de *estimación de voto* que se basan en la información anterior a la cual se le añaden otras variables cualitativas de la encuesta, además luego de los comicios se tienen los llamados *exit polls* que son realizadas a los electores posteriormente a que hayan sufragado y al final, se tienen resultados oficiales otorgados por el organismo correspondiente[4]. Todos éstos resultados suelen presentar el número de votos o intención de voto hacia cierto partido político a través de tablas o gráficos porcentuales.

En ocasiones se desea realizar un análisis mas profundo de los resultados, como por ejemplo, dividir la población en clases como: edad, socioeconómicas, nivel de educación, sexo, etc. y medir el apoyo que obtuvo cierto partido político por cada una de éstas. Dado que el voto es secreto, en principio, no se conoce el apoyo por clases, sin embargo, éste análisis puede ser realizado con ayuda de la inferencia ecológica, la cual, a través de modelos estadísticos, nos permite inferir las proporciones de apoyo de cierta clase hacia cierto partido político.

En 2013 Alianza País ganó la presidencia en primera vuelta con más del 50% de apoyo de los electores [3]. Los resultados oficiales, que se pueden ver en la figura 1, fueron presentados a través del CNE (Consejo Nacional Electoral). El presente trabajo consiste en inferir las proporciones de apoyo de las diferentes clases a Alianza País en

las elecciones presidenciales del año 2013, mediante la aplicación de modelos estadísticos bayesianos, junto con técnicas computacionales como la simulación usando el algoritmo Markov Chain Monte Carlo (MCMC) en el cual se basa el lenguaje de programación para inferencia estadística bayesiana Stan.

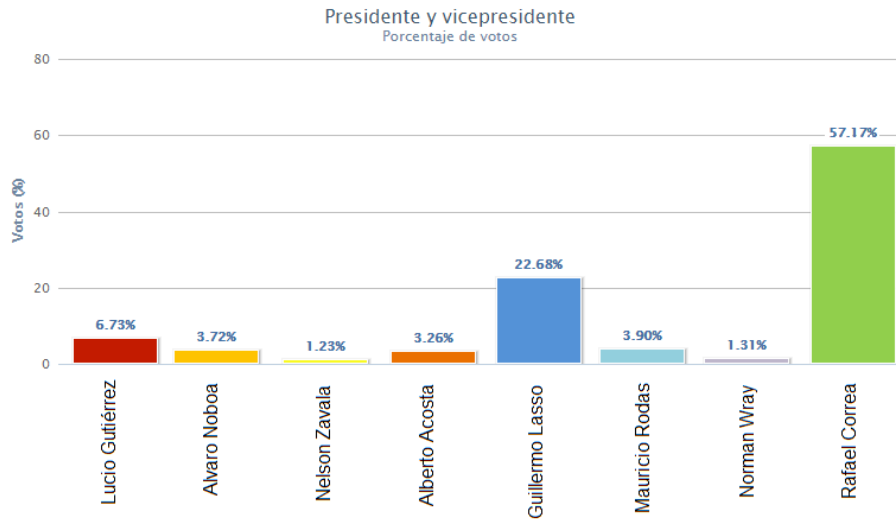


Figura 1: Resultados electorales presidenciales en Ecuador 2013. Fuente: CNE

## 2. Preliminares

### 2.1. La Inferencia Ecológica

Inferencia ecológica es el proceso de inferir el comportamiento desagregado (individual) a partir de datos agregados [5]. El nombre de inferencia ecológica se refiere a la idea de correlaciones ecológicas, es decir, las correlaciones entre las variables observadas para un grupo de individuos, en lugar de las correlaciones individuales, donde justamente los individuos son la unidad de análisis [32].

El objetivo de la inferencia ecológica es hacer predicciones a nivel desagregado a partir de datos agregados. Una manera de entender éste problema es considerar una tabla de contingencia, cuyas entradas dentro de la tabla sean desconocidas y sus marginales conocidas; el objetivo es entonces, estimar las intersecciones que dan origen a los totales marginales por filas y columnas en dicha tabla de contingencia. Éste método estadístico ha sido aplicado en varias áreas como: políticas públicas, epidemiología, economía,

ciencias de la administración y marketing, historia, sociología, etc.

Dado que en una tabla de contingencia trabajamos con dos o más variables, éstas estarán de algún modo correlacionadas entre sí. Por ejemplo, se podría estudiar la correlación entre la calidad de alimentación y la estatura entre los niños de un nivel escolar dado. Un estudio a nivel individual podría hacer uso de la información que proveen 100 niños, luego medir la calidad de alimentación y su estatura; la correlación entre las dos variables sería a nivel individual. Por el contrario, otro estudio podría hacer uso de 50 grupos de estudiantes del mismo nivel escolar, a continuación, medir la media de la calidad de alimentación y la estatura media de cada uno de los 50 grupos. Una correlación entre éstos medios grupales sería un ejemplo de una correlación ecológica.

Debido a que una correlación describe el grado de relación lineal entre variables, las correlaciones a nivel de grupo pueden ser mucho más altas que las de nivel individual. Pensar que ambos son iguales es un ejemplo de *falacia ecológica*.

La falacia ecológica[31] puede referirse a la siguiente falacia estadística: la correlación entre las variables individuales se deduce de la correlación de las variables recogidas para el grupo al que pertenecen esos individuos. Por ejemplo, si la puntuación media de un grupo de valores es mayor que cero, no significa que un miembro aleatorio de ése grupo tenga más probabilidad de tener una puntuación positiva que negativa (a menos que éste grupo de valores tenga más puntuaciones negativas que puntuaciones positivas). Matemáticamente, esto viene del hecho de que una distribución puede tener una media positiva pero una mediana negativa. Esta propiedad está vinculada a la asimetría de la distribución de los datos.

Otro limitante para la inferencia ecológica es el *sesgo de agregación* que podría definirse como la ocurrencia de un error que no aparece como un hecho provocado por el azar, sino que ocurre en forma sistemática, ya que una asociación observada entre variables en un nivel agregado no necesariamente representa la asociación que existe en el ámbito desagregado.

Gran parte de la ocurrencia de la falacia ecológica se podría deber por ejemplo, a la imposibilidad de controlar eventuales variables de confusión en el análisis, puesto que el estudio podría no considerar la medición de covariables. Como ejemplo, en un estudio ecológico sobre los efectos del hábito alcohólico en el desempeño laboral, algunas variables de confusión (capacitación laboral y estrés) podrían producir un sesgo de agregación al

no ser considerada su medición.

Dadas las limitaciones para la inferencia ecológica, ésta en general no puede ser usada para hacer inferencias concernientes al comportamiento desagregado. Sin embargo, los investigadores siguen utilizando correlaciones ecológicas para realizar inferencias, lo importante es analizar bajo que circunstancias o condiciones éstas inferencias son justificadas[10]. Como veremos en el desarrollo de los modelos, se deben cumplir ciertas condiciones para que éstos funcionen adecuadamente.

Desde la década de 1950 se han dado varios enfoques para la solución de la inferencia ecológica como el método de los intervalos, el método de la regresión ecológica y de forma más reciente, Gary King (1997) combinó el método de los intervalos con procedimientos de simulación que es conocida como EI (acrónimo de Inferencia Ecológica)[6]. La velocidad del desarrollo de la investigación estadística sobre la inferencia ecológica ha sido paralela al progreso de las aplicaciones, y en los últimos años se ha visto numerosos modelos nuevos, métodos innovadores y nuevos esquemas de cálculo [7]. El modelo que utiliza éste trabajo es un modelo jerárquico bayesiano tanto para el caso de tablas 2x2, así como para el caso de tablas RxC. Sin embargo, es importante tener una reseña de los modelos anteriores ya que éstos contienen aspectos importantes de los cuales se nutren los nuevos modelos, esa es la razón justamente de revisarlos.

### 2.1.1. Inferencia Ecológica: Tablas 2x2.

El caso más simple para el análisis de inferencia se presenta cuando se tiene una tabla de contingencia 2x2 como se muestra en el cuadro 1, para esto se ha dividido al electorado en dos clases: Hombres-Mujeres y el Apoyo o No Apoyo a cierto partido político. El análisis de los modelos se realizará con los datos obtenidos en las elecciones del 2013.

Para el caso 2x2 se ha utilizado la terminología usada por King en su modelo básico, es decir se ha usado el superíndice  $b$  para representar a los hombres y el superíndice  $w$  para representar a las mujeres.

La información que proveen los censos nacionales es la cantidad de empadronados por recinto electoral y los resultados oficiales de las elecciones están dadas por el organismo correspondiente para cada recinto  $i$  de un total de  $p$  recintos, cada recinto  $i$  es una unidad poblacional geográfica (estado, provincia, parroquia, etc) tomada por el investigador.

Cada recinto es una unidad agregada y cada variable individual que yace dentro de éstas unidades es dicotómica. En el cuadro 1 se observa para cada recinto  $i$  ( $i = 1, \dots, p$ ), la fracción de personas que votaron por AP ( $T_i$ ) y la fracción de hombres que acudieron a las urnas ( $X_i$ ) de un total de  $N_i$  individuos. Las cantidades que quedan sin observar (dado que el voto es secreto) son: la proporción de hombres que votaron por AP ( $\beta_i^b$ ) y la proporción de mujeres que votaron por AP ( $\beta_i^w$ ). Se tienen entonces dos variables observadas:  $X_i$  y  $T_i$  y dos cantidades de interés no observadas:  $\beta_i^b$  y  $\beta_i^w$  para cada una de las  $p$  observaciones.

	Apoya a AP	No Apoya a AP	Total
Hombres (b)	$\beta_i^b$	$1 - \beta_i^b$	$X_i$
Mujeres (w)	$\beta_i^w$	$1 - \beta_i^w$	$1 - X_i$
Total	$T_i$	$1 - T_i$	1

Cuadro 1: Tabla de contingencia para el análisis de inferencia ecológica y su notación: Caso 2x2.

Cada término se define a continuación:

- $\beta_i^b$ : proporción de hombres inscritos en los padrones electorales de la región  $i$ , que votaron por AP.
- $\beta_i^w$ : proporción de mujeres inscritas en los padrones electorales de la región  $i$ , que votaron por AP.
- $1 - \beta_i^b$ : proporción de hombres inscritos en los padrones electorales de la región  $i$ , que no votaron por AP.
- $1 - \beta_i^w$ : proporción de mujeres inscritas en los padrones electorales de la región  $i$ , que no votaron por AP.
- $X_i$ : proporción de hombres inscritos en los padrones electorales de la región  $i$ .
- $1 - X_i$ : proporción de mujeres inscritas en los padrones electorales de la región  $i$ .
- $T_i$ : proporción de individuos inscritos en los padrones electorales de la región  $i$  que votaron por AP.
- $1 - T_i$ : proporción de individuos inscritos en los padrones electorales de la región  $i$  que no votaron por AP.

El objetivo de algunos métodos es solo obtener la fracción promedio ponderada de las correspondientes cantidades de interés a nivel de recinto[7], es decir:

$$B^b = \frac{\sum_{i=1}^p N_i X_i \beta_i^b}{\sum_{i=1}^p N_i X_i} \quad (1)$$

así como también:

$$B^w = \frac{\sum_{i=1}^p N_i (1 - X_i) \beta_i^w}{\sum_{i=1}^p N_i (1 - X_i)} \quad (2)$$

Se hará un análisis breve de los enfoques más importantes que se han dado desde el siglo pasado al problema de la inferencia ecológica. El modelo que utiliza éste trabajo es un modelo jerárquico bayesiano tanto para el caso 2x2 como para el caso RxC, sin embargo es importante tener una reseña de los modelos anteriores ya que éstos contienen aspectos importantes de los cuales se nutren los nuevos modelos, esa es la razón justamente de revisar los enfoques anteriores. Se hará un recorrido breve por los métodos de Duncan-Davis, Goodman y King para contar con ideas básicas acerca del problema y su solución.

Antes de King (1997), se habían desarrollado dos enfoques importantes para la solución de la inferencia ecológica: el método determinístico de los límites propuesto por Duncan y Davis (1953) y el modelo estadístico de Goodman (1959) [7].

El objetivo del método de los límites es extraer la información determinística acerca de las cantidades de interés que yace en las observaciones. Por ejemplo, si se conoce que en cierto recinto, la cantidad de personas que apoyaron a AP es de 150 y se sabe también que en dicho recinto acudieron a las urnas 60 mujeres, entonces con 100% de certeza se puede decir que la cantidad de hombres que apoyaron a AP es un número entre 0 y 90, es decir,  $[0,90]$  es un intervalo de confianza al 100%. Sin embargo, esta técnica muchas veces no ha permitido establecer rangos acotados y, por lo tanto, útiles para las investigaciones [8].

Está dicho que los parámetros de interés  $\beta_i^b$  y  $\beta_i^w$  están acotados en el intervalo unitario y una vez que observa  $X_i$  y  $T_i$ , éstos estarán acotados de forma más estrecha [9] mediante:

$$\beta_i^b \in \left[ \max \left( 0, \frac{T_i - (1 - X_i)}{X_i} \right), \min \left( \frac{T_i}{X_i}, 1 \right) \right] \quad (3)$$

y

$$\beta_i^w \in \left[ \max \left( 0, \frac{T_i - X_i}{1 - X_i} \right), \min \left( \frac{T_i}{1 - X_i}, 1 \right) \right] \quad (4)$$

Éstas son las expresiones determinísticas para los límites en los cuales se encontrarán los parámetros de interés.

Goodman por su parte miró las observaciones desde un enfoque estadístico, observando la variación de  $X_i$  y  $T_i$  por medio de una regresión [10] que fué denominada *Regresión Ecológica*. Éste enfoque provee en general estimadores más acotados por lo que ha tenido mayor difusión. La esencia de éste enfoque plantea que realizar una regresión entre las variables ecológicas agregadas puede ser útil para hacer inferencias sobre las cantidades de interés[8].

Para cada distrito se tiene:

$$T_i = X_i \beta_i^b + (1 - X_i) \beta_i^w \quad (5)$$

Ésta identidad contiene tanto las variables de interés ( $\beta_i^b$  y  $\beta_i^w$ ) así como las observaciones por distrito ( $X_i$  y  $T_i$ ). Para inferir los  $\beta$ 's, Goodman propone trabajar con una regresión lineal a partir de la información dada, bajo el supuesto de que éstas proporciones se mantienen constantes en todos los recintos, lo que implica decir que los parámetros de interés no están correlacionados con  $X_i$ , o sea  $\text{Cov}(\beta_i^b, X_i) = \text{Cov}(\beta_i^w, X_i) = 0$ .

Si el supuesto se cumple, la estimación por mínimos cuadrados de  $T_i$  dado  $X_i$  provee un estimador insesgado de  $\beta_i^b$  y  $\beta_i^w$ . La regresión de Goodman ha sido criticada por autores como Boudon[11], dado el hecho de que es difícil sostener que las proporciones de interés se mantengan constantes en todos los recintos. Como Goodman reconoció, cuando ésta suposición clave no se cumple, las estimaciones del modelo serán sesgadas. De hecho, pueden ser muy sesgadas, fuera de los límites deterministas e incluso fuera del intervalo unitario. Sin embargo, lo que Goodman logró fue formalizar la lógica del enfoque en un modelo de regresión simple y dar las condiciones bajo las cuales las estimaciones de dicho modelo funcionarían como una solución al problema[7].

Fue a partir de la enumeración de estas dificultades que Gary King desarrolló una nueva propuesta, la misma que utiliza los dos enfoques anteriores: la información determinística contenida en los límites, y un modelo estadístico que permite realizar una estimación al interior de dichos límites. Para ello, se mueve sucesivamente del nivel de



recintos al nivel regional y viceversa. Primero, al tratar cada recinto por separado, el método utiliza toda la información disponible para dar un rango de posibles valores para estas cantidades de interés a ese nivel, luego, a fin de aproximarse a una respuesta más acotada, el modelo estadístico "toma fuerza" de todos los datos de los otros recintos para dar la probable ubicación de cada verdadera cantidad de interés dentro de los conocidos límites determinísticos [8].

La figura 2 muestra éste enfoque donde se puede observar la información que proporcionan los recintos. El primero es un diagrama de dispersión en el que se muestra  $T_i$  contra  $X_i$  (donde cada punto es un recinto), es decir la proporción del electorado que acudió a las urnas para cada recinto  $i$  contra la proporción del electorado que apoyó a AP en ése recinto. En cambio, en el segundo gráfico, cada punto es ahora una línea sobre la cual se ubicarán las cantidades de interés reales ( $\beta_i^b$  y  $\beta_i^w$ ).

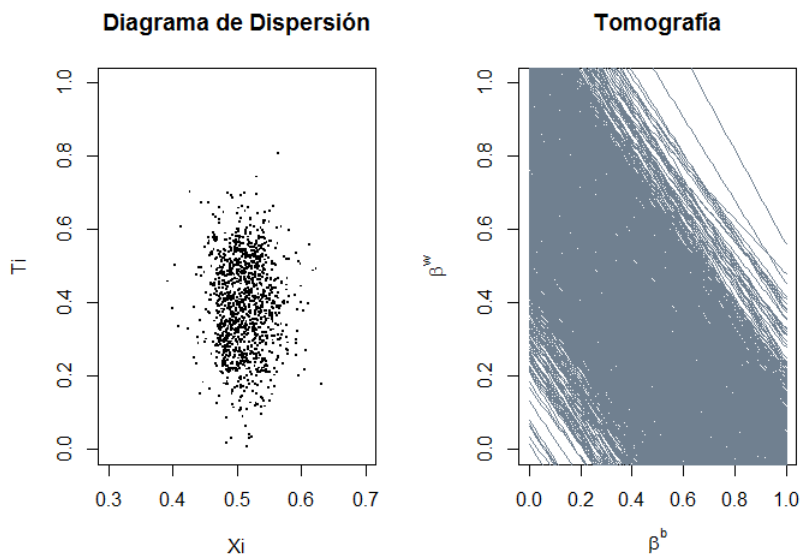


Figura 2: El gráfico de la izquierda muestra el diagrama de dispersión de  $X_i$  vs  $T_i$  y el gráfico de la derecha muestra la misma información como una tomografía de las cantidades de interés  $\beta_i^b$  y  $\beta_i^w$ . Los datos utilizados provienen de las elecciones presidenciales del año 2013. beta b corresponde a la proporción de hombres y beta w a la proporción de mujeres que apoyaron a AP.

Al graficar las líneas determinísticas indicadas se puede ver el efecto de la pérdida de información debida a la agregación. La inferencia ecológica consiste en localizar justamente los parámetros de interés para cada recinto que se encuentran en éste cuadrado

unitario, ya que valores fuera de éste, serían lógicamente imposibles [7].

Para realizar éstas gráficas, King [5] tomó la identidad de Goodman (ecuación 5) y encontró que si resuelve para uno de los parámetros desconocidos, ambos están linealmente relacionados mediante:

$$\beta_i^w = \left( \frac{T_i}{1 - X_i} \right) - \left( \frac{X_i}{1 - X_i} \right) \beta_i^b \quad (6)$$

En ésta ecuación la pendiente y el intercepto son conocidos es decir, cada punto en el gráfico izquierdo de la figura 2, puede ser expresado, sin suposiciones o pérdida de información, como el gráfico de la derecha, lo que King llamó una línea de *tomografía* dentro del cuadrado unitario.

Debido a la agregación de los datos se pierde información, lo que obliga a trazar una línea completa (sobre la cual, los valores  $\beta_i^b$  y  $\beta_i^w$  deben ubicarse). Sin embargo, el hecho de tener la certeza de que para cada punto en la figura 2a existe una recta en la figura 2b que contiene las verdaderas cantidades de interés ya es un gran paso hacia el objetivo y más aún, King observó que para inferir las cantidades de interés en un recinto  $i$  se puede "tomar fuerza" de los demás recintos, es decir existe información adicional en las demás observaciones del mismo conjunto de datos ( $X_j$  y  $T_j$  para  $i \neq j$ ) que bajo los supuestos correctos pueden ser usados para aprender sobre  $\beta_i^b$  y  $\beta_i^w$  para cada recinto [7].

El modelo de King requiere de tres supuestos para su funcionamiento, en primer lugar, el conjunto de puntos ( $\beta_i^b$ ,  $\beta_i^w$ ) deben caer en un solo grupo dentro de la unidad cuadrada el cuál puede ser ampliamente o estrechamente disperso o muy variable en uno de los parámetros y estrecho en el otro; y las dos incógnitas pueden estar positivamente, negativamente, o no correlacionadas en absoluto sobre  $i$ . La expresión matemática específica de esta suposición (de que se tiene un solo grupo de puntos) es que  $\beta_i^b$  y  $\beta_i^w$  siguen una densidad conjunta normal bivariada truncada:

$$TN(\beta_i^b, \beta_i^w | \check{B}, \check{\Sigma}) = N(\beta_i^b, \beta_i^w | \check{B}, \check{\Sigma}) \frac{1(\beta_i^b, \beta_i^w)}{R(\check{B}, \check{\Sigma})} \quad (7)$$

donde su núcleo es la distribución bivalente normal no truncada:

$$N(\beta_i^b, \beta_i^w | \check{B}, \check{\Sigma}) = (2\pi)^{-1} |\check{\Sigma}|^{-1/2} \exp \left[ -\frac{1}{2} (\beta_i - \check{B})' \check{\Sigma}^{-1} (\beta_i - \check{B}) \right] \quad (8)$$

La función indicadora:  $1(\beta_i^b, \beta_i^w)$  es uno si  $\beta_i^b$  y  $\beta_i^w$  se encuentran en el intervalo [0,1]

y cero en otro caso.  $R(\check{B}, \check{\Sigma})$  es un factor de normalización y equivale al volumen debajo de la distribución no truncada y encima del cuadrado unitario.

La segunda suposición es la ausencia de autocorrelación espacial [5] y es necesaria para construir la función de máxima verosimilitud y establece que condicionando en  $X_i$ ,  $T_i$  y  $T_j$  son independientes en la media (la cual es una condición mas débil que estocásticamente independientes [12]). Las violaciones de esta suposición de maneras empíricamente razonables (e incluso algunas no razonables) no parecen inducir mucho sesgo [7].

La tercera suposición y a la vez la más crítica establece que  $X_i$  es independiente de  $\beta_i^b$  y  $\beta_i^w$ . Esta hipótesis implica que las cantidades de interés varían aleatoriamente en torno a medias constantes y por tanto son independientes de  $X_i$  [13]. Esta hipótesis es equivalente a asumir la inexistencia de sesgo en la agregación, lo que es necesario para obtener estimaciones consistentes para los parámetros de la distribución, aunque esta suposición no siempre sea necesaria para obtener estimaciones precisas para las cantidades de interés.

### 2.1.2. Inferencia Ecológica: tablas RxC.

En muchos casos como los de contiendas electorales y en variadas aplicaciones es necesario trabajar con extensiones de la tabla 2x2, éstas se encuentran en la literatura como tablas RxC ( $R \geq 2$ ,  $C > 2$ ). Varios autores como Grofman (1985), Achen and Shively (1997), King (1997), Rosen (2001), Benoit (2004) han trabajado en la inferencia ecológica de éstas tablas. Si bien hoy en día se han evaluado empíricamente métodos para tablas 2x2 (por ejemplo, Wakefield 2004, Hudson 2010), casi ninguna evaluación empírica ha caracterizado métodos que enfrentan la cuestión de la estimación de tablas RxC. Esto es sorprendente, dado que el mundo real normalmente tiende a presentar situaciones en las que los datos necesitan ser agrupados en tablas con más de dos columnas y filas [14].

En éste trabajo se ha realizado un análisis para tablas RxC usando la información de la misma contienda electoral (febrero 2013). Se tomó en cuenta el apoyo de diferentes grupos de votantes agregados por edad hacia los dos partidos mas importantes en dichas elecciones: Alianza País y CREO, quedando el resto de votos hacia otros partidos junto con votos blancos y nulos. En forma general la extensión a tablas mas largas RxC (Cuadro 2) queda configurada de la siguiente manera:

	Apoyo a AP	Apoya a CREO	OTROS	Total
16 a 29	$\beta_{11}^i$	$\beta_{12}^i$	$1 - \sum_{c=1}^2 \beta_{1c}^i$	$X_{1i}$
30 a 44	$\beta_{21}^i$	$\beta_{22}^i$	$1 - \sum_{c=1}^2 \beta_{2c}^i$	$X_{2i}$
45 a 60	$\beta_{31}^i$	$\beta_{32}^i$	$1 - \sum_{c=1}^2 \beta_{3c}^i$	$X_{3i}$
más de 60	$\beta_{41}^i$	$\beta_{42}^i$	$1 - \sum_{c=1}^2 \beta_{4c}^i$	$1 - \sum_{r=1}^3 X_{ri}$
Total	$T_{1i}$	$T_{2i}$	$1 - \sum_{c=1}^2 T_{ci}$	

Cuadro 2: Tabla de contingencia para el análisis de inferencia ecológica y su notación: Caso RxC.

Para un recinto dado  $i$  ( $i = 1, \dots, p$ ), se tiene como información observada las fracciones del electorado que apoyan a cierto partido político ( $T_{1i}, \dots, T_{C,i}$ ) y las fracciones del electorado que pertenecen a cierta clase ( $X_{1i}, \dots, X_{R,i}$ ). Mientras que  $(\beta_{rc}^i, r = 1, \dots, R, c = 1, \dots, C - 1)$  indican las cantidades de interés.

Goodman demostró que la identidad mostrada en la ecuación 5 se mantiene exactamente para tablas más grandes. Nuevamente, la suposición más importante, necesaria y problemática a la vez y que evita resultados sesgados es que los parámetros de interés y los  $X_r^i$  no estén correlacionados. Cuando esta suposición no se cumple, las estimaciones serán sesgadas, e incluso fuera de los límites deterministas [14]. Así como en el caso 2x2 ésto se traduce en una suposición de que, a nivel de recintos electorales, la tendencia a emitir un voto por un partido (el *beta* para una clase dada) no debería correlacionarse con el tamaño de la clase en el recinto. Afortunadamente esta correlación se puede observar en un diagrama de dispersión de  $X_i$  vs  $T_i$  como se verá mas adelante.

## 2.2. Inferencia Bayesiana

La inferencia bayesiana es el proceso de ajustar un modelo de probabilidad a un conjunto de datos y resumir el resultado mediante una distribución de probabilidad sobre los parámetros del modelo y sobre cantidades no observadas tales como predicciones para nuevas observaciones [15].

Supongamos que se desea conocer los valores de  $k$  cantidades desconocidas  $\Theta = (\theta_1, \theta_2, \dots, \theta_k)$  con  $k \geq 1$  y se tiene conocimiento a priori de sus valores, lo cual se podría expresar en términos de una pdf (función de densidad de probabilidad)  $p(\Theta)$ .

Además suponiendo que se tiene  $n$  observaciones acerca de las cantidades desconocidas:  $\mathbf{X} = (X_1, X_2, \dots, X_n)$  las cuales tienen una distribución de probabilidad cuyos parámetros son éstas  $k$  cantidades desconocidas entonces dado que el vector  $\mathbf{X}$  es una variable

aleatoria, se puede escribir su dependencia en términos de una pdf:  $p(\mathbf{X}|\Theta)$ .

Ahora, para expresar la creencia acerca de  $\Theta$  teniendo en cuenta tanto el conocimiento a priori y las observaciones, la herramienta básica necesaria es el teorema de Bayes para variables aleatorias (generalizadas para tratar con vectores aleatorios).

Partiendo del concepto de probabilidad condicional, el teorema de Bayes en forma general se desarrolla de la siguiente manera:

$$p(\Theta|\mathbf{X}) = \frac{p(\Theta, \mathbf{X})}{p(\mathbf{X})}$$

pasando de probabilidad conjunta a marginal:

$$= \frac{p(\mathbf{X}|\Theta)p(\Theta)}{p(\mathbf{X})}$$

Usando la ley de probabilidad total:

$$= \frac{p(\mathbf{X}|\Theta)p(\Theta)}{\int_{\Theta} p(\mathbf{X}, \Theta)d\Theta}$$

pasando de probabilidad conjunta marginal en el denominador:

$$= \frac{p(\mathbf{X}|\Theta)p(\Theta)}{\int_{\Theta} p(\mathbf{X}|\Theta)p(\Theta)d\Theta}$$

y por lo tanto:

$$p(\Theta|\mathbf{X}) \propto p(\Theta)p(\mathbf{X}|\Theta) \tag{9}$$

Es posible que las creencias a priori de cierto investigador sobre  $\Theta$  puedan diferir de las de otro, pero muy a menudo estarán de acuerdo en la forma en que los datos están relacionados con  $\Theta$  (es decir, en la forma de  $p(\mathbf{X}|\Theta)$ ). Si esto es así, diferirán en sus creencias a posteriori (es decir, en sus creencias después de haber obtenido los datos), pero resultará que si recopilan suficientes datos, sus creencias a posteriori generalmente serán muy similares [16].

Los matemáticos llamaron *verosimilitud* al tratar a  $p(\mathbf{X}|\Theta)$  como función de  $\Theta$ , es decir que:  $l(\Theta|\mathbf{X}) = p(\mathbf{X}|\Theta)$  donde  $l(\Theta|\mathbf{X})$  es la verosimilitud de los datos. Muchas veces es más natural considerar la *log-verosimilitud*, es decir  $L(\Theta|\mathbf{X}) = \log[l(\Theta|\mathbf{X})]$ . Con ésta definición y teniendo en cuenta que  $p(\Theta)$  es la pdf a priori para  $\Theta$  y que  $p(\Theta|\mathbf{X})$  es la

pdf a posteriori para  $\Theta$  dado  $\mathbf{X}$  se tiene que *Posteriori*  $\propto$  *Priori*  $\times$  *Verosimilitud*.

### 2.2.1. Modelos Jerárquicos Bayesianos

Teniendo conocimiento de que los datos siguen una pdf  $p(\mathbf{X}|\Theta)$ , a veces, se puede intuir que existe una extensión de éste conocimiento a priori. Por ejemplo, en una primera etapa las observaciones tienen una densidad que depende de  $k$  parámetros desconocidos  $\Theta = (\theta_1, \theta_2, \dots, \theta_k)$  para los cuales se conoce dicha densidad  $p(\Theta)$ . Entonces se puede hacer una o más suposiciones acerca de las relaciones entre los diferentes parámetros, por ejemplo, que se distribuyen de forma idéntica e independiente (i.i.d) ó que están en orden creciente, etc. Tales relaciones se refieren a menudo como estructurales[16].

Es natural modelar un problema de éste tipo jerárquicamente, con resultados observables modelados condicionalmente en ciertos parámetros, que a su vez reciben una especificación probabilística en términos de parámetros adicionales, conocidos como *hiperparámetros*. Tal pensamiento jerárquico ayuda en la comprensión de los problemas multiparámetros y también desempeña un papel importante en el desarrollo de estrategias computacionales. Quizá más importante aún en la práctica es que los modelos no jerárquicos simples suelen ser inadecuados para los datos jerárquicos ya que con pocos parámetros, generalmente no pueden ajustar con gran precisión los conjuntos de datos grandes, mientras que con muchos parámetros tienden a sobreajustarlos (overfit), es decir, producen modelos que ajustan bien los datos existentes, pero conducen a malas predicciones para los nuevos datos. Por el contrario, los modelos jerárquicos pueden tener suficientes parámetros para ajustar bien los datos, al tiempo que se utiliza una distribución de la población para estructurar cierta dependencia en los parámetros, evitando así problemas de sobreajuste[15].

### 2.3. Computación Bayesiana y el lenguaje de programación Stan

Las estadísticas bayesianas se desarrollan de forma sencilla siempre que se trabaja con distribuciones bien conocidas y distribuciones a priores conjugadas. Pero en la vida real las cosas son distintas, teniendo distribuciones muy complejas en las cuales por ejemplo, no es fácil encontrar el múltiplo constante en  $p(\Theta|\mathbf{X}) \propto p(\Theta)l(\Theta|\mathbf{X})$  y se tiene que recurrir a alguna técnica numérica, las cuales en los últimos años han tenido un avance impresionante como las técnicas MCMC (Markov Chain Monte Carlo). Ésto es

aún más cierto cuando por ejemplo se desea evaluar la densidad marginal de un solo componente  $\theta_i$  del vector  $\Theta$ [16] el cual puede ser muy difícil de calcular sin ayuda de métodos computacionales. Ésta sección presenta las técnicas usadas y características lenguaje de programación usado en éste trabajo.

Stan es un lenguaje de programación probabilístico para especificar modelos estadísticos. Un programa Stan define imperativamente una función de probabilidad sobre parámetros condicionados a datos y constantes especificadas. A partir de la versión 2.14.0, Stan proporciona una inferencia bayesiana completa para modelos de variable continua a través de métodos MCMC, específicamente forma adaptativa de muestreo Hamiltoniano Monte Carlo. Las estimaciones de máxima verosimilitud se calculan utilizando métodos de optimización avanzados [27]. Stan puede ser llamado desde la línea de comandos usando el paquete cmdstan, a través de R usando el paquete RStan.

### 2.3.1. Métodos Monte Carlo

Los métodos de Monte Carlo se desarrollaron para aproximar numéricamente las integrales que no son manejables analíticamente pero para las cuales la evaluación de la función que es integrada es manejable[18]. Como ejemplo: si se quiere calcular la media  $\mu$  de una pdf se lo puede hacer mediante  $\mu = \int_{\Theta} \theta \times p(\theta) d\theta$ . Para modelos incluso moderadamente complejo, las densidades a posterior conducen a una integral que es imposible evaluar analíticamente. Ahora suponiendo que se pueden tomar muestras independientes de  $p(\theta)$  y sean  $\theta^{(1)}, \theta^{(2)}, \dots, \theta^{(N)}$   $N$  de éstas muestras, el estimador de Monte Carlo  $\hat{\mu}$  acerca del parámetro  $\mu$  está dado en forma básica por  $\hat{\mu} = \frac{1}{N} \sum_{n=1}^N \theta^{(n)}$ .

Si la función de probabilidad  $p(\theta)$  tiene media y varianza finita, la ley de grandes números asegura que la estimación de Monte Carlo converge al valor correcto a medida que aumenta el número de muestras, es decir que  $\lim_{N \rightarrow \infty} \hat{\mu} = \mu$ .

Asumiendo la media y la varianza finita, el error de estimación se rige por el teorema del límite central, de modo que el error de estimación disminuye cuando la raíz cuadrada de  $N$  aumenta, así:  $|\mu - \hat{\mu}| \propto \frac{1}{\sqrt{N}}$ .

Por lo tanto, la estimación de una media a un lugar decimal adicional de precisión requiere cien veces más muestras, agregar dos decimales significa diez mil veces más muestras, etc. Esto hace que los métodos de Monte Carlo sean más útiles para estimaciones aproximadas a unas pocas cifras decimales que las estimaciones muy precisas. En aplica-

ciones prácticas, no es tan necesario estimar una cantidad más allá de la incertidumbre de la muestra de datos sobre la que se basa, por lo que ésta "falta de muchos decimales" de exactitud rara vez es un problema en la práctica para los modelos estadísticos[20].

### 2.3.2. Métodos de muestreo Markov Chain Monte Carlo (MCMC)

Los métodos Markov Chain Monte Carlo fueron desarrollados para situaciones en las que no es fácil extraer muestras independientes de una distribución objetivo[25].

Una cadena de Markov es una secuencia de variables aleatorias  $\theta^{(1)}, \theta^{(2)}, \dots$  donde cada variable es condicionalmente independiente de todas las demás variables dado el valor de su estado anterior. Si se tiene  $\theta^{(1)}, \theta^{(2)}, \dots, \theta^{(N)}$  entonces,  $p(\theta) = p(\theta^{(1)}) \prod_{n=2}^N p(\theta^{(n)} | \theta^{(n-1)})$ .

Dado que se desea generar muestras aleatorias de una distribución objetivo, se debe identificar una forma de construir una buena cadena de Markov tal que su distribución de probabilidad de equilibrio se convierta en nuestra distribución objetivo. Si podemos construir tal cadena entonces arbitrariamente comenzamos desde algún punto e iteramos la cadena de Markov muchas veces entonces eventualmente, las muestras que generamos aparecerían como si vinieran de nuestra distribución objetivo. Luego, se aproximan las cantidades de interés (por ejemplo, la media) tomando el promedio de una muestra a partir de la distribución objetivo. En éstos métodos hay varias técnicas de construir cadenas de Markov con buenas características, por ejemplo, el muestreador de Gibbs ó el algoritmo de Metrópolis-Hastings. Éstas técnicas mejoran la precisión y la tasa de convergencia de la integración Monte Carlo sin aumentar el número de experimentos.

Stan usa el método Monte Carlo Hamiltoniano para generar el siguiente estado en una cadena de Markov[20]. Ésta técnica de muestreo es mas eficiente y robusta que los métodos Gibbs y Metrópolis-Hastings puros.

Uno de los métodos MCMC más populares es el algoritmo Metrópolis-Hastings; de hecho la mayoría de los algoritmos de MCMC pueden ser interpretados como casos especiales de éste algoritmo. El problema básico que intenta resolver el algoritmo Metropolis-Hastings es proporcionar un método para generar muestras aproximadamente independientes de alguna distribución genérica,  $p(\theta)$  [30].

La idea básica detrás de este algoritmo es la de construir una cadena de Markov cuya distribución invariante sea la distribución de muestreo que deseamos, es decir  $p(\theta)$ . Supongamos que el estado actual de la cadena de Markov es  $\theta^n$  y queremos generar  $\theta^{n+1}$ ,



entonces el algoritmo genera primeramente un candidato  $\theta^*$ . El valor de  $\theta^*$  se genera a partir de una distribución propuesta denotada como  $Q(\theta^*|\theta^n)$  la cual depende del estado actual de la cadena de Markov,  $\theta^n$ . Luego se tiene una etapa de aceptación-rechazo, para esto se calcula la probabilidad de aceptación así [30]:

$$A(\theta^n \rightarrow \theta^*) = \min \left( 1, \frac{p(\theta^*)}{p(\theta^n)} \cdot \frac{Q(\theta^n|\theta^*)}{Q(\theta^*|\theta^n)} \right) \quad (10)$$

Para aceptar ó rechazar el candidato generamos un número aleatorio uniformemente distribuido entre 0 y 1 que llamamos  $u$ , entonces:

$$\theta^{n+1} = \begin{cases} \theta^* & \text{si } u \leq A(\theta^n \rightarrow \theta^*) \\ \theta^n & \text{si } u > A(\theta^n \rightarrow \theta^*) \end{cases}$$

Las cadenas de Markov que Stan y otros samplers MCMC generan son ergódicas, esto significa que hay una probabilidad razonable de alcanzar un valor de  $\theta$  desde otro y que sus probabilidades de transición de  $n$  pasos convergiran a las probabilidades del estado estable conforme  $n$  se haga mas grande. Las cadenas de Markov son también estacionarias, lo que significa que las probabilidades de transición no cambian para las diferentes posiciones en la cadena, de modo que para  $n, n' \geq 0$ , la función de probabilidad  $p(\theta^{(n+1)}|\theta^{(n)})$  es la misma que  $p(\theta^{(n'+1)}|\theta^{(n')})$ .

Las cadenas de Markov estacionarias tienen una distribución de equilibrio de estados en la que cada uno tiene la misma función de probabilidad marginal, por lo que  $p(\theta^{(n)})$  es la misma función de probabilidad que  $p(\theta^{(n+1)})$ . En Stan, ésta distribución de equilibrio  $p(\theta^{(n)})$  es la función de probabilidad  $p(\theta)$  que está siendo muestreada, típicamente una densidad a posteriori bayesiana [20].

Los métodos MCMC implican dos dificultades: el primer problema es determinar cuándo una cadena de Markov inicialmente ha convergido a su distribución de equilibrio y el segundo problema es que las muestras de una cadena de Markov están correlacionadas y, por tanto, el teorema del límite central vinculado al error de estimación ya no se aplica.

Una cadena de Markov genera muestras de la distribución objetivo sólo después de haber convergido al equilibrio. Por desgracia, esto sólo se garantiza en el límite en teoría. En la práctica debemos aplicar ciertos diagnósticos para vigilar si la(s) cadena(s) de Markov han convergido.

Una manera de monitorear si una cadena ha convergido a la distribución de equilibrio

es comparar su comportamiento con otras cadenas inicializadas al azar. Esta es la motivación para el estadístico de reducción de escala potencial  $\hat{R}$  de Gelman y Rubin (1992) [26]. El estadístico  $\hat{R}$  mide la relación de la varianza media de las muestras dentro de cada cadena con la varianza de las muestras agrupadas a través de las cadenas; si todas las cadenas están en equilibrio, éstas (varianza dentro de cada cadena y la varianza a través de las cadenas) serán las mismas y  $\hat{R}$  será uno. Si las cadenas no han convergido a una distribución común, el estadístico  $\hat{R}$  será mayor que uno. La recomendación de Gelman y Rubin es que las cadenas de Markov independientes se inicialicen con valores iniciales difusos para los parámetros y se muestreen hasta que todos los valores de  $\hat{R}$  estén por debajo de 1,1. Stan permite a los usuarios especificar valores iniciales para los parámetros.

El estadístico  $\hat{R}$  se define para un conjunto de  $M$  cadenas de Markov,  $\theta_m$ , cada una de las cuales tiene  $N$  muestras  $\theta_m^{(n)}$ . La estimación de la varianza entre-muestras [20] es:

$$B = \frac{N}{M-1} \sum_{m=1}^M (\bar{\theta}_m^{(\bullet)} - \bar{\theta}_{\bullet}^{(\bullet)})^2$$

donde:

$$\bar{\theta}_m^{(\bullet)} = \frac{1}{N} \sum_{n=1}^N \theta_m^{(n)}$$

$$\bar{\theta}_{\bullet}^{(\bullet)} = \frac{1}{M} \sum_{m=1}^M \bar{\theta}_m^{(\bullet)}$$

La varianza-dentro de la muestra es

$$W = \frac{1}{M} \sum_{m=1}^M s_m^2$$

donde:

$$s_m^2 = \frac{1}{N-1} \sum_{n=1}^N (\theta_m^{(n)} - \bar{\theta}_m^{(\bullet)})^2$$

El estimador de la varianza es:

$$\widehat{Var}^+(\theta|y) = \frac{N-1}{N} W + \frac{1}{N} B$$

Finalmente el estadístico de reducción de escala potencial es definido como

$$\hat{R} = \sqrt{\frac{\widehat{Var}^+(\theta|y)}{W}} \quad (11)$$

La segunda dificultad técnica planteada por los métodos MCMC es que las muestras típicamente se autocorrelacionan dentro de una cadena. Esto aumenta la incertidumbre de la estimación de las cantidades de interés posteriores, como las medias, las varianzas o los cuantiles [20].

La cantidad por la cual la autocorrelación dentro de las cadenas aumenta la incertidumbre en las estimaciones se puede medir por el tamaño efectivo de la muestra. Dado muestras independientes, el teorema del límite central limita la incertidumbre en estimaciones basadas en el número de muestras  $N$ . Dadas muestras dependientes, el número de muestras independientes se reemplaza con el tamaño de muestra efectivo  $n_{eff}$ , que es el número de muestras independientes con el mismo poder de estimación que las  $N$  muestras autocorrelacionadas; es decir, el error de estimación es proporcional a  $\frac{1}{\sqrt{n_{eff}}}$  en lugar de  $\frac{1}{\sqrt{N}}$  [20].

El tamaño de muestra efectivo de una secuencia se define en términos de las autocorrelaciones dentro de la secuencia en diferentes tiempos. La autocorrelación  $\rho_t$  para el tiempo  $t \geq 0$  para una cadena con función de probabilidad conjunta  $p(\theta)$  con media  $\mu$  y varianza  $\sigma^2$  se define como:

$$\rho_t = \frac{1}{\sigma^2} \int_{\Theta} (\theta^{(n)} - \mu)(\theta^{(n+t)} - \mu)p(\theta)d\theta$$

Esta es la correlación entre las dos cadenas retardadas por  $t$  posiciones. Se conoce que cuando  $\theta^{(n)}$  y  $\theta^{(n+t)}$  logran estacionarse en una MCMC, tienen la misma distribución marginal; por lo que podemos reducirla a:

$$\rho_t = \frac{1}{\sigma^2} \int_{\Theta} \theta^{(n)}\theta^{(n+t)}p(\theta)d\theta$$

Entonces el tamaño muestral efectivo para  $N$  muestras generado por un proceso con autocorrelación  $\rho_t$  es definido por:

$$n_{eff} = \frac{N}{\sum_{t=-\infty}^{\infty} |\rho_t|} = \frac{N}{1 + 2 \sum_{t=1}^{\infty} |\rho_t|} \quad (12)$$

En la práctica, la función de probabilidad en cuestión no puede integrarse de forma fácil y, por lo tanto, no se puede calcular la autocorrelación ni el tamaño de muestra efectivo. En su lugar, estas cantidades deben estimarse a partir de las muestras mismas.

### 2.3.3. Muestreo Hamiltoniano Monte Carlo (HMC)

Stan utiliza el algoritmo Hamiltoniano Monte Carlo para obtener las muestras de la distribución objetivo. Éste es un método Markov Chain Monte Carlo que utiliza las derivadas de la función de densidad que se muestrea para generar transiciones más eficientes que con otros métodos.

Sea  $p(\theta)$  nuestra densidad objetivo, típicamente una distribución a posteriori bayesiana  $p(\theta|y)$  dado los datos  $y$ , el método HMC introduce variables *momentum*  $\rho$  que son auxiliares y toma muestras de la densidad conjunta  $p(\rho, \theta) = p(\rho|\theta)p(\theta)$  [28].

En la mayoría de aplicaciones HMC, ésta densidad auxiliar es una normal multivariada que no depende de los parámetros  $\theta$ , es decir que  $\rho \sim \text{Multinormal}(0, \Sigma)$  donde la matriz de covarianza  $\Sigma$  actúa como una métrica Euclídeana para rotar y escalar la distribución objetivo [29]. En Stan, ésta matriz puede ser escogida como una matriz identidad y su inversa  $\Sigma^{-1}$  es conocida como la matriz de masa [20].

La densidad conjunta  $p(\rho, \theta)$  define un hamiltoniano como:

$$\begin{aligned} H(\rho, \theta) &= -\log p(\rho, \theta) \\ &= -\log p(\rho|\theta) - \log p(\theta) \\ &= T(\rho|\theta) + V(\theta) \end{aligned}$$

Donde  $T(\rho|\theta) = -\log p(\rho|\theta)$  es llamada la *energía cinética* y el término  $V(\theta) = -\log p(\theta)$  es llamada *energía potencial*, la cual es especificada por el programa Stan.

Estando  $\theta$  en su valor actual, una transición a un nuevo estado es generada en dos partes antes de ser sujeta a un paso de aceptación o rechazo de Metrópolis. Primero, se muestrea un valor para el momentum, independientemente de los valores de los parámetros actuales,  $\rho \sim \text{Multinormal}(0, \Sigma)$ . Luego, el sistema conjunto  $(\rho, \theta)$  que está hecho de los valores de los parámetros actuales  $\theta$  y de un nuevo momentum  $\rho$  es evolucionado vía las ecuaciones de Hamilton [20]:

$$\frac{d\theta}{dt} = + \frac{\partial T}{\partial \rho} \quad (13)$$

$$\frac{d\rho}{dt} = - \frac{\partial V}{\partial \theta} \quad (14)$$

Ésto nos deja por resolver un sistema de ecuaciones diferenciales, para el cual, Stan utiliza el integrador Leapfrog, que es un algoritmo de integración numérica que está específicamente adaptado para proporcionar resultados estables para los sistemas de ecuaciones hamiltonianos [20]. Lo que se hace en la práctica para considerar los errores numéricos durante la integración es aplicar un paso de aceptación-rechazo de Metropólis, donde la probabilidad de guardar la densidad propuesta  $(\rho^*, \theta^*)$  generada por  $(\rho, \theta)$ :

$$\min(1, \exp(H(\rho, \theta) - H(\rho^*, \theta^*))) \quad (15)$$

Si la propuesta no es aceptada, el valor del parámetro anterior es retornado para la siguiente muestra y se utiliza para inicializar la siguiente iteración.

## 2.4. Método con Regresión Multinomial Penalizada.

Otra forma de encontrar las proporciones de apoyo por clase desde un recinto  $i$  a un partido político determinado, es tratarlo como un problema de regresión. Suponiendo que se tiene múltiples salidas  $Y_1, Y_2, \dots, Y_K$  que se desean predecir a partir de las entradas  $X_0, X_1, X_2, \dots, X_p$ , se asume un modelo lineal para cada salida[33]:

$$Y_k = \beta_{0k} + \sum_{j=1}^p X_j \beta_{jk} + \varepsilon_k \quad (16)$$

$$= f_k(X) + \varepsilon_k \quad (17)$$

$$Y = XB + E \quad (18)$$

Donde  $Y$  representa la matriz de salidas de tamaño  $N \times K$ ,  $X$  es la matriz de datos  $N \times (p + 1)$ ,  $B$  representa la matriz de parámetros desconocidos de tamaño  $(p + 1) \times K$  y  $E$  es la matriz de tamaño  $N \times K$  de errores.

Tradicionalmente, esto implica  $n < p$  y normalmente se estima una regresión con mínimos cuadrados. Sin embargo, si  $p \gg n$  (como sucede en muchas aplicaciones) la

regresión falla [34]. En éstos casos, lo que interesa es reducir el número de covariables para obtener la solución. Esto implica forzar a que alguna(s) componentes del vector de parámetros  $\beta$  sean contraídas o eventualmente se hagan cero.

El fin de la penalización del vector  $\beta$  es reducir el *salvajismo de los coeficientes* a costa de perder un poco en la bondad de ajuste del modelo [34]. Para esto, se han creado varios métodos de regresión penalizados como el Ridge y Lasso.

La regresión Ridge reduce los coeficientes de regresión al imponer una penalización en su tamaño [33], el método es muy similar a la de mínimos cuadrados a excepción de que los coeficientes se estiman minimizando una cantidad diferente:

$$\hat{\beta}^{ridge} = \underset{\beta}{\operatorname{argmin}} \left\{ \sum_{i=1}^N \left( y_i - \beta_0 - \sum_{j=1}^p x_{ij} \beta_j \right)^2 + \lambda \sum_{j=1}^p \beta_j^2 \right\} \quad (19)$$

donde el primer sumando representa el error de mínimos cuadrados *RSS* (*ResidualSumSquares*) y el segundo sumando representa la el factor de penalización,  $\lambda \geq 0$  es el parámetro que controla la cantidad de penalización en los parámetros. Éste método tiende a contraer los coeficientes de regresión al incluir el término de penalización en la función objetivo: cuanto mayor sea , mayor penalización y por tanto, mayor contracción de los coeficientes.

El método Ridge produce un conjunto diferente de  $\beta$ 's para cada valor de  $\lambda$  y no un único vector de coeficientes estimados. Si  $\lambda = 0$  se tiene el caso de mínimos cuadrados ordinarios. En otro caso si  $\lambda \rightarrow \infty$  entonces  $\hat{\beta} \rightarrow 0$ .

Uno de los inconvenientes de este método es que contrae todos los coeficientes hacia cero, pero no consigue la nulidad de alguno de ellos. Por tanto, no se produce selección de variables, permaneciendo en el modelo todas las variables. Este hecho resulta un inconveniente en aquellos estudios que tienen un elevado número  $p$  de variables explicativas o predictores. Para eludir este inconveniente se propuso la regresión Lasso, que se explica a continuación.

Tibshirani [1996] propuso la técnica Lasso (least absolute shrinkage and selection operator) para calcular los coeficientes de regresión resolviendo el problema de mínimos cuadrados penalizado. A partir de cierto valor del parámetro  $\lambda$ , el estimador de Lasso produce estimaciones nulas para algunos coeficientes y no nulas para otros, con lo cual Lasso realiza una especie de selección de variables. Lasso reduce la variabilidad de las estimaciones por la reducción de los coeficientes y al mismo tiempo produce modelos

interpretables por la reducción de algunos coeficientes a cero. El método Lasso estaría dado por:

$$\hat{\beta}^{lasso} = \underset{\beta}{\operatorname{argmin}} \left\{ \frac{1}{2} \sum_{i=1}^N \left( y_i - \beta_0 - \sum_{j=1}^p x_{ij} \beta_j \right)^2 + \lambda \sum_{j=1}^p |\beta_j| \right\} \quad (20)$$

En general, los modelos generalizados Lasso son mucho más fáciles de interpretar que los obtenidos mediante Ridge. Al igual que antes, queda por buscar el mejor valor del parámetro  $\lambda$  usando el método de validación cruzada. Al igual que la regresión Ridge, cuando las estimaciones de mínimos cuadrados tienen excesivamente alta varianza, la solución Lasso puede producir una reducción de la varianza a expensas de un pequeño aumento de sesgo, y por tanto, puede generar predicciones más exactas.

Como puede observarse, todas estas técnicas de mínimos cuadrados penalizados dependen de un parámetro de penalización  $\lambda$ , que controla la importancia dada a la penalización en el proceso de optimización. Cuanto mayor es  $\lambda$ , mayor es la penalización en los coeficientes de regresión y más se contraen éstos hacia cero. Un método automático para la estimación de  $\lambda$ , pero intensivo computacionalmente, consiste en usar la técnica de validación cruzada.

El método de validación cruzada (*cross-validation*) consiste en dividir el modelo en un set de entrenamiento (training set) para ajustar un modelo y un set de prueba (test set) para evaluar su capacidad predictiva, mediante el error de predicción u otra medida. La forma en que se aplica la validación cruzada es mediante la división del conjunto de datos disponibles de manera aleatoria en  $k$  subconjuntos de igual tamaño y mutuamente excluyentes. Uno de los subconjuntos se utiliza como datos de prueba y el resto ( $K - 1$ ) como datos de entrenamiento. El proceso de validación cruzada es repetido durante  $k$  iteraciones, con cada uno de los posibles subconjuntos de datos de prueba. Finalmente se realiza la media aritmética de los resultados de cada iteración para obtener un único resultado. Este método es muy preciso puesto que evaluamos a partir de  $K$  combinaciones de datos de entrenamiento y de prueba, pero aun así tiene una desventaja, y es que es lento desde el punto de vista computacional. La validación cruzada con  $k = 10$  es una de las más utilizadas, pero hay que tener en cuenta el número de observaciones del que disponemos.

### 3. Datos y métodos aplicados

En ésta sección se presenta de forma general los procedimientos que se realizaron con la información obtenida desde las diferentes fuentes que pueden estar abiertas al público en el caso de los censos nacionales o puede ser más restringida como en el caso de la información por recintos de los resultados electorales. Generalmente, los datos llegan al investigador en forma cruda, es decir, sin ningún tipo de acondicionamiento como para ser tratados directamente por lo que se vuelve necesario realizar un proceso para tenerlos en forma *normada* en el sentido de que sean uniformes y completos.

#### 3.1. Datos

El análisis y la inferencia realizados en éste trabajo se han hecho en base a los datos obtenidos en el Consejo Nacional Electoral tomados de las elecciones presidenciales del año 2013 y del último censo a nivel nacional realizado por el Instituto Ecuatoriano de Estadísticas y Censos del año 2010. Los datos correspondientes a las elecciones presidenciales están dispuestos a nivel de juntas electorales, lo cual significa tener varios miles de datos que al ser procesados, requieren una gran cantidad de tiempo usado por el algoritmo correspondiente, por lo que, para efectos de implementación se procesó la información para obtener datos agregados a nivel de parroquia teniendo un total de 1222 observaciones. La información dispuesta en ésta base indica que la inferencia se hará con un promedio del 85 % de la población total empadronada en cada parroquia.

Los datos obtenidos del censo nacional presentan información de alrededor de 1000 parroquias, por lo que para unir las dos bases de datos de modo que podamos un conjunto de datos completo y uniforme, se tuvieron que eliminar aquellas parroquias presentes en la base de datos de las elecciones y que no aparecen en la base del censo nacional. Debido a que la información disponible del censo es del año 2010, se utilizó la tasa de crecimiento poblacional a nivel parroquial [17] para empatar con la información de las elecciones presidenciales del año 2013 [3]. En este trabajo no se toma en cuenta la población que vive fuera del país y para el análisis de la variable *edad* se considera tanto la población obligada a votar así como también la que tiene voto facultativo (16 a 18 y mayores a 65 años).

Las variables demográficas utilizadas en éste estudio son: sexo, edad, etnias, población económicamente activa e inactiva (PEA y PEI respectivamente), población con necesi-



dades básicas insatisfechas (NBI) y población que tiene acceso a un seguro de vida como el IESS, ISSFA, ISSPOL y Seguro Campesino. El método propuesto en éste trabajo se puede comparar solamente para la variable *sexo* que está presente en la base de datos de las elecciones, aunque se puede realizar una comparación para otras variables, si se tiene resultados de *exit polls* realizados en el año 2013, pero ese trabajo está fuera del alcance de ésta tesis.

La representación gráfica de los parámetros inferidos se muestra en un mapa del Ecuador dividido a nivel de parroquias el mismo que se ha obtenido de la página web: Global Administrative Areas GADM y tiene extensión *.rds* (*SpatialPolygonsDataFrame*) el cual puede utilizarse en *R* mediante el paquete *sp*.

## 3.2. Metodología

Los métodos desarrollados en éste trabajo están basados en dos soportes estadísticos principales: La inferencia ecológica a través de modelos bayesianos y un conjunto de técnicas computacionales que permiten determinar las cantidades de interés por medio de distribuciones a posteriori. La inferencia ecológica se la realizó a través de un modelo Binomial- Beta para el caso de tablas de contingencia 2x2 y un modelo Multinomial-Dirichlet para el caso de tablas RxC. El cálculo computacional necesario se lo realizó usando métodos MCMC ya establecidos en la sección anterior usando Stan.

El segundo método usado es la regresión multinomial penalizada como se describe en la sección 2.4, cabe señalar que para este proceso se utilizó la librería *glmnet* cuyo procedimiento se describen en la sección 3.5.

### 3.2.1. El modelo Bayesiano

El modelado Bayesiano se puede resumir en tres pasos [15]:

1. Establecer un modelo de probabilidad completa para todas las cantidades observables y no observables. Este modelo debe ser consistente con el conocimiento existente de los datos que se modelan.
2. Calcular la probabilidad posterior de cantidades desconocidas condicionadas a las cantidades observadas. Las incógnitas pueden incluir cantidades no observables tales como parámetros y cantidades potencialmente observables tales como predicciones para futuras observaciones.

3. Evaluar el ajuste del modelo a los datos.

Típicamente, este ciclo se repetirá hasta que se alcance un ajuste suficiente en la tercera etapa. Stan automatiza los cálculos involucrados en el segundo y tercer pasos. Ahora se presenta los modelos Binomial-Beta y Multinomial-Dirichlet para los casos de inferencia ecológica 2x2 y RxC respectivamente.

### 3.2.2. Algoritmo HMC

El algoritmo que realiza el método Hamiltoniano Monte Carlo se resume en los siguientes pasos y cuyo flujograma se muestra en la figura 3 [20]:

1. El algoritmo arranca en un conjunto inicial de parámetros  $\theta$ . En Stan, éste conjunto puede ser especificado por el usuario o generado aleatoriamente.
2. Luego, para un número de iteraciones dado, un nuevo vector momentum es muestreado y el valor actual del parámetro  $\theta$  se actualiza usando el método integrador leapfrog con un tiempo de discretización y número de pasos dado, de acuerdo a la dinámica Hamiltoniana.
3. Entonces, se aplica un paso de aceptación-rechazo dado por el algoritmo de Metrópolis explicado en la sección 2.3.2, para decidir si se actualiza al nuevo estado  $(\rho^*, \theta^*)$  ó si mantenemos el estado actual.

A continuación se muestran los modelos jerárquicos bayesianos para tablas de tamaño 2x2 y RxC que se han utilizado en este trabajo.

### 3.2.3. Modelo Jerárquico Bayesiano para tablas 2x2 sin covariable

El modelo jerárquico Beta - Binomial utilizado para tablas de contingencia 2x2 e implementado en RStan para obtener inferencias sobre las proporciones  $\beta^b$  y  $\beta^w$  se describe a continuación.

Si se tiene  $p$  recintos (parroquias), se denomina  $T'_i$ , con  $i = 1 \dots p$ , al número total de votantes que apoyan a AP en cada recinto, entonces en el primer nivel de jerarquía se asume que:

$$T'_i \sim Bin(N_i, \theta_i)$$

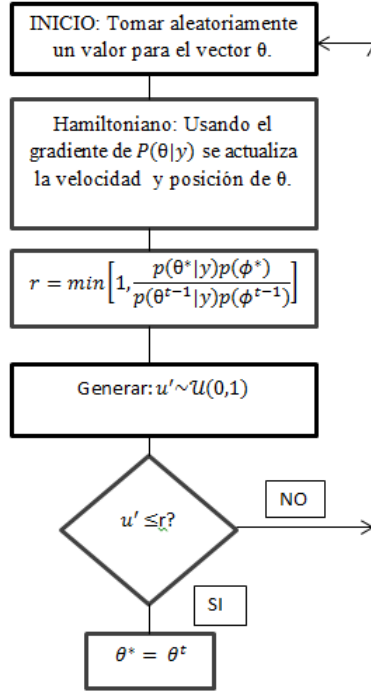


Figura 3: Diagrama de flujo del algoritmo HMC.

donde:  $N_i$  indica el número total de votantes en cada recinto  $i$  y la probabilidad  $\theta_i$ , expresada como  $\theta_i = X_i\beta_i^b + (1 - X_i)\beta_i^w$  indica la probabilidad de que los votantes apoyen a AP.

Ya que la distribución *Beta* puede ser usada para modelar un fenómeno aleatorio cuyo conjunto de posibles valores están en un intervalo finito  $[c, d]$  [19], en este caso los posibles valores de  $\beta^b$  y  $\beta^w$  están el intervalo  $[0, 1]$ , también debido a que las distribuciones *Beta* y *Binomial* pertenecen a la misma familia exponencial y por ser distribuciones conjugadas (ver demostración en Anexos), entonces en el segundo nivel del modelo jerárquico se asume que  $\beta^b$  y  $\beta^w$  se muestrean independientemente de una distribución *Beta* con parámetros  $c_b$ ,  $d_b$ ,  $c_w$  y  $d_w$  respectivamente. La comparación de las distribuciones *Beta* y *Binomial* sugiere que la distribución a priori *Beta* es equivalente a  $c_b - 1$  éxitos y  $d_b - 1$  fallos a priori, lo mismo con  $c_w$  y  $d_w$ , entonces:

$$\beta_i^b \sim Beta(c_b, d_b)$$

$$\beta_i^w \sim Beta(c_w, d_w)$$

Debido a que no se tiene información sobre los hiperparámetros de la distribución *Beta*, se utiliza una distribución exponencial con media  $1/\lambda = 2$  para los mismos, lo que significa usar una distribución a priori no informativa en este nivel [24], es decir:

$$c_b, d_b, c_w, d_w \sim \exp(\lambda)$$

Como se menciona en la sección 2.2 y por el teorema de Bayes, la distribución posterior de los parámetros es proporcional a la verosimilitud multiplicada por la distribución a priori, es decir:

$$Posteriori \propto Priori \times Verosimilitud$$

por lo tanto, para el modelo jerárquico especificado anteriormente la distribución posterior de los parámetros es [24]:

$$\begin{aligned} p(\beta_i^b, \beta_i^w | \text{datos}) &\propto p(\text{datos} | (\beta_i^b, \beta_i^w)) \times p((\beta_i^b, \beta_i^w) | c_b, d_b, c_w, d_w) \times p(c_b, d_b, c_w, d_w) \\ &= \prod_{i=1}^p (X_i \beta_i^b + (1 - X_i) \beta_i^w)^{T_i'} (1 - X_i \beta_i^b - (1 - X_i) \beta_i^w)^{(N_i - T_i')} \\ &\quad \times \prod_{i=1}^p \frac{\Gamma(c_b + d_b)}{\Gamma(c_b) \Gamma(d_b)} (\beta_i^b)^{c_b - 1} (1 - \beta_i^b)^{d_b - 1} \prod_{i=1}^p \frac{\Gamma(c_w + d_w)}{\Gamma(c_w) \Gamma(d_w)} (\beta_i^w)^{c_w - 1} \\ &\quad (1 - \beta_i^w)^{d_w - 1} \times \exp(-\lambda c_b) \times \exp(-\lambda c_w) \times \exp(-\lambda d_b) \times \exp(-\lambda d_w) \end{aligned}$$

Cabe señalar que, con el fin de obtener muestras de la distribución posterior mediante simulación a través de RStan, la distribución a priori y la verosimilitud no se escriben de forma explícita como se muestra en el anterior párrafo sino que mas bien se especifican independientemente la distribución de cada parámetro como se muestra en el bloque *model* de la figura 18 en el apéndice A1.

### 3.2.4. Modelo Jerárquico Bayesiano para tablas 2x2 con el uso de covariables

Incluir una apropiada covariable en el modelo puede mejorar la inferencia de los parámetros, éste es un punto clave ya que se puede incluir una amplia cantidad de información la cual no está representada en  $X_i$ ,  $T_i$  o  $N_i$ . Mediante el uso de covariables es posible complementar la información que se pierde en la agregación de los datos y por lo tanto alcanzar inferencias mas confiables [21] [24]. Enfocar el modelo de esta manera

también permite tener una amplia familia de modelos de los cuales se puede escoger el mejor.

Por otra parte, usar covariables permite tener distribuciones posteriores de formas mas complicadas y al condicionar en  $X_i$ , o alguna correlación de este, se puede modelar la relación que hay entre esta información y el parámetro  $\beta^b$  o  $\beta^w$  en vez de asumir que son a priori independientes.[24]

Siguiendo la notación de la sección anterior se asume que  $Z_i$  es la covariable asociada al recinto  $i$ , aunque  $Z_i$  puede ser un vector, en este estudio se usará como un escalar.

En el primer nivel del modelo jerárquico se asume nuevamente que:

$$T'_i \sim Bin(N_i, \theta_i)$$

donde:  $N_i$  indica el número de votantes en el recinto  $i$  y  $\theta_i = X_i \beta_i^{b(Z_i)} + (1 - X_i) \beta_i^{w(Z_i)}$  indica la probabilidad de que los votantes de los recintos  $i$  apoyen a AP y tanto  $\beta_i^{b(Z_i)}$  como  $\beta_i^{w(Z_i)}$  indican la dependencia de los parámetros  $\beta_i^b$  y  $\beta_i^w$  en la covariable  $Z_i$ .

En esta sección se suprime la dependencia en  $Z_i$  de los parámetros  $\beta_i^b$  y  $\beta_i^w$  por facilidad de la notación.

Por las mismas razones establecidas para el modelo de tablas 2x2 sin covariables, en el segundo nivel de este modelo se utiliza una distribución *Beta* como se muestra a continuación:

$$\beta_i^b \sim Beta(d_b \exp(\alpha + \beta Z_i), d_b)$$

$$\beta_i^w \sim Beta(d_w \exp(\gamma + \delta Z_i), d_w)$$

La interpretación de los parámetros de la distribución *Beta*, en términos del número de *éxitos* (apoyo a AP) y *fracasos* (apoyo a otros partidos), dependen de la función que se utilice para tener una estimación puntual de la distribución, en este caso se utiliza la media y por lo tanto, los parámetros de la distribución *Beta* representan  $d_b \exp(\alpha + \beta Z_i)$  éxitos y  $d_b$  fracasos.

Por otra parte, tanto la probabilidad  $\theta_i$  como los parámetros  $\beta^b$  y  $\beta^w$  de la distribución *Binomial* del primer nivel, dependen de la covariable  $Z_i$  de la forma como a continuación se describe: dado que la media de una distribución  $Beta(a, b)$  es  $\frac{a}{a+b}$ , entonces la media

de  $\beta_i^b$ , que puede escribirse como  $E(\beta_i^b)$ , es:

$$\begin{aligned} E(\beta_i^b) &= \frac{d_b \exp(\alpha + \beta Z_i)}{d_b + d_b \exp(\alpha + \beta Z_i)} \\ &= \frac{\exp(\alpha + \beta Z_i)}{1 + \exp(\alpha + \beta Z_i)} \\ &= \frac{1}{1 + \exp^{-(\alpha + \beta Z_i)}} \\ e^{\alpha + \beta Z_i} &= \frac{E(\beta_i^b)}{1 - E(\beta_i^b)} \end{aligned}$$

Finalmente se tiene que:

$$\alpha + \beta Z_i = \ln \left( \frac{E(\beta_i^b)}{1 - E(\beta_i^b)} \right)$$

donde:  $\ln\left(\frac{E(\beta_i^b)}{1 - E(\beta_i^b)}\right)$  representa el *logit* de las probabilidades binomiales desconocidas, lo que indica que el logaritmo natural de la razón de momios dependen linealmente en la covariable  $Z_i$ , siendo  $\alpha$  y  $\beta$  el intercepto y la pendiente de la regresión respectivamente [23].

De igual manera para  $\beta_i^w$  se tiene:

$$\gamma + \delta Z_i = \ln \left( \frac{E(\beta_i^w)}{1 - E(\beta_i^w)} \right)$$

Debido a que no se tiene información respecto a los hiperparámetros de la distribución *Beta*, se asume que los parámetros de regresión:  $\alpha$ ,  $\beta$ ,  $\gamma$  y  $\delta$  siguen distribuciones *flat prior* independientes y de forma similar que en el modelo de la sección anterior, los parámetros  $d_b$  y  $d_w$  siguen distribuciones exponenciales con media  $1/\lambda = 2$ , es decir, una distribución a priori no informativa[24] mostrando la falta de conocimiento sobre los mismos.

$$d_b \sim \exp(\lambda)$$

$$d_w \sim \exp(\lambda)$$

### 3.2.5. Modelo Jerárquico Bayesiano para tablas de tamaño $R \times C$

El modelo Beta - Binomial descrito en la sección 3.2.2 se extiende en ésta sección a un modelo *Multinomial - Dirichlet*, permitiendo de ésta forma el análisis de tablas  $R \times C$  con  $C > 2$  y  $R \geq 2$ .

En el primer nivel del modelo se asume que el vector que indica el número total de

votantes por cada uno de los partidos políticos, (tabla 2), en la parroquia  $i$ , es decir el vector  $(T'_{1i}, T'_{2i}, T'_{3i})$ , sigue una distribución multinomial con un vector de probabilidades  $\theta_i = (\theta_{1i}, \theta_{2i}, \theta_{3i})^t$  y  $N_i$ , donde  $\theta_{ci} = \sum_{r=1}^R \beta_{rci} X_{ri}$ ,  $c = 1, \dots, 3$

$$(T'_{1i}, T'_{2i}, T'_{3i}) \sim \text{Multinomial} \left( N_i, \sum_{r=1}^R \beta_{r1i} X_{ri}, \sum_{r=1}^R \beta_{r2i} X_{ri}, \sum_{r=1}^R \beta_{r3i} X_{ri} \right)$$

$N_i$  es el total de votantes en cada parroquia y se tiene la restricción  $\sum_{c=1}^3 \theta_{ci} = 1$ .

En el segundo nivel del modelo jerárquico se asume que:

$$(\beta_{r1i}, \beta_{r2i}, \beta_{r3i}) \sim \text{Dirichlet}(\alpha_{r1}, \alpha_{r2}, \alpha_{r3})$$

donde: el vector  $(\beta_{r1i}, \beta_{r2i}, \beta_{r3i})^t$  corresponde a los parámetros de interés a inferir por cada grupo de edad para la parroquia  $i$ , es decir: si  $r = 1$  entonces  $(\beta_{11i}, \beta_{12i}, \beta_{13i})^t$  representa el vector de parámetros que indican las proporciones de apoyo del grupo de votantes entre 16 y 29 años para los partidos AP, CREO y OTROS.

Finalmente en el tercer nivel de jerarquía del modelo se tiene que la distribución a *priori* de los hiperparámetros  $\alpha_{rc}$  siguen una distribución Gamma:

$$\alpha_{rc} \sim \text{Gamma}(\lambda_1, \lambda_2)$$

En esta tesis se utilizó este modelo para determinar las proporciones de apoyo a los diferentes partidos políticos usando la variable *edad* cuyos resultados se verán mas adelante. En el tercer nivel del modelo se fijaron los valores de  $\lambda_1 = 4$  y  $\lambda_2 = 2$  para la distribución Gamma [21][22].

Al igual que para tablas 2x2, el modelo puede ajustarse utilizando una covariable conveniente que permita mejorar la inferencia de los parámetros, en tal caso, la distribución del vector  $\beta_{ri}$  es:

$$(\beta_{r1i}, \beta_{r2i}, \beta_{r3i}) \sim \text{Dirichlet}(d_r \exp(\gamma_{r1} + \delta_{r1} Z_i), d_r \exp(\gamma_{r2} + \delta_{r2} Z_i), d_r)$$

Si un determinado vector  $\theta = (\theta_1, \dots, \theta_k)$  sigue una distribución  $\text{Dirichlet}(\alpha_1, \dots, \alpha_k)$

entonces:  $E(\theta_j) = \frac{\alpha_j}{\alpha_0}$ , donde  $\alpha_0 = \sum_{j=1}^k \alpha_j$  [15], por lo tanto para este caso:

$$\begin{aligned} E(\beta_{rc}^i) &= \frac{d_r \exp(\gamma_{rc} + \delta_{rc} Z_i)}{d_r (1 + \sum_{j=1}^{C-1} \exp(\gamma_{rj} + \delta_{rj} Z_i))} \\ &= \frac{\exp(\gamma_{rc} + \delta_{rc} Z_i)}{1 + \sum_{j=1}^{C-1} \exp(\gamma_{rj} + \delta_{rj} Z_i)} \end{aligned}$$

para  $i = 1, \dots, p$ ,  $r = 1, \dots, 4$  y  $c = 1, \dots, 3$ , lo que implica (de forma similar que en la sección 3.2.3) que:

$$\gamma_{rc} + \delta_{rc} Z_i = \ln \left( \frac{E(\beta_{rc}^i)}{E(\beta_{rC}^i)} \right)$$

donde  $C = 3$  y  $\ln \left( \frac{E(\beta_{rc}^i)}{E(\beta_{rC}^i)} \right)$  indica que el logaritmo natural de la razón de momios depende linealmente de la covariable  $Z_i$ .

Por lo tanto, la contribución de los datos de la parroquia  $i$  a la verosimilitud es:

$$\theta_{1i}^{T'_{1i}} \times \theta_{2i}^{T'_{2i}} \times \left( 1 - \sum_{c=1}^2 \theta_{ci} \right)^{N_i - \sum_{c=1}^2 T'_{ci}}$$

La distribución de los hiperparámetros de la regresión,  $\gamma_{rc}$  y  $\delta_{rc}$  se muestrean independientemente a partir de una distribución *flat prior* y finalmente para los hiperparámetros  $d_r$  se asume [21] [22]:

$$d_r \sim \exp(\lambda)$$

En esta tesis no se abordó el problema RxC con covariables debido a la falta de información que sea conveniente para mejorar la inferencia de la variable *edad*.

### 3.3. Resultados

Una vez procesada la información necesaria, se cuenta con alrededor de 1000 observaciones (dependiendo de las variables a considerar).

Cada observación consta de los valores marginales de la tabla de contingencia descrita en la tabla 2 y también cada observación corresponde a una parroquia a nivel nacional.

#### 3.3.1. Resultados para tablas de contingencia 2x2 sin covariable

En primer lugar se realiza la inferencia del electorado usando la variable *sexo*: hombres y mujeres, para analizar el apoyo que obtuvo Alianza País de parte de éstas clases.



De forma general se tratará a las proporciones de apoyo de hombres y mujeres como  $\beta_i^b$  y  $\beta_i^w$  respectivamente para cada parroquia  $i$ .

Tomando como ejemplo la parroquia de Cumbayá, ubicada en la provincia de Pichincha, el modelo infiere los siguientes resultados:  $\beta^b=0.32$  y  $\beta^w=0.57$  con desviaciones estándar de  $\sigma^b=0.15$  y  $\sigma^w=0.13$  y con intervalos de credibilidad del 95% de (0,15;0,64) y (0,26;0,79) respectivamente.

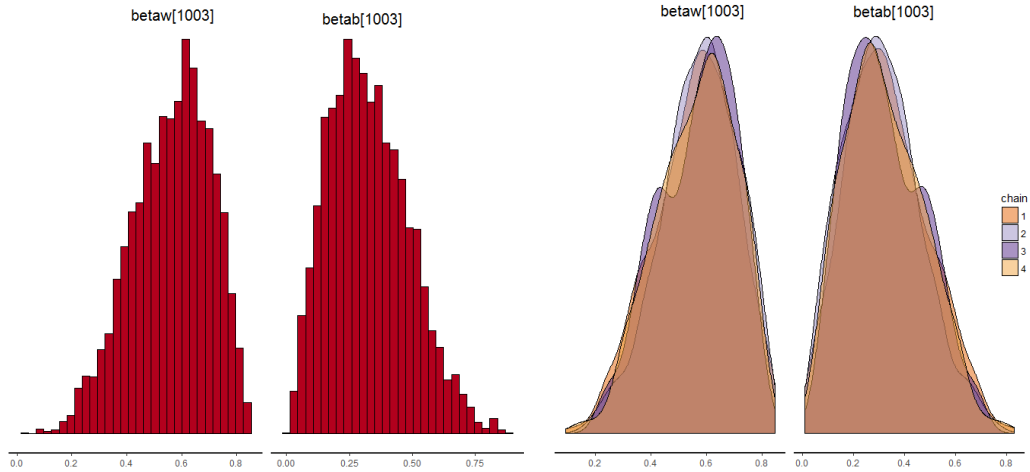


Figura 4: Histogramas y densidad posterior para las proporciones  $\beta^b$  y  $\beta^w$ .

La figura 4 muestra, los histogramas de las proporciones de interés ( $\beta^b$  y  $\beta^w$ ) con sus correspondientes densidades a posteriori luego de haber simulado el modelo en RStan. Uno de los principales problemas al inferir parámetros en tablas de contingencia es que existe el riesgo de que aparezca lo que se denomina *bias aggregation*, lo cual afortunadamente se puede evaluar usando un diagrama de dispersión para determinar el nivel de correlación que existe entre la proporción de votantes hombres ( $X_i$ ) y la proporción de votantes hombres que apoyan a AP ( $\beta^b$ ) así como también la correlación presente entre la proporción de votantes mujeres ( $1 - X_i$ ) y la proporción de mujeres que apoyaron a AP ( $\beta^w$ ).

La figura 5 muestra los diagramas de dispersión de las proporciones mencionadas en el párrafo anterior cuyos valores de correlación entre  $X_i$  y  $\beta^b$  y entre  $1 - X_i$  y  $\beta^w$  son: -0.69 y 0.33 respectivamente lo que indica que existe una importante correlación entre  $X_i$  y  $\beta^b$  lo cual significa que los parámetros  $\beta^b$  inferidos son sesgados debido a la agregación

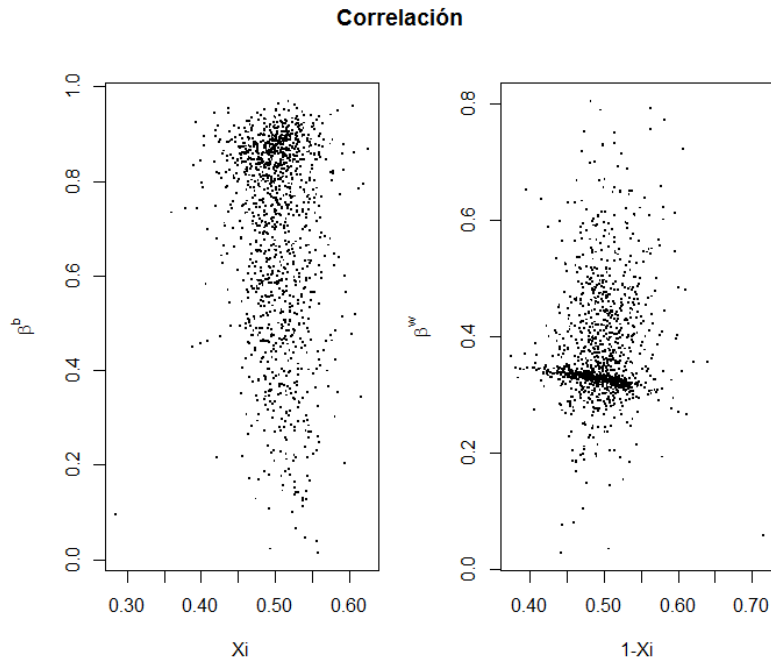


Figura 5: Correlación entre las proporciones  $X_i - \beta^b$  y  $(1-X_i) - \beta^w$ .

de los datos.

Una forma de reducir el sesgo de agregación y mejorar la inferencia es apoyarse en una o varias covariables que permitan clasificar de mejor manera el grupo en estudio, una covariable adecuada también puede ser la misma proporción  $X_i$  y en este trabajo se utiliza esta covariable con el fin de mejorar la inferencia utilizando para ello el modelo *Beta-Binomial* con covariables, cuyos resultados se muestran en la siguiente sección.

Dado que la información real acerca del apoyo que recibió AP por parte de hombres y mujeres está disponible en la base de datos de las elecciones presidenciales, se evalúa el ajuste del modelo al comparar la información real con los parámetros inferidos. Esta comparación se puede ver en la figura 6 la cual muestra el ajuste que existe entre los  $\beta$ 's reales y los  $\beta$ 's predichos, al graficarlos como un par ordenado y observar que tan distantes están de la recta con pendiente unitaria.

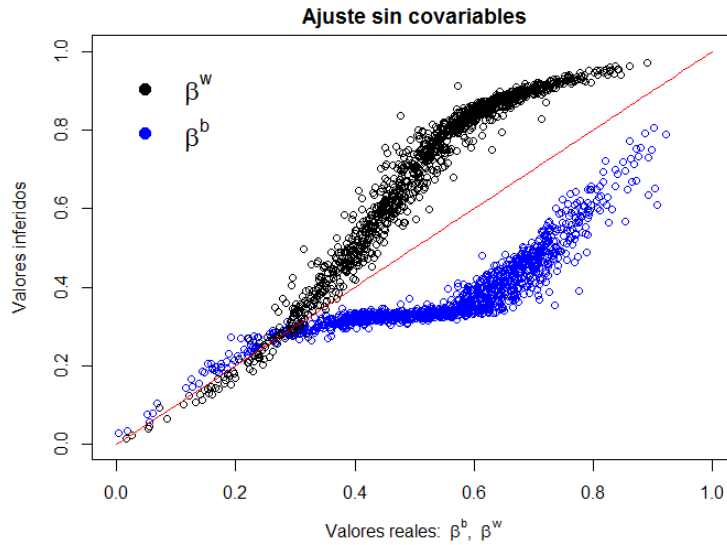


Figura 6: Ajuste entre las proporciones reales e inferidas.

Como se mencionó antes, la inferencia de las proporciones se puede mejorar utilizando  $X_i$  como covariable, lo cual es necesario en vista del resultado observado en la figura 6, que muestra que la inferencia no es buena con el modelo sin uso de covariables.

### 3.3.2. Resultados para tablas de contingencia 2x2 con covariable

Con la finalidad de mejorar los resultados de la sección anterior, se usa el modelo descrito en la sección 3.2.4 utilizando como covariable la proporción  $X_i$ .

De forma similar que en la sección anterior, se toma como ejemplo la parroquia Cumbayá para mostrar los resultados de la simulación.

La figura 7 expone la distribución posterior junto a los histogramas los cuales son el resultado de la simulación en RStan y cuyos valores inferidos son:  $\beta^b = 0,44$  y  $\beta^w = 0,47$  con desviaciones estándar  $\sigma^b = 0,17$  y  $\sigma^w = 0,14$  y con intervalos de credibilidad de 95 % de (0,16; 0,74) y (0,17; 0,76) respectivamente.

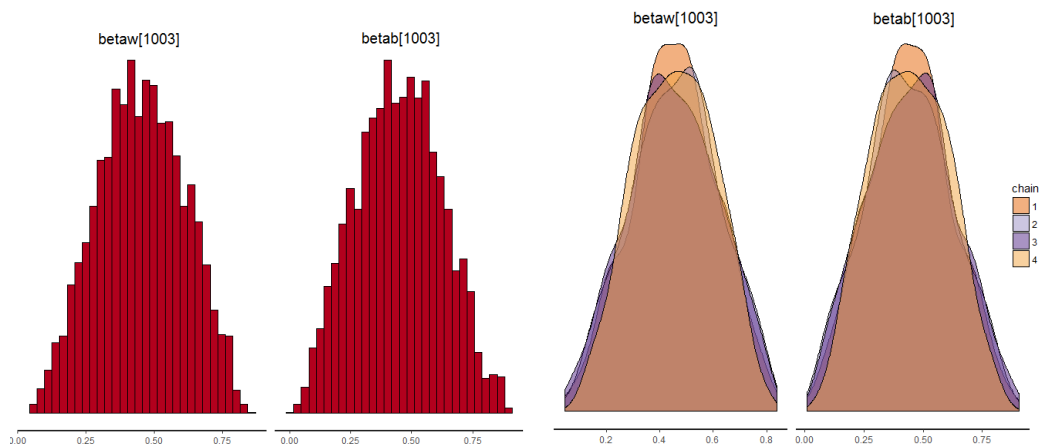


Figura 7: Histogramas de las proporciones de interés y sus respectivas densidades a posteriori con 4 cadenas de Markov, para la parroquia de Cumbayá.

De igual forma se muestran las distribuciones posteriores de los parámetros de regresión  $\alpha$ ,  $\beta$ ,  $\gamma$  y  $\delta$  en la figura 8. Los parámetros  $\beta$  y  $\delta$ , los cuales representan las pendientes de *log odds* para las poblaciones masculina y femenina respectivamente que apoyaron a AP en la parroquia de Cumbayá tienen medias  $\mu^\beta = 0,14$  y  $\mu^\delta = 0,08$  con desviaciones estándar  $\sigma^\beta = 0,1$  y  $\sigma^\delta = 0,09$  respectivamente. Lo cual indica que no hay evidencia que muestre que hay dependencia entre  $X_i$  y  $\beta^b$  así como tampoco existe dependencia entre  $1 - X_i$  y  $\beta^w$  como indica el diagrama de dispersión entre estas variables en la figura 8.

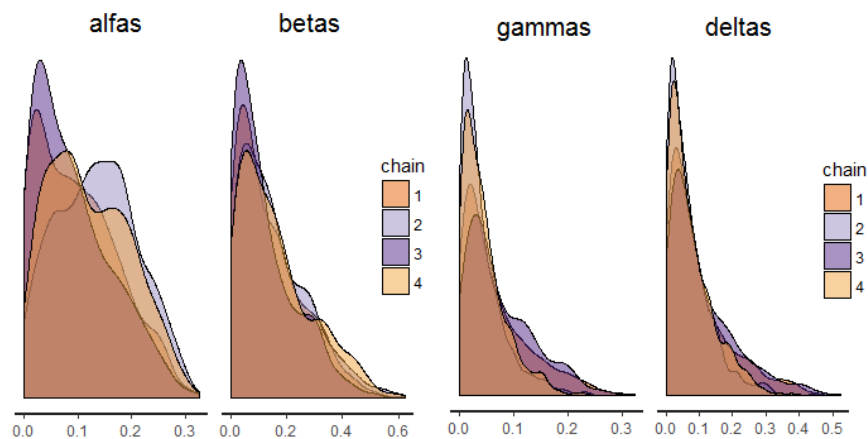


Figura 8: Distribución posterior de los parámetros de regresión para el modelo aplicado en tablas de tamaño 2x2 con el uso de covariables.

La figura 9 muestra los diagramas de correlación entre las proporciones  $\beta^b$  y  $X_i$  así como también entre  $\beta^w$  y  $1 - X_i$  y cuyos valores de correlación son: -0.24 y 0.22 respectivamente, lo cual indica que, tanto estos valores de correlación como el sesgo de agregación han disminuido respecto a los valores del modelo sin covariable, por lo que las proporciones  $\beta^b$  y  $\beta^w$  son menos segadas y por lo tanto más exactas.

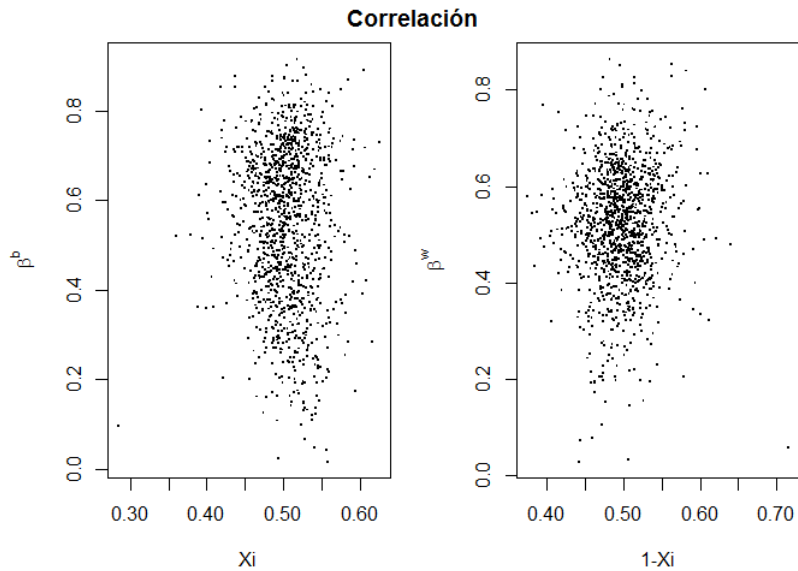


Figura 9: Correlación entre las proporciones  $X_i - \beta^b$  y  $(1-X_i) - \beta^w$ .

La figura 10 muestra los intervalos de credibilidad al 95% para cada uno de los parámetros inferidos  $\beta^b$  y  $\beta^w$ .

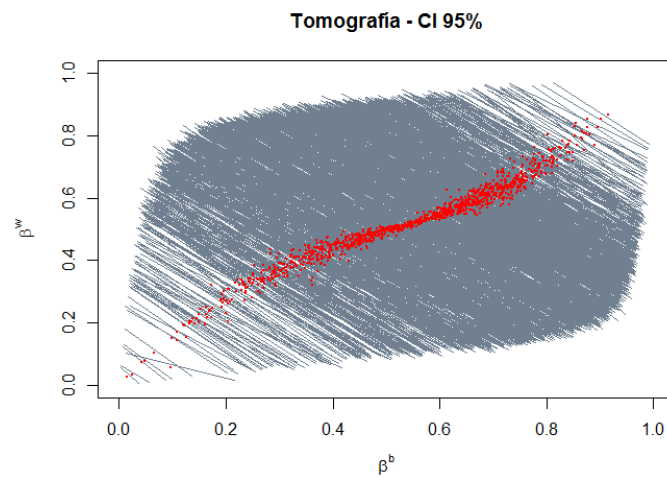


Figura 10: Intervalos de credibilidad del 95% para  $\beta^b$  y  $\beta^w$ .

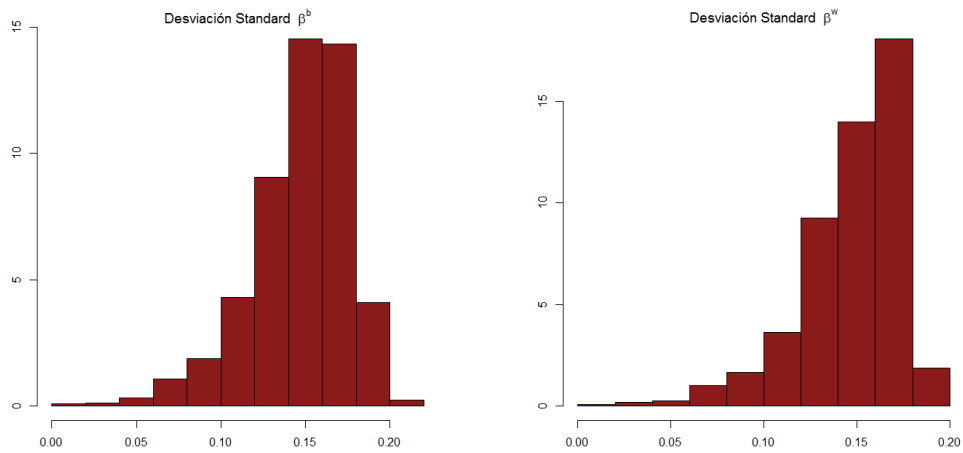


Figura 11: Histogramas de las desviaciones estándar.

La figura 11 muestra las desviaciones estándar de los parámetros inferidos  $\beta^b$  y  $\beta^w$  con  $\mu_{\beta^b} = 0,1468$  y  $\mu_{\beta^w} = 0,1473$  con desviaciones estándar  $\sigma_{\beta^b} = 0,03$  y  $\sigma_{\beta^w} = 0,026$  respectivamente.

Con la finalidad de evaluar el efecto del uso de la covariable  $X_i$  utilizada en este modelo, la figura 12 muestra el ajuste comparando las proporciones  $\beta^b$  y  $\beta^w$  inferidas con las respectivas proporciones reales obtenidas de la base de datos original.

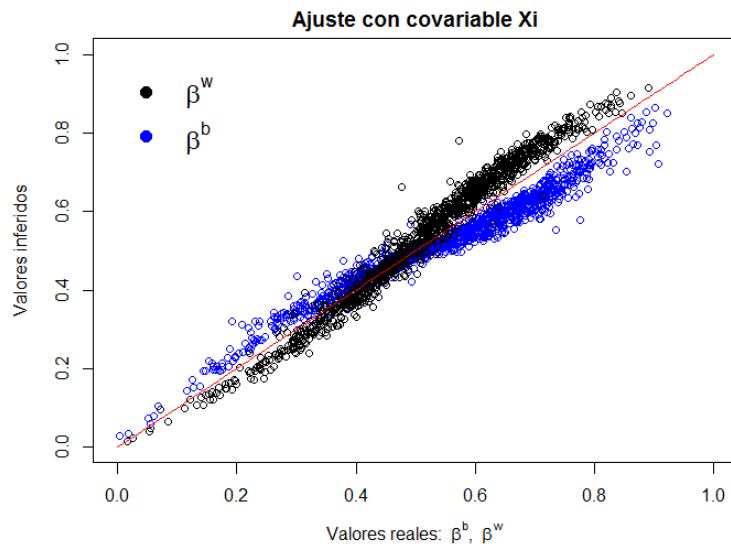


Figura 12: Ajuste entre las proporciones reales e inferidas utilizando covariables.

Finalmente la figura 13 muestra la distribución geográfica de los resultados, en la cual se ha pintado cada parroquia de acuerdo al valor del parámetro inferido, es decir, de acuerdo al porcentaje de apoyo tanto de hombres, mujeres y la diferencia entre ellos.

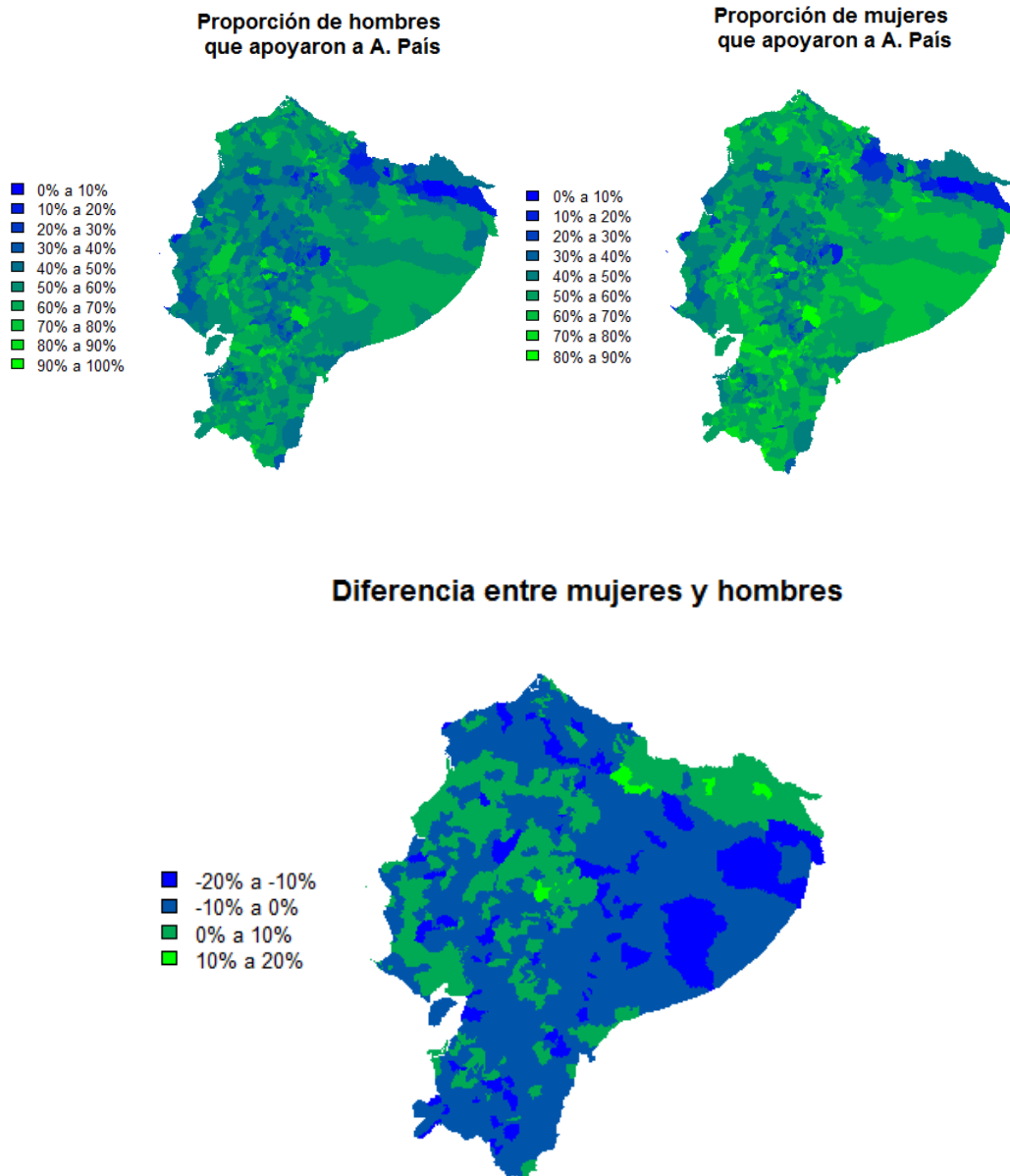


Figura 13: Mapa a nivel parroquial. Superior: Porcentaje de apoyo a AP de hombres y mujeres. Inferior: Diferencia porcentual del apoyo para AP de mujeres menos el apoyo para AP de hombres.

### 3.3.3. Resultados para tablas de contingencia de tamaño $R \times C$

Para ilustrar el modelo jerárquico bayesiano Multinomial - Dirichlet se ha utilizado la variable *edad* la cual se ha disgregado en 4 grupos de diferente edad: grupo 1 votantes entre 16 y 29 años, el grupo 2 de 30 a 44 años, grupo 3 votantes de 45 a 60 años y grupo 4 votantes mayores de 60 años. Tomando como ejemplo la parroquia Cumbayá y considerando que los subíndices  $_{11}$ ,  $_{21}$ ,  $_{31}$  y  $_{41}$  se refieren al apoyo que tuvo AP por los grupos antes mencionados, los resultados de la inferencia bayesiana son los siguientes:  $\beta_{11} = 0,45$ ;  $\beta_{21} = 0,56$ ;  $\beta_{31} = 0,49$  y  $\beta_{41} = 0,42$  con desviaciones estándar:  $\sigma_{11} = 0,11$ ;  $\sigma_{21} = 0,13$ ;  $\sigma_{31} = 0,15$ ;  $\sigma_{41} = 0,15$  respectivamente.

La figura 14 muestra las distribuciones posteriores de éstos parámetros utilizando cada una de las 4 cadenas simuladas, donde la etiqueta  $betas[i,j,k]$  representa la proporción  $\beta$  de la población que apoya al partido político  $k$ , que pertenece al grupo de edad  $j$  y de la parroquia  $i$ .

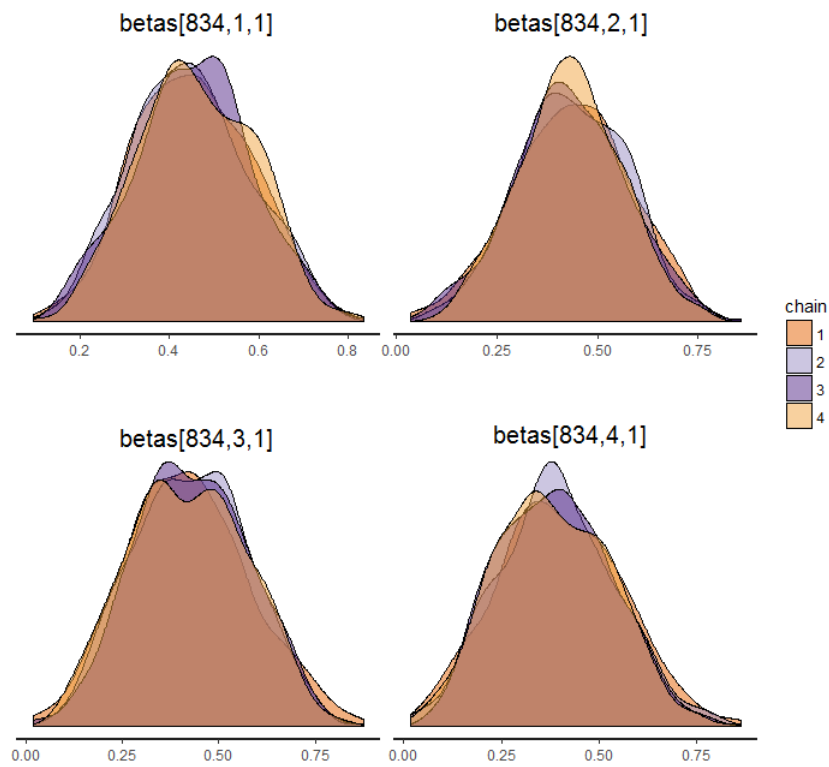


Figura 14: Densidad posterior para cada uno de los grupos de edad pertenecientes a la parroquia de Cumbayá.



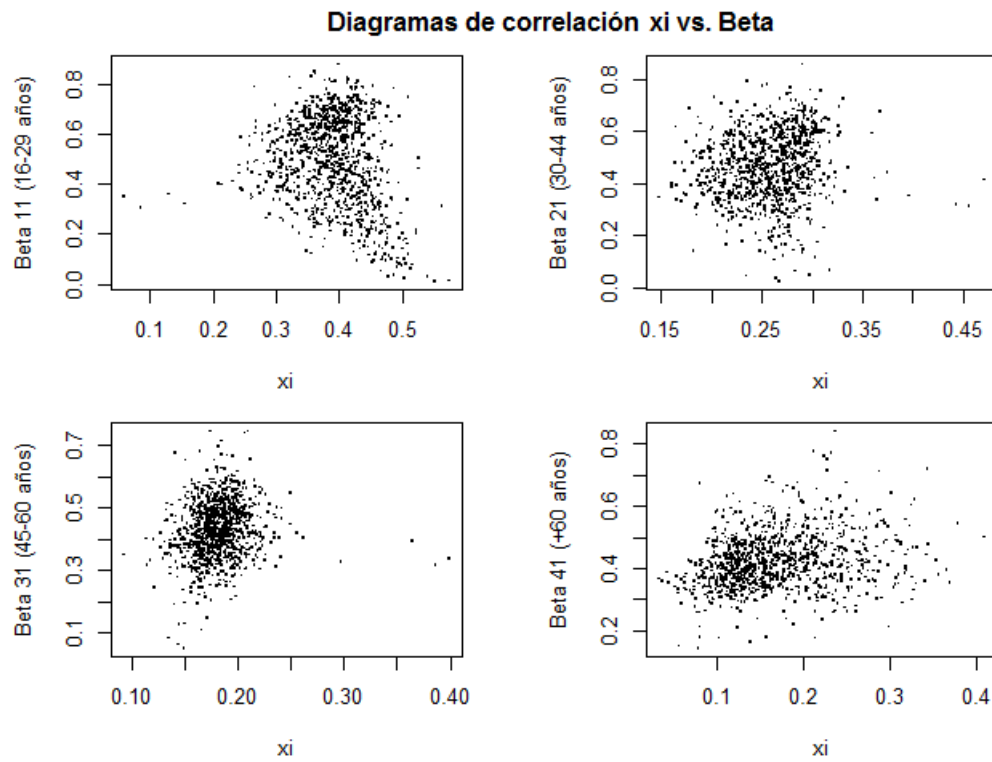


Figura 15: Correlación entre los parámetros inferidos  $\beta$ 's y las proporciones  $X_i$  para cada grupo.

En la figura 15 se muestran la correlación entre los parámetros inferidos  $\beta$ 's con la proporción total de votantes  $X_i$  en cada grupo. Los valores de estas correlaciones son: -0.65; 0.42; 0.58 y 0.32 respectivamente lo cual indica que existe sesgo debido a la agregación de los datos para los grupos 1 y 3 especialmente. Igual que en el modelo para tablas 2x2, se podría disminuir este sesgo utilizando una covariable apropiada que permita mejorar la inferencia pero esa información no está disponible.

La figura 16, muestra de forma geográfica los resultados inferidos por el modelo para la variable *edad*.

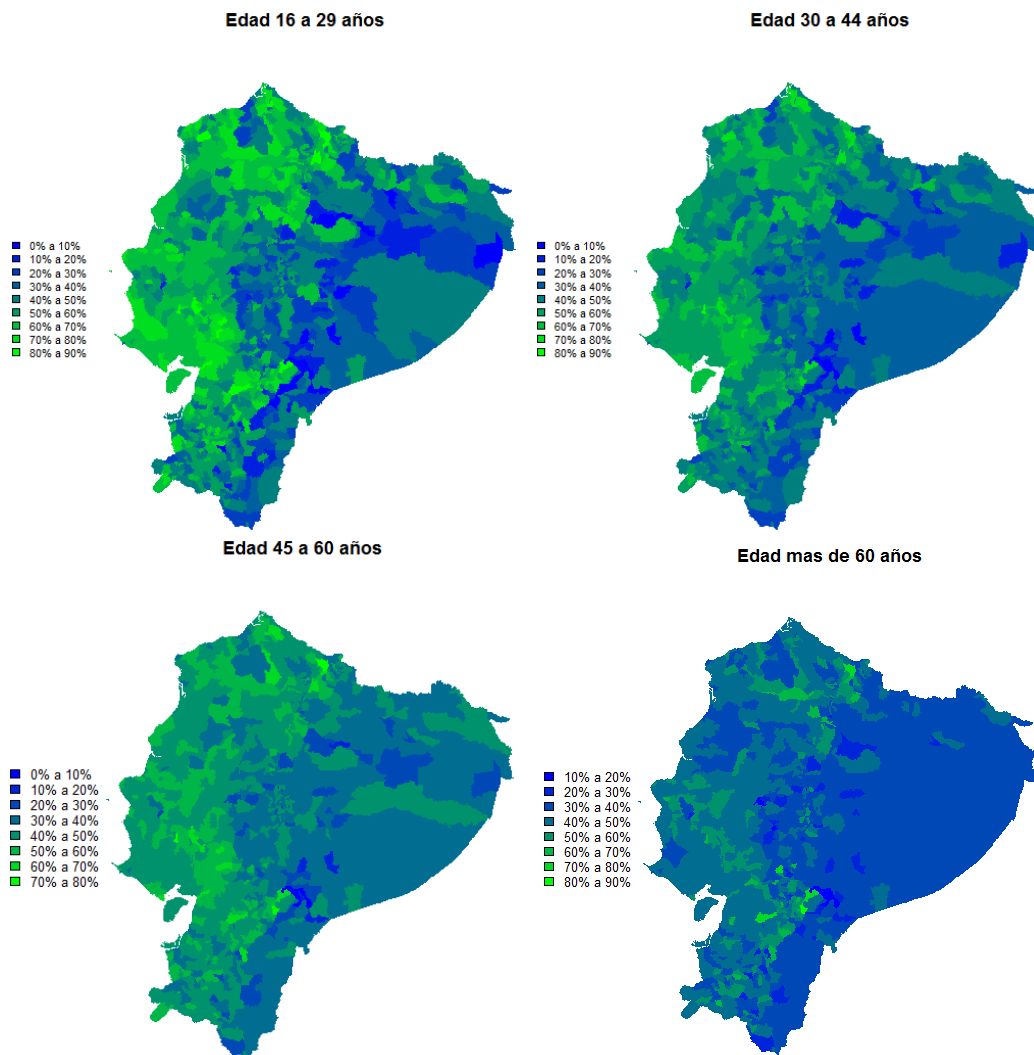


Figura 16: Mapa geográfico con las proporciones de apoyo para AP de los diferentes grupos de edad.

### 3.4. Evaluación de las simulación

Como se ha mencionado anteriormente en la sección 2.3.2, uno de los problemas inherentes a la simulación, es la inferencia mediante muestras simuladas que estén correlacionadas y que pueden existir dentro de cada cadena, es decir que la inferencia mediante muestras simuladas que están correlacionadas es menos precisa que la inferencia mediante el mismo número de muestras independientes [15]. Este problema puede evaluarse mediante un diagrama de correlación como el que se muestra en la figura 17

que corresponde al modelo aplicado en tablas de contingencia de tamaño 2x2 con la variable *sexo* y covariable  $X_i$ , el cual indica que las muestras simuladas en cada cadena tienen una baja correlación, lo que significa que la simulación está explorando de manera eficiente la distribución posterior de los parámetros.

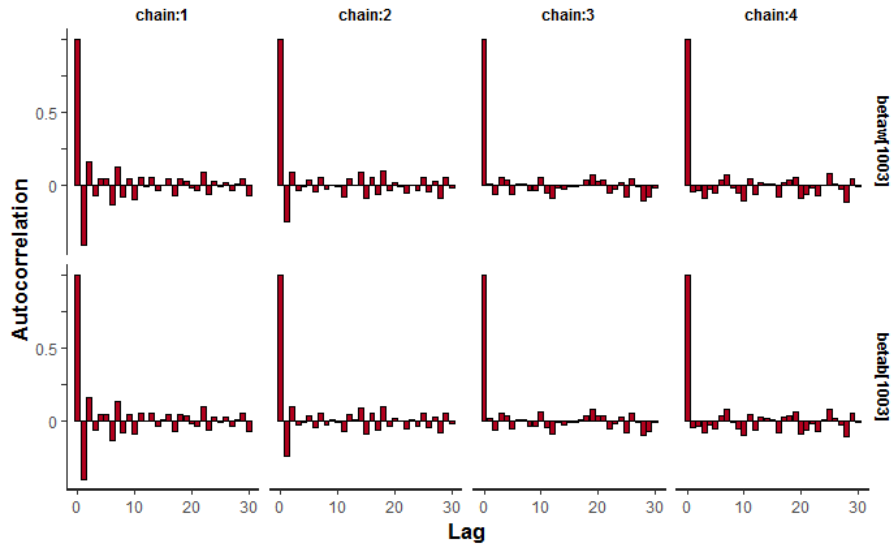


Figura 17: Diagrama de autocorrelación para cada una de las cadenas del modelo aplicado a la variable *sexo*. La parte superior muestra la correlación en las cadenas de  $\beta^w$  y la parte inferior corresponde a las cadenas de  $\beta^b$  de la parroquia Cumbayá.

Los valores  $n_{eff}$  y  $Rhat$  que se muestran en los resultados de la simulación para cada uno de los parámetros estimados sirven para evaluar la convergencia del proceso de simulación. Como se mencionó anteriormente, el valor  $n_{eff}$  indica la medida del tamaño efectivo de la muestra, el cual da un sentido de la precisión obtenida de las simulaciones [15]. En las simulaciones realizadas en éste estudio se fijó en 1000 iteraciones por cadena y con un número de cadenas igual a 4, de tal manera que se obtiene una muestra de 2000 simulaciones de la distribución posterior de los parámetros, lo cual es suficiente para alcanzar la convergencia. La figura 18 muestra la secuencia de las 4 cadenas simuladas para la parroquia Cumbayá usando la variable *sexo* y covariable  $X_i$ .

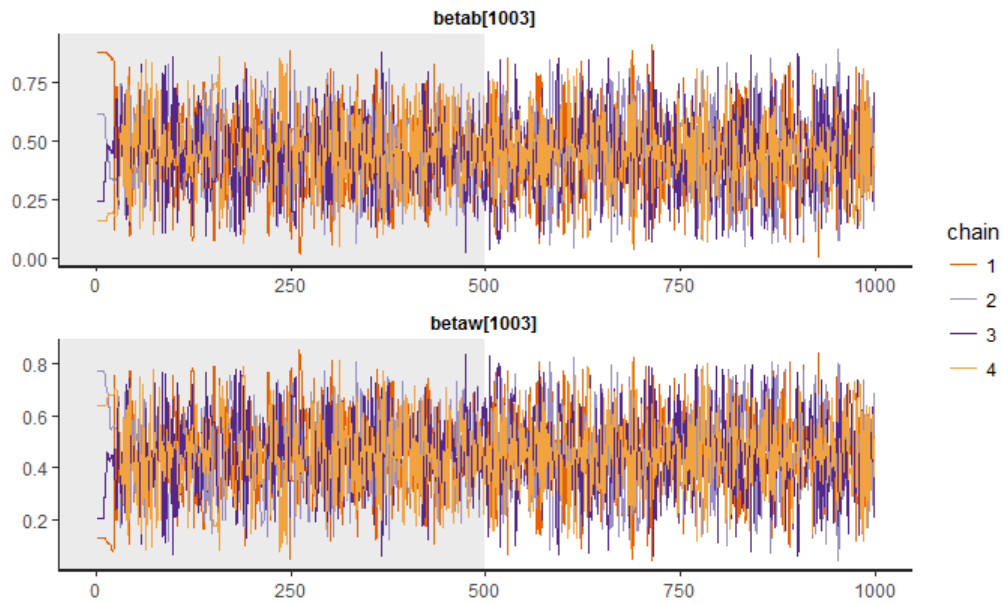


Figura 18: Traceplot de los parámetros  $\beta^b$  y  $\beta^w$  para la parroquia Cumbayá, el cual muestra que las cadenas están bien mezcladas y han alcanzado la convergencia en 1000 iteraciones.

Por otra parte se puede monitorear la convergencia de la simulación iterativa mediante  $Rhat$  (potential scale reduction factor) el mismo que indica el factor por el cual la escala (raíz cuadrada de la varianza) de la distribución de un parámetro estimado podría reducirse si las simulaciones se aproximan al infinito. Cuando la simulación alcanza la convergencia de la distribución posterior del parámetro, entonces  $Rhat=1$ , esto también indica que las cadenas simuladas han sido bien mezcladas antes de alcanzar la convergencia. Si el factor  $Rhat$  es alto, entonces lo recomendable es realizar un número mayor de simulaciones para mejorar la inferencia de la distribución del parámetro a estimar. Cabe recalcar que estos controles no son pruebas de hipótesis. No existe *valor-p* ni significancia estadística, por lo que se evalúa la discrepancia de la convergencia mediante una significancia práctica o alguna convención de la misma tal como:  $Rhat > 1.1$ . [15] En la figura 19 se muestran los resultados de los primeros parámetros de la simulación para el modelo de tablas de contingencia 2x2 con la variable *sexo* y covariable  $X_i$ .

Inference for Stan model: hombre\_mujer\_Zxi.  
 4 chains, each with iter=1000; warmup=500; thin=1;  
 post-warmup draws per chain=500, total post-warmup draws=2000.

	mean	se_mean	sd	2.5%	25%	50%	75%	97.5%	n_eff	Rhat
betab[1]	0.69	0.00	0.14	0.41	0.59	0.69	0.79	0.95	2000	1.00
betab[2]	0.56	0.00	0.17	0.24	0.44	0.56	0.68	0.87	2000	1.00
betab[3]	0.63	0.00	0.17	0.31	0.51	0.63	0.76	0.93	2000	1.01
betab[4]	0.64	0.00	0.18	0.28	0.51	0.64	0.77	0.95	2000	1.00
betab[5]	0.80	0.00	0.11	0.57	0.72	0.81	0.88	0.97	2000	1.02
betab[6]	0.85	0.00	0.08	0.68	0.79	0.86	0.92	0.98	2000	1.02
betab[7]	0.52	0.00	0.22	0.11	0.35	0.52	0.69	0.91	2000	1.00
betab[8]	0.68	0.00	0.16	0.38	0.57	0.69	0.80	0.95	2000	1.01
betab[9]	0.65	0.00	0.17	0.31	0.52	0.65	0.78	0.94	2000	1.01
betab[10]	0.61	0.00	0.18	0.26	0.48	0.62	0.75	0.94	2000	1.00
betab[11]	0.81	0.00	0.10	0.61	0.74	0.82	0.89	0.97	2000	1.02
betab[12]	0.65	0.00	0.17	0.32	0.53	0.66	0.78	0.95	2000	1.01
betab[13]	0.66	0.00	0.17	0.32	0.54	0.66	0.79	0.95	2000	1.01
betab[14]	0.58	0.00	0.19	0.19	0.44	0.59	0.73	0.92	2000	1.00
betab[15]	0.61	0.00	0.18	0.26	0.47	0.61	0.74	0.93	2000	1.00
betab[16]	0.67	0.00	0.16	0.38	0.55	0.67	0.80	0.94	2000	1.01
betab[17]	0.73	0.00	0.14	0.46	0.63	0.74	0.84	0.97	2000	1.01
betab[18]	0.57	0.00	0.18	0.21	0.43	0.58	0.71	0.90	2000	1.00
betab[19]	0.67	0.00	0.17	0.34	0.55	0.67	0.80	0.95	2000	1.01
betab[20]	0.69	0.00	0.15	0.39	0.57	0.69	0.80	0.95	2000	1.01
betab[21]	0.61	0.00	0.16	0.30	0.49	0.60	0.72	0.91	2000	1.00
betab[22]	0.68	0.00	0.16	0.36	0.56	0.68	0.80	0.95	2000	1.01
betab[23]	0.68	0.00	0.16	0.38	0.57	0.69	0.80	0.95	2000	1.01
betab[24]	0.75	0.00	0.13	0.49	0.65	0.76	0.86	0.97	2000	1.01
betab[25]	0.72	0.00	0.15	0.42	0.61	0.73	0.83	0.97	2000	1.01
betab[26]	0.65	0.00	0.17	0.31	0.52	0.65	0.78	0.95	2000	1.01
betab[27]	0.60	0.00	0.19	0.22	0.47	0.61	0.75	0.94	2000	1.00
betab[28]	0.69	0.00	0.15	0.40	0.58	0.69	0.81	0.96	2000	1.01
betab[29]	0.61	0.00	0.16	0.30	0.49	0.61	0.73	0.91	2000	1.01
betab[30]	0.58	0.00	0.19	0.21	0.45	0.59	0.72	0.92	2000	1.00
betab[31]	0.71	0.00	0.14	0.43	0.60	0.72	0.83	0.96	2000	1.01
betab[32]	0.62	0.00	0.17	0.30	0.50	0.63	0.75	0.93	2000	1.00
betab[33]	0.77	0.00	0.12	0.53	0.69	0.78	0.87	0.97	2000	1.01
betab[34]	0.69	0.00	0.16	0.40	0.58	0.70	0.81	0.97	2000	1.01
betab[35]	0.78	0.00	0.12	0.54	0.69	0.78	0.87	0.97	2000	1.01

Figura 19: Resultado numérico de la simulación con 4 cadenas de 1000 iteraciones cada una, para el modelo *Beta-Binomial* aplicado en tablas de contingencia de tamaño 2x2

La tabla anterior muestra en la primera columna el nombre del parámetro  $\beta^b$  junto al cual se observa la media estimada para dicho parámetro. La cuarta columna muestra la desviación estándar de la inferencia, las siguientes muestran diferentes cuantiles y las dos últimas columnas muestran los valores  $n_{eff}$  y  $Rhat$  mencionados anteriormente. Como se puede observar,  $Rhat$  es menor que 1.1 lo que significa que no existe divergencia en las simulaciones para cada uno de los parámetros estimados.

Estos valores de  $Rhat$  y  $n_{eff}$  se resumen para todos los parámetros estimados utilizando histogramas los cuales se muestran en la figura 19.

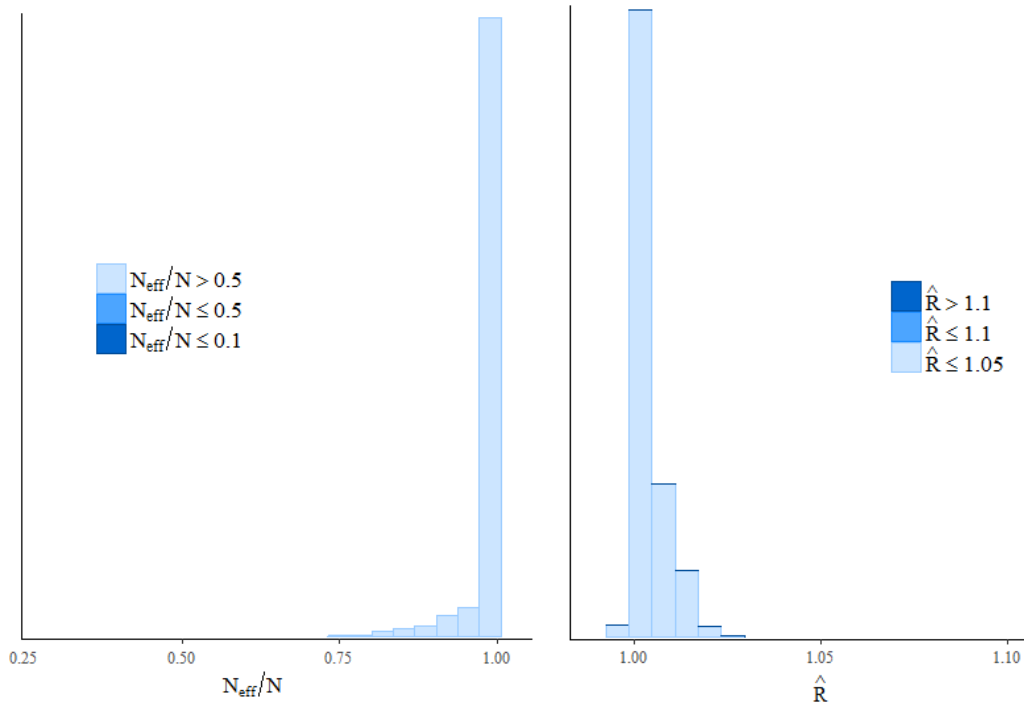


Figura 20: Izquierda: histograma de la razón entre  $n_{eff}$  y  $N$ . Derecha: histograma para el valor  $\hat{R}$  correspondiente al modelo Beta-Binomial con el uso de covariables.

El tamaño efectivo de la muestra  $n_{eff}$  es un estimado del número de muestras independientes de la distribución posterior de un parámetro estimado. Ya que las muestras dentro de una cadena de Markov no son independientes, si es que existe correlación, se espera que  $n_{eff}$  sea menor que  $N$  para cada parámetro estimado. Mientras mas cercana a 1 sea la razón  $n_{eff}/N$  entonces se tiene un mejor tamaño de  $n_{eff}$  y por lo tanto un mayor número de muestras independientes en la simulación. En este modelo se realizó la simulación fijando un número de iteraciones igual a 1000 con 4 cadenas y 500 iteraciones de precalentamiento, lo que da como resultado un valor de  $N = 2000$  muestras, por lo que el primer histograma de la figura 20 indica que las muestras de la distribución posterior no tienen correlación. Los valores 0.1 y 0.5 de este histograma no tienen un significado teórico en particular pero de forma heurística hay que preocuparse si la razón  $n_{eff}/N$  es menor que 0.1 [28].

Como se mencionó antes, el valor  $\hat{R}$  mide la razón entre la varianza promedio de las muestras dentro de cada cadena y la varianza de las muestras mezcladas a través de las

cadenas, si todas las cadenas están en equilibrio, éstas prácticamente serían las mismas y  $\hat{R}$  sería igual a 1. Si las cadenas no han convergido a una distribución común,  $\hat{R}$  será mayor que 1. En el histograma de la derecha de la figura 18 se observa que  $\hat{R}$  es menor que 1.05 para todos los parámetros estimados, lo que significa que la simulación ha sido eficiente [28].

### 3.5. Resultados del segundo método

El modelo multinomial penalizado fue ajustado utilizando la librería **glmnet** en **R** como en [32], la cual ajusta modelos *lasso*, *ridge* y *elastic-net*, específicamente se realizó la regresión utilizando la función *cv.glmnet* la cual realiza el ajuste utilizando *cross-validation* con  $k = 10$  *folds*.

Los parámetros de ajuste para la función *cv.glmnet* son:  $x$  y  $y$  corresponden a las matrices de variables y respuesta respectivamente, el parámetro *family* indica el tipo de respuesta que en este caso es multinomial, el parámetro *alpha* ( $0 \leq \alpha \leq 1$ ) define el tipo de proceso que se usará para seleccionar las variables que dan el menor error (*Residual Sum Square RSS*) [33], si  $\alpha = 0$  se tiene el proceso *ridge*, si  $\alpha = 1$  el proceso es *lasso* y si ( $0 < \alpha < 1$ ) entonces el proceso es *elastic-net*, el parámetro *parallel*, si está fijado como *TRUE*, entonces la ejecución del comando se hará en múltiples procesadores o cores.

En primera instancia, se ejecuta el comando para obtener el parámetro  $\alpha$  mas adecuado de tal manera que se minimice el error de la regresión. En este caso se obtuvo  $\alpha = 0$  (*ridge*) con un error promedio mínimo  $cvm = 0,119$  donde:  $cvm = \text{cross-validated mean}$ .

La figura 21 muestra la curva del proceso *cross-validation* junto con las curvas superior e inferior de la desviación estándar en función de los valores  $\lambda$  utilizados. Como se puede observar, la mínima desviación se da cuando  $\lambda = 0,011$ , por otra parte el valor  $\lambda_{1se} = 0,083$  es el valor  $\lambda$  mas grande de tal manera que el error está dentro de 1 error estándar a partir del mínimo. En la parte superior de la gráfica se muestra el número de variables que se utilizan para la regresión a lo largo de los valores  $\lambda$  utilizados. Ya que se usó un modelo *ridge* el número de variables se mantienen pero el valor de los coeficientes de regresión se reducen mediante la penalización.

A partir del modelo ajustado y utilizando la función *predict* de la misma librería *glmnet* se obtuvo la predicción del apoyo a los tres partidos AP, CREO y Otros, en cada una de las parroquias, las proporciones que resultan de la predicción se muestran en el apéndice 4.

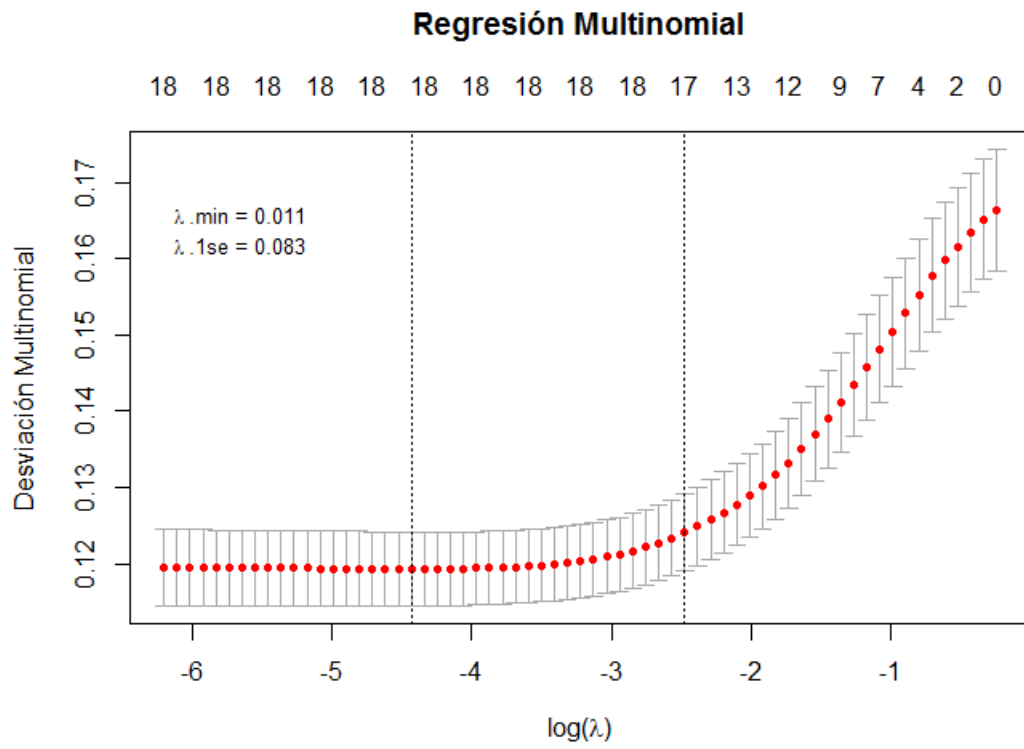


Figura 21: Desviación estándar en función de Lambda



## 4. Conclusiones

Éste trabajo integra dos ramas de la estadística: la inferencia bayesiana con modelos jerárquicos que resuelven el problema de la inferencia ecológica y la estadística computacional que básicamente se refiere a los métodos MCMC implementados en el lenguaje Stan. Al finalizar éste estudio se tienen las siguientes conclusiones:

1. Varios enfoques se han desarrollado para la solución del problema de la inferencia ecológica, en éste trabajo se ha tomado un enfoque bayesiano. A diferencia de los métodos estadísticos frecuentistas, los métodos bayesianos se basan en la formulación de un conjunto de distribuciones previas para los parámetros desconocidos que se basan en *creencias* a priori del investigador . Tales distribuciones previas son parte del modelo estadístico así como la parte que expresa la distribución de probabilidad de las observaciones dadas a través del cálculo de su verosimilitud.
2. Para el cálculo de las distribuciones a posteriori se ha recurrido a los métodos MCMC, los cuales se basan en el diseño de una adecuada cadena de Markov que bajo ciertas condiciones convergen a una densidad estacionaria invariante en el tiempo que coincida con la densidad que se quiere muestrear. Las cadenas de Markov son procesos estocásticos en el que la probabilidad de que ocurra un evento depende del evento inmediatamente anterior. Las técnicas MCMC pueden ser aplicadas en teoría a cualquier densidad de probabilidad (las densidades en los modelos jerárquicos bayesianos suelen ser muy complejas), ésto las hace particularmente interesantes aplicables en varios campos como la optimización estocástica y otros campos diferentes al muestreo.
3. Una desventaja de los métodos MCMC es que producen números aleatorios correlacionados entre sí. Éste problema se puede solucionar si se usan algoritmos más especializados, en el caso de Stan, usa el algoritmo Hamiltoniano Monte Carlo que permite al proceso ser más eficiente en cuanto al problema de la correlación y a la rapidez de convergencia. Éste caso especial de los métodos MCMC se apoya en una variable adicional que le permite encontrar más rápidamente la dirección de máxima probabilidad, evitando así el paseo aleatorio en la cadena de Markov.
4. El sesgo de agregación afecta a la inferencia ecológica en el sentido de confundir las relaciones entre las variables involucradas. En nuestro caso, si las variables

$X_i$  y  $(\beta^b, \beta^w)$  están correlacionadas, entonces los parámetros de interés resultarían sesgados. Al realizar el análisis de correlación para las clases hombres-mujeres se obtuvo una correlación entre  $X_i$  y  $\beta^w$  de -0.64, el cual no es despreciable. Una posible solución a la correlación entre éstos parámetros es modelar dicha dependencia, haciendo que los parámetros  $(\beta^b, \beta^w)$  varíen como funciones lineales de  $X_i$  cuyos resultados se muestran en la sección 3.3.2. verificando que efectivamente el uso de covariables adecuadas mejora la inferencia de los parámetros. En este modelo el resultado de la correlación obtenida fue de -0.22 el cual es mucho menor que -0.64. En éste trabajo no se contó con mucha información que pueda ser usada como covariables, salvo en el modelo de tablas de tamaño 2x2 para las variables PEA y PEI, cuyos resultados se muestran en el apéndice 2, en las cuales usamos como covariable la información de si los votantes cuentan o no con algún tipo de seguro social.

5. Los parámetros estimados en el modelo Beta- Binomial y Multinomial-Dirichlet para las variables *hombre – mujer* y *edad* presentan desviaciones estándar altas, lo cual podría ser una señal de falta de información que permita realizar una inferencia más exacta de los parámetros estimados, tal información puede estar presente en una o varias covariables adecuadas para lograr tal objetivo. Por otra parte, como se indica en [8], se puede presentar un problema cuando existe escasa variabilidad en la proporción  $X_i$ , esto significa que las líneas en la tomografía son casi paralelas como se puede ver en la figura 2, porque las pendientes de éstas líneas son calculadas mediante  $-X_i/(1 - X_i)$ . Cuánto menor es la variación en  $X_i$ , mayor será la imprecisión asociada a los resultados del modelo. Otro de los problemas que se deben considerar es que se puede presentar valores bajos de  $X_i$  en todos los recintos, esto se traduce en que las líneas en la tomografía de la figura 2 serían prácticamente horizontales permitiendo a  $\beta^b$  tomar todos los valores posibles en el rango  $[0,1]$ .
6. Este trabajo podría ampliarse y profundizarse si se tuviese mayor información respecto a variables demográficas que estén cruzadas con las bases de datos del CNE, de tal manera que se puedan emplear los modelos aquí mostrados para realizar inferencias mejoradas con covariables para un grupo más grande de variables de estudio como por ejemplo votantes de diferentes nacionalidades, con diferentes

niveles de ingresos económicos, diferente nivel de estudios, etc.

7. Éste trabajo logró su objetivo al inferir las proporciones de apoyo para un candidato dado a nivel de recinto (parroquia). Éstos resultados podrían ser de gran interés para el marketing político ya que con éstos se podría conocer los niveles de apoyo por regiones. Ésto permitiría diseñar de mejor manera los planes de propaganda política ya que aún tomando solo encuestas antes de elecciones se podrían inferir las cantidades de interés.

## Apéndice 1: Script para los modelos jerárquicos bayesianos en Stan

```

data{
  int <lower=0> N;
  real <lower=0> xi[N];
  real <lower=0> unomenosxi[N];
  int <lower=0> Ti[N];
  int <lower=0> ni[N];
}
parameters{
  real <lower=0,upper=1> betab[N];
  real <lower=0,upper=1> betaw[N];
  real <lower=0> cb;
  real <lower=0> db;
  real <lower=0> cw;
  real <lower=0> dw;
}
transformed parameters{
  real <lower=0> theta[N];
  for (n in 1:N){
    theta[n]=xi[n]*betab[n]+unomenosxi[n]*betaw[n];
  }
}
model{
  cb~exponential(0.5);
  db~exponential(0.5);
  cw~exponential(0.5);
  dw~exponential(0.5);
  for(n in 1:N){
    betab[n]~beta(cb,db);
    betaw[n]~beta(cw,dw);
  }
  for (n in 1:N){
    Ti[n]~binomial(ni[n],theta[n]);
  }
}

```

Figura 22: Modelo Jerárquico bayesiano Beta Binomial para tablas de contingencia de tamaño 2x2 sin covariables.

```

data{
  int N;
  vector[N] xi;
  vector[N] unomenosxi;
  int Ti[N];
  int ni[N];
  vector[N] Zi;
}
parameters{
  vector <lower=0,upper=1>[N] betab;
  vector <lower=0,upper=1>[N] betaw;
  real <lower=0.01> db;
  real <lower=0.01> dw;
  real <lower=0> alfas;
  real <lower=0> betas;
  real <lower=0> gammas;
  real <lower=0> deltas;
}
transformed parameters{
  vector <lower=0,upper=1>[N] theta;
  theta=xi .*betab+unomenosxi .*betaw;
}
model{
  db~exponential(0.5);
  dw~exponential(0.5);
  betab~beta(db *exp(alfas+betas *Zi),db);
  betaw~beta(dw *exp(gammas+deltas *Zi),dw);
  for (n in 1:N){
    Ti[n]~binomial(ni[n],theta[n]);
  }
}

```

Figura 23: Modelo Jerárquico bayesiano Beta Binomial para tablas de contingencia de tamaño 2x2 con el uso de covariables.

```

data{
  int N;          //numero de observaciones
  int R;          //número de grupos de edades
  int C;          //categorías de clasificación
  int fcorrea[N]; //personas que apoyan a correa
  int flasso[N];  //personas que apoyan a lasso
  int otros[N];   //personas que apoyan a otros
  int ni[N];      //total de votantes
  real g1[N];     //núm. personas entre 16 y 29
  real g2[N];     //núm. personas entre 29 y 44
  real g3[N];     //núm. personas entre 45 y 60
  real g4[N];     //núm. personas entre 60 y mas
}
transformed data{
  int Ti[N,3];
  for (n in 1:N){
    Ti[n,1]=fcorrea[n];
    Ti[n,2]=flasso[n];
    Ti[n,3]=otros[n];
  }
}
parameters{
  simplex [C] betas[N,R];
  vector <lower=0.01> [C] alfas[N,R];
}
transformed parameters{
  simplex[C] theta[N];
  for (n in 1:N){
    theta[n,1]=betas[n,1,1]*g1[n]+betas[n,2,1]
*g2[n]+betas[n,3,1]*g3[n]+betas[n,4,1]*g4[n];
    theta[n,2]=betas[n,1,2]*g1[n]+betas[n,2,2]
*g2[n]+betas[n,3,2]*g3[n]+betas[n,4,2]*g4[n];
    theta[n,3]=betas[n,1,3]*g1[n]+betas[n,2,3]
*g2[n]+betas[n,3,3]*g3[n]+betas[n,4,3]*g4[n];
  }
}
model{
  for (n in 1:N){
    for (i in 1:R){
      for (j in 1:C){
        alfas[n,i,j]~gamma(4,2);
      }
    }
  }
  for (n in 1:N){
    for(i in 1:R){
      betas[n,i]~dirichlet(alfas[n,i]);
    }
  }
  for (n in 1:N){
    Ti[n]~multinomial(theta[n]);
  }
}

```

Figura 24: Modelo Jerárquico bayesiano Multinomial - Dirichlet para casos RxC sin covariable usado para la variable *edad*.

## Apéndice 2: Resultados de las simulaciones para cada modelo

### Modelo Beta - Binomial usando covariable

En este apéndice se muestran los resultados del modelo Beta-Binomial para las variables PEA y PEI usando como covariable la población que tiene acceso al seguro social como el IESS, ISSFA, Seguro Social Campesino, etc.

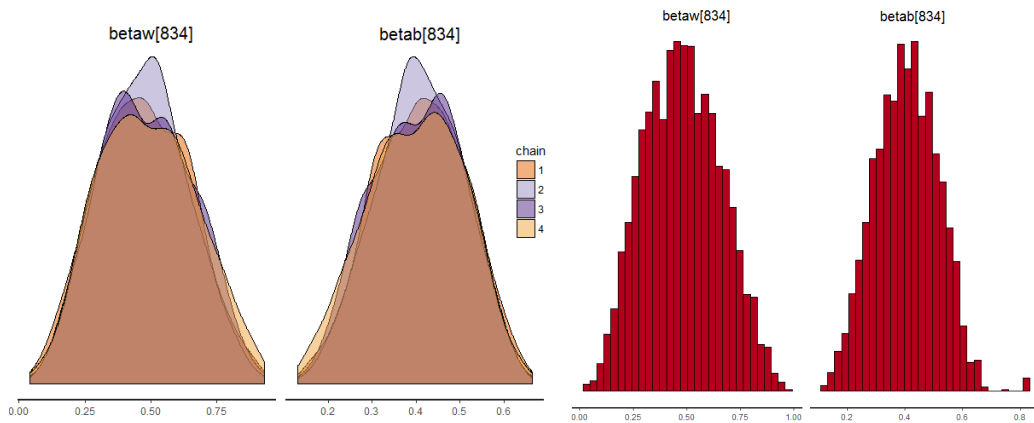


Figura 25: Densidad posterior e histogramas de los parámetros  $\beta^b$  y  $\beta^w$  que corresponden a la variable de población económicamente activa e inactiva usando covariable.

La proporción  $\beta^b$  corresponde a la proporción de votantes con condición *PEA* y  $\beta^w$  indica la proporción de votantes *PEI* que apoyaron a AP. Los resultados de éste modelo y particularmente para la parroquia Cumbayá son:  $\beta^b = 0,4$ ;  $\beta^w = 0,47$  con desviaciones estándar  $\sigma^b = 0,1$  y  $\sigma^w = 0,17$  como se muestra en la figura 25.

De igual manera se muestran en la figura 26 las distribuciones posteriores de los parámetros de regresión  $\alpha$ ,  $\beta$ ,  $\gamma$  y  $\delta$ . El parámetro  $\beta$ , que indica la pendiente de regresión de *log odds* para la población económicamente activa PEA, tiene un valor medio de 0.008 lo cual indica que no hay evidencia para asegurar que existe un efecto de regresión, lo mismo sucede con el parámetro  $\delta$ , pendiente de la regresión para el *log odds* para la PEI, cuya media es 0.009 sugiriendo nuevamente que no hay evidencia de que exista dependencia entre  $X_i$  y  $\beta^b$  así como tampoco entre  $1 - X_i$  y  $\beta^w$  [24] como muestra la figura 27.

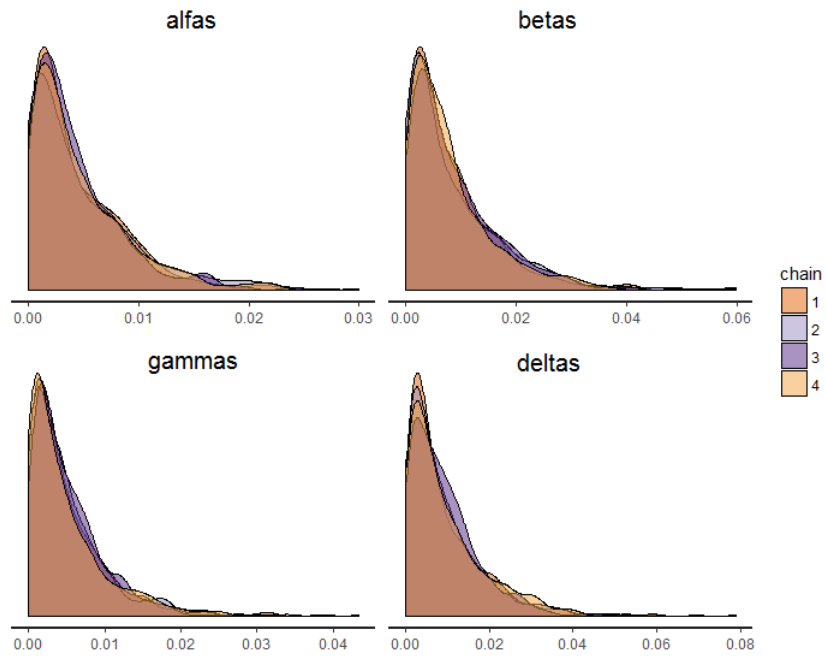


Figura 26: Distribución posterior de los parámetros de regresión para el modelo 2x2 con covariable.

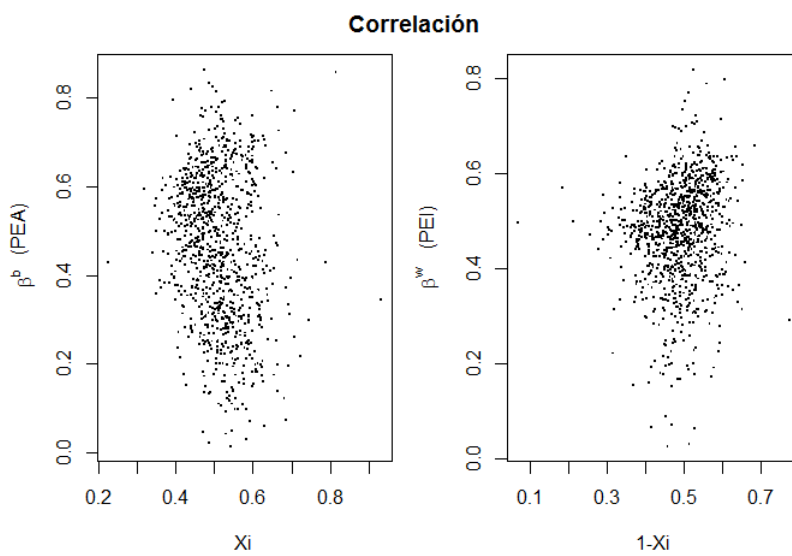


Figura 27: Diagramas de dispersión para las variables PEA y PEI.



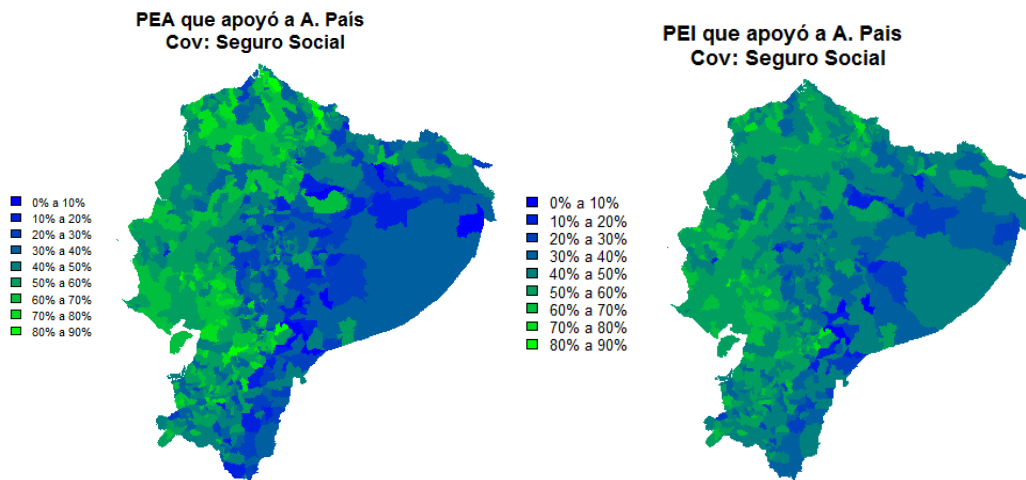


Figura 28: Mapa geográfico con las proporciones de población económicamente activa e inactiva usando covariable.

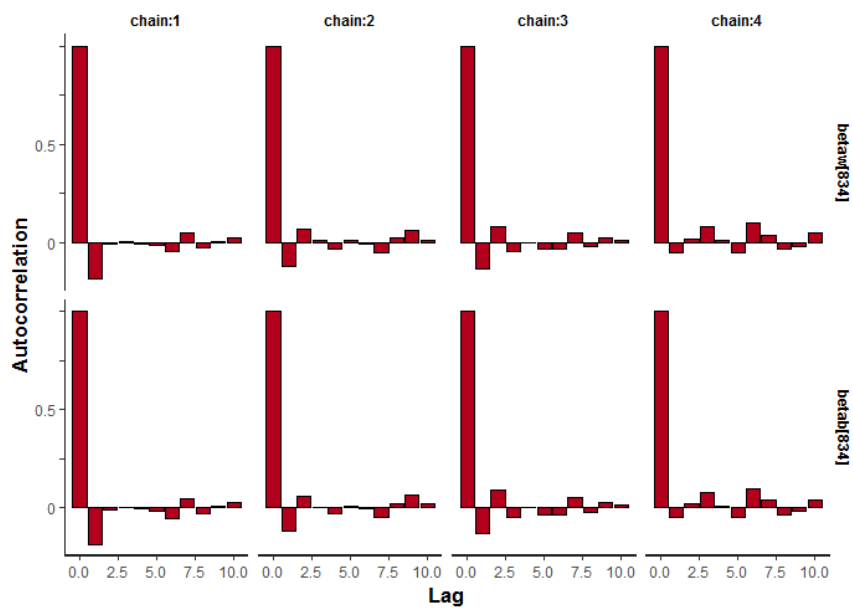


Figura 29: Diagrama de autocorrelación para los parámetros  $\beta^b$  y  $\beta^w$  de la variable *Población económicamente activa e inactiva*.

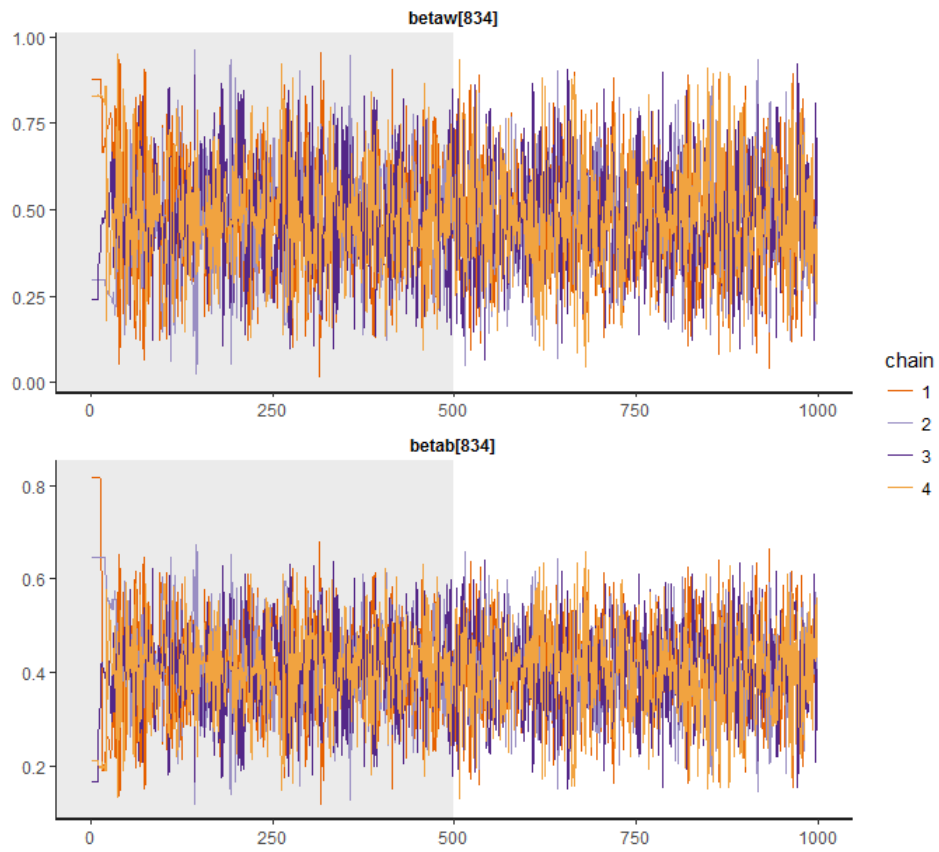


Figura 30: *Traceplot* de los parámetros  $\beta^b$  y  $\beta^w$  para las variables PEA y PEI.

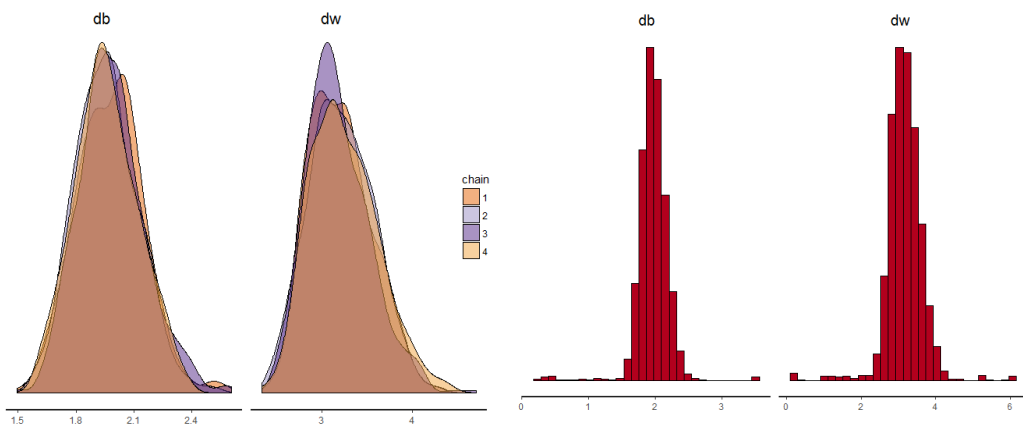


Figura 31: Densidad e histograma de los hiperparámetros  $d_b$  y  $d_w$  de la distribución *Beta*.

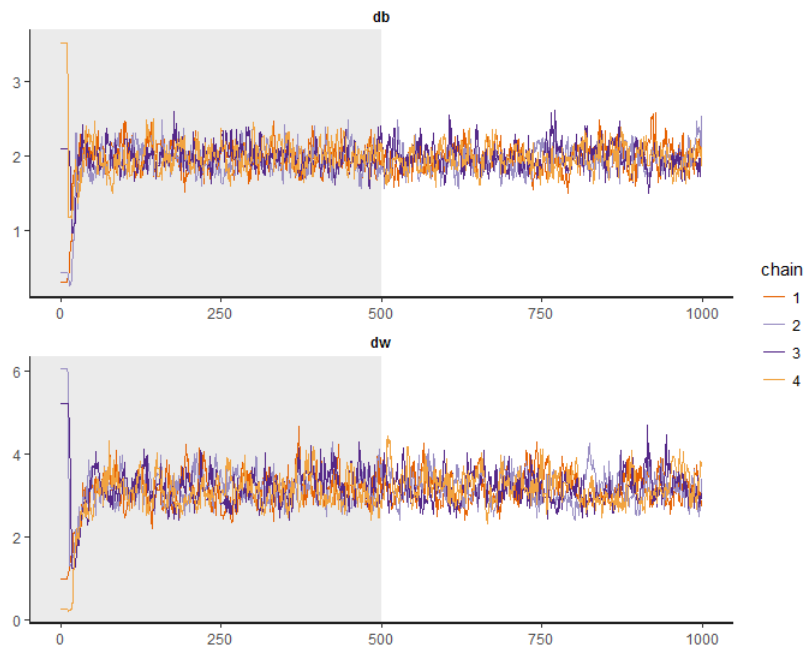


Figura 32: *Tracepot* de los parámetros  $d_b$  y  $d_w$  de la distribución *Beta*.

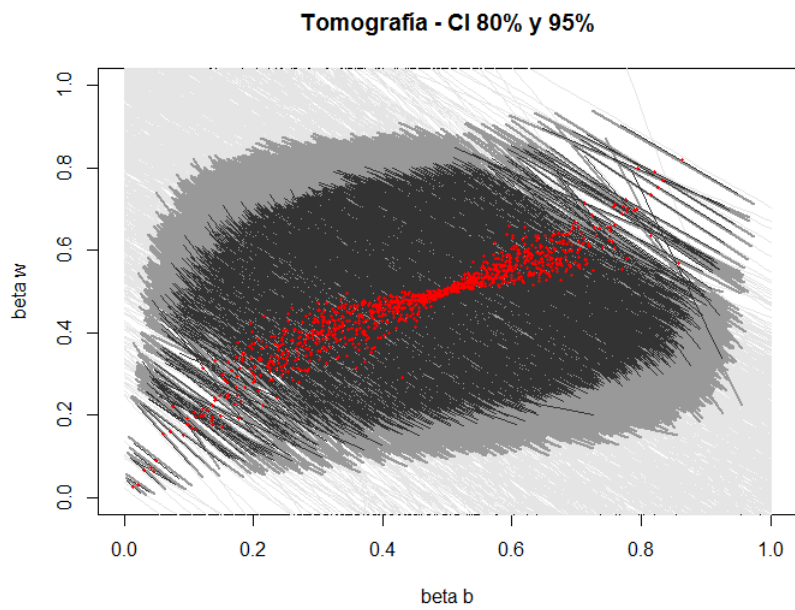


Figura 33: Diagrama de los intervalos de credibilidad al 80% y al 95% para los parámetros  $\beta^b$  y  $\beta^w$ .

Inference for Stan model: pea\_Z\_seguro.  
 4 chains, each with iter=1000; warmup=500; thin=1;  
 post-warmup draws per chain=500, total post-warmup draws=2000.

	mean	se_mean	sd	2.5%	25%	50%	75%	97.5%	n_eff	Rhat
betab[1]	0.62	0.00	0.12	0.39	0.53	0.61	0.70	0.85	2000	1.00
betab[2]	0.48	0.00	0.14	0.22	0.38	0.48	0.58	0.74	2000	1.00
betab[3]	0.55	0.00	0.14	0.30	0.45	0.55	0.66	0.81	2000	1.00
betab[4]	0.48	0.00	0.11	0.26	0.40	0.48	0.56	0.69	2000	1.00
betab[5]	0.67	0.00	0.14	0.39	0.56	0.67	0.77	0.93	2000	1.00
betab[6]	0.82	0.00	0.08	0.67	0.75	0.82	0.88	0.96	2000	1.00
betab[7]	0.41	0.00	0.14	0.15	0.31	0.41	0.50	0.67	2000	1.00
betab[8]	0.64	0.00	0.13	0.39	0.54	0.64	0.74	0.89	2000	1.00
betab[9]	0.66	0.00	0.13	0.41	0.56	0.66	0.77	0.91	2000	1.01
betab[10]	0.60	0.00	0.16	0.29	0.48	0.60	0.72	0.90	2000	1.00

Figura 34: Resultado de la simulación de los 10 primeros parámetros  $\beta^b$  que muestran la convergencia y la mezcla adecuada de las cadenas simuladas.

### Modelo Multinomial - Dirichlet sin covariable

Las siguientes figuras muestran los resultados de la simulación para la variable *edad* evidenciando que el proceso de simulación ha sido eficiente.

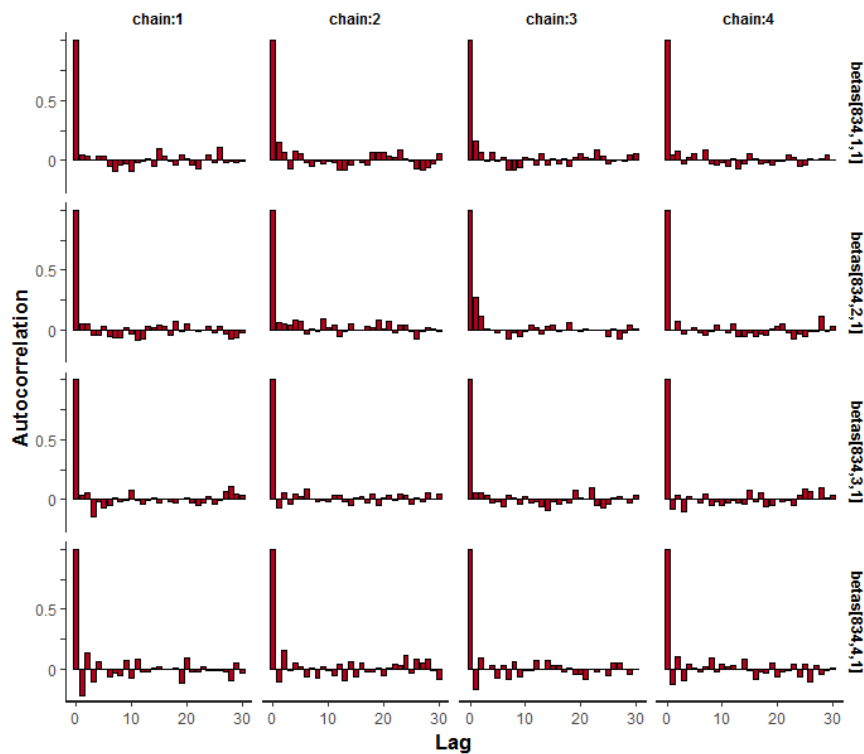


Figura 35: Diagrama de autocorrelación de las simulaciones para cada cadena de los parámetros de cada grupo de edad.

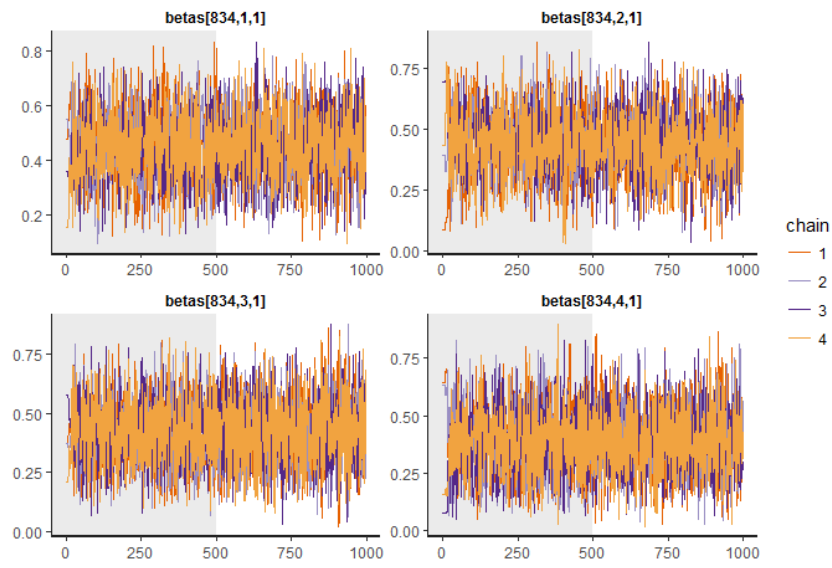


Figura 36: *Traceplot* de las cadenas simuladas para cada parámetro de los grupos de la variable *edad*.

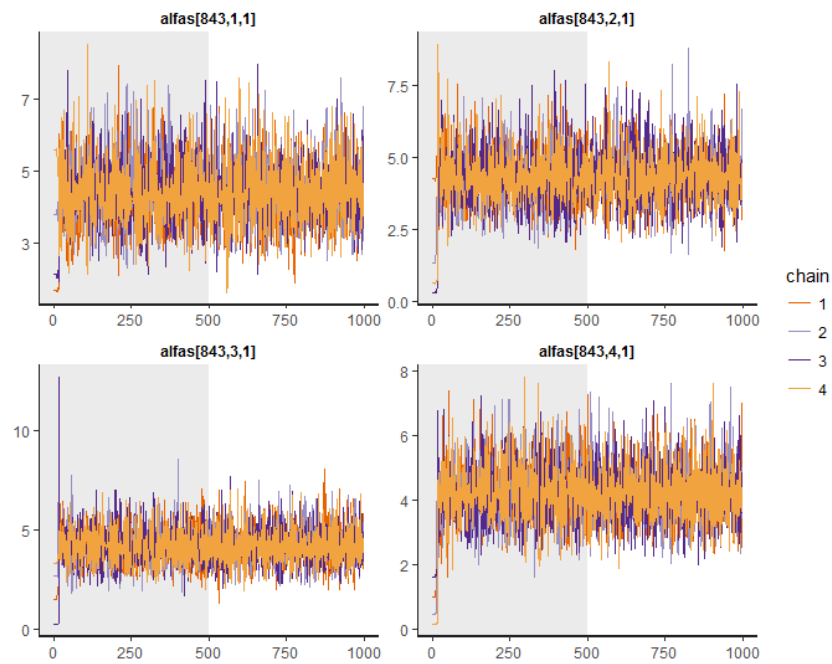


Figura 37: *Traceplot* de los hiperparámetros  $\alpha$  de la distribución *Dirichlet* de la variable *edad*.

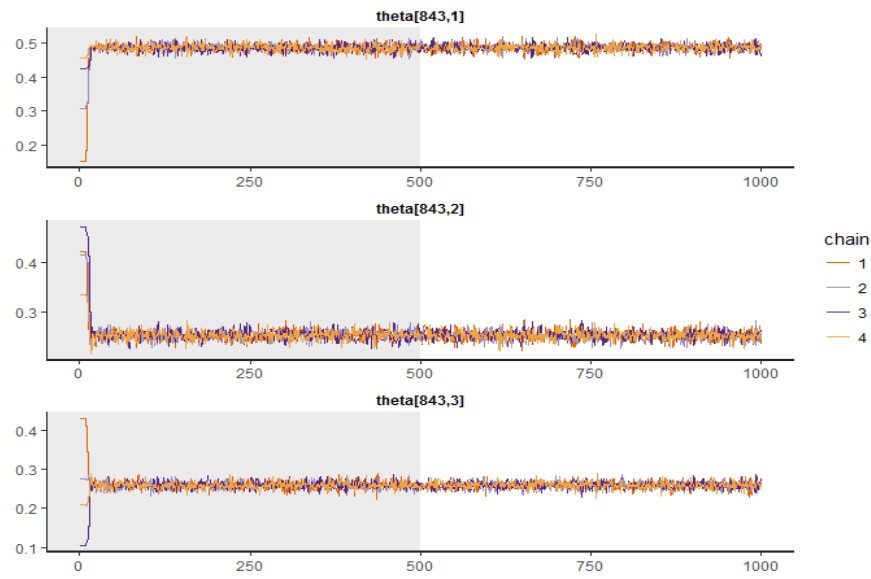


Figura 38: *Traceplot* de las cadenas simuladas para los parámetros  $\theta$  de la distribución *Multinomial* de la variable *edad*.

Inference for Stan model: edad.

4 chains, each with iter=1000; warmup=500; thin=1;

post-warmup draws per chain=500, total post-warmup draws=2000.

	mean	se_mean	sd	2.5%	25%	50%	75%	97.5%	n_eff	Rhat
betas[1,1,1]	0.67	0.00	0.09	0.51	0.61	0.67	0.74	0.84	1302	1.00
betas[1,1,2]	0.12	0.00	0.05	0.03	0.08	0.12	0.16	0.22	1628	1.00
betas[1,1,3]	0.20	0.00	0.07	0.06	0.15	0.20	0.26	0.35	1212	1.00
betas[1,2,1]	0.55	0.00	0.13	0.29	0.46	0.56	0.64	0.79	1313	1.00
betas[1,2,2]	0.18	0.00	0.08	0.04	0.12	0.17	0.23	0.34	1636	1.00
betas[1,2,3]	0.27	0.00	0.11	0.08	0.19	0.26	0.34	0.51	1155	1.00
betas[1,3,1]	0.43	0.00	0.15	0.15	0.32	0.43	0.54	0.73	2000	1.00
betas[1,3,2]	0.25	0.00	0.12	0.06	0.16	0.24	0.33	0.52	2000	1.00
betas[1,3,3]	0.32	0.00	0.14	0.09	0.22	0.31	0.41	0.60	2000	1.00
betas[1,4,1]	0.39	0.00	0.15	0.11	0.27	0.39	0.50	0.70	2000	1.00
betas[1,4,2]	0.29	0.00	0.13	0.08	0.19	0.27	0.38	0.57	2000	1.00
betas[1,4,3]	0.32	0.00	0.14	0.08	0.21	0.31	0.42	0.63	2000	1.00

Figura 39: Resultados de la simulación de los parámetros para la parroquia 1 para cada uno de los grupos de la variable *edad*.

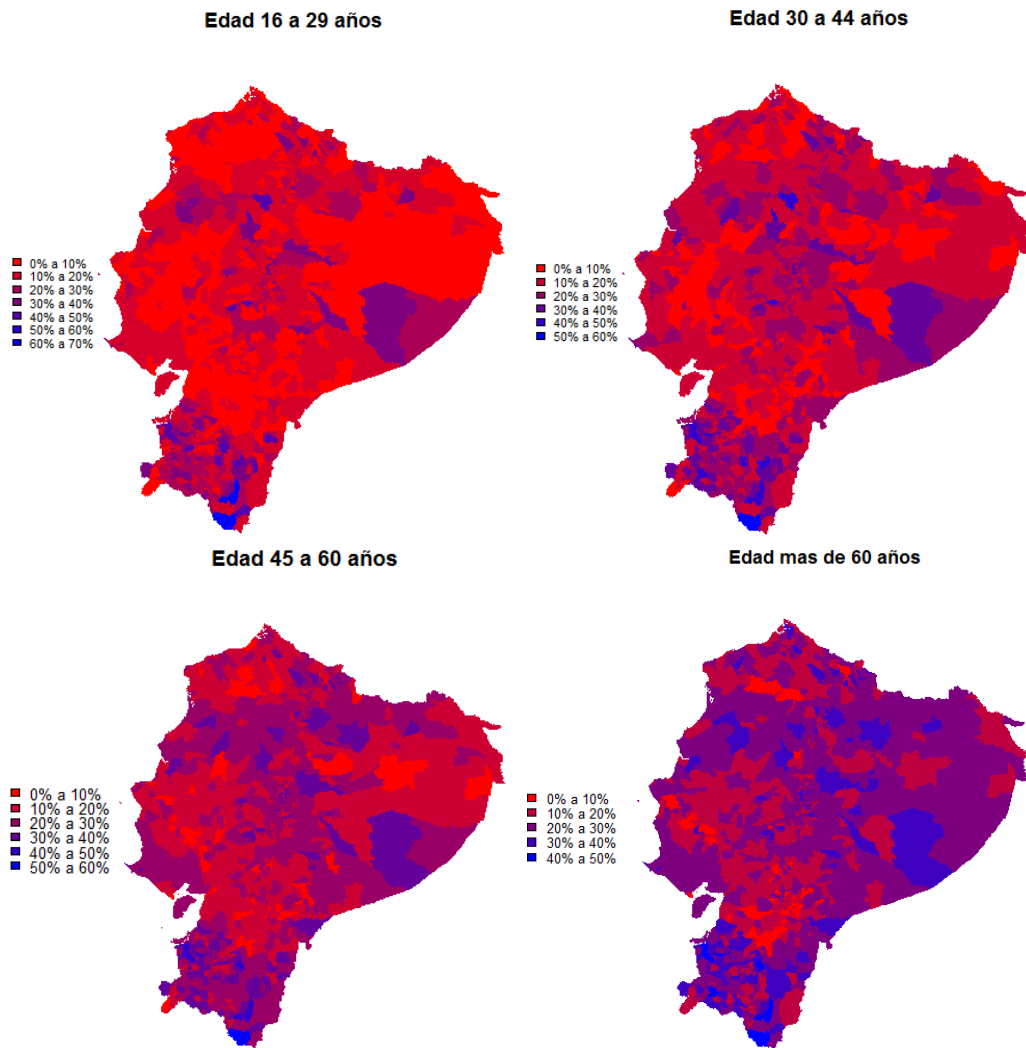


Figura 40: Mapa geográfico con las proporciones de apoyo de los diferentes grupos de edad para el partido político CREO.

### Apéndice 3: Gráfica y Demostración de que la distribución Beta es conjugada con la distribución Binomial

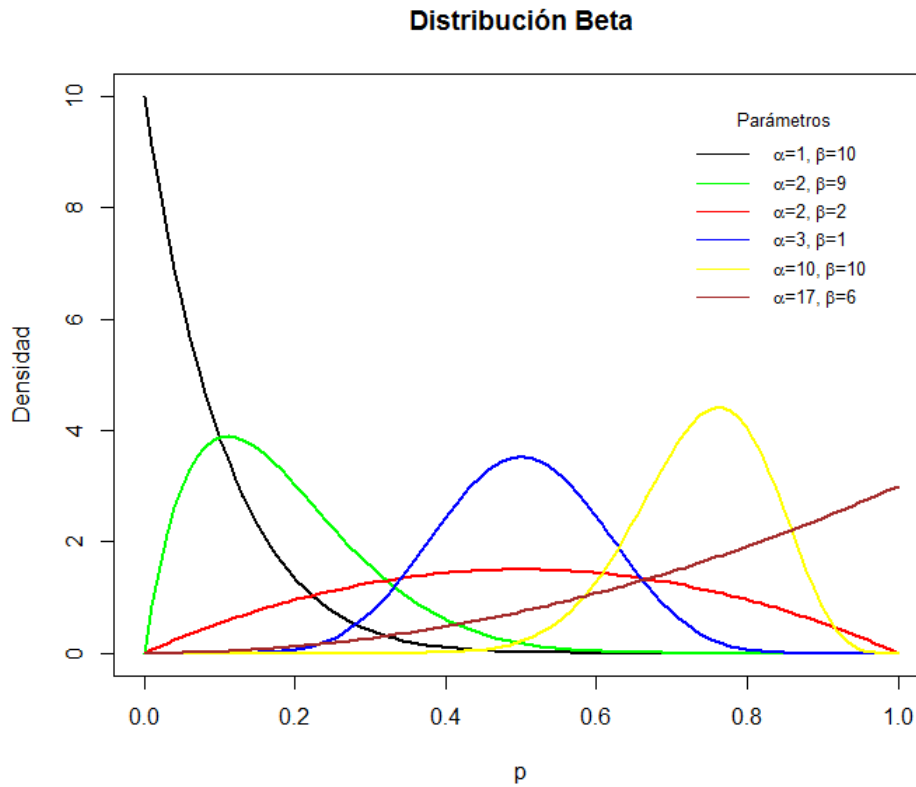


Figura 41: Distribución Beta con distintos parámetros.

En teoría Bayesiana, si la distribución posterior  $p(\Theta|x)$  pertenece a la misma familia que la distribución de probabilidad a priori  $p(\Theta)$ , entonces las distribuciones a priori y posteriori se denominan *conjugadas* [15].

Si la distribución de verosimilitud de  $y$  corresponde a una distribución *Binomial* cuyo parámetro  $\theta$  sigue una distribución a priori *Beta*, entonces:



$$p(y|\theta) = \binom{n}{y} \theta^y (1-\theta)^{n-y}$$
$$p(\theta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1} (1-\theta)^{\beta-1}$$

Y la distribución a posteriori  $p(\theta|y)$  sería:

$$p(\theta|y) \propto \theta^y (1-\theta)^{n-y} \times \theta^{\alpha-1} (1-\theta)^{\beta-1}$$
$$= \theta^{y+\alpha-1} (1-\theta)^{n-y+\beta-1}$$
$$= \text{Beta}(y + \alpha, n - y + \beta)$$

## Apéndice 4: Resultados de la regresión multinomial

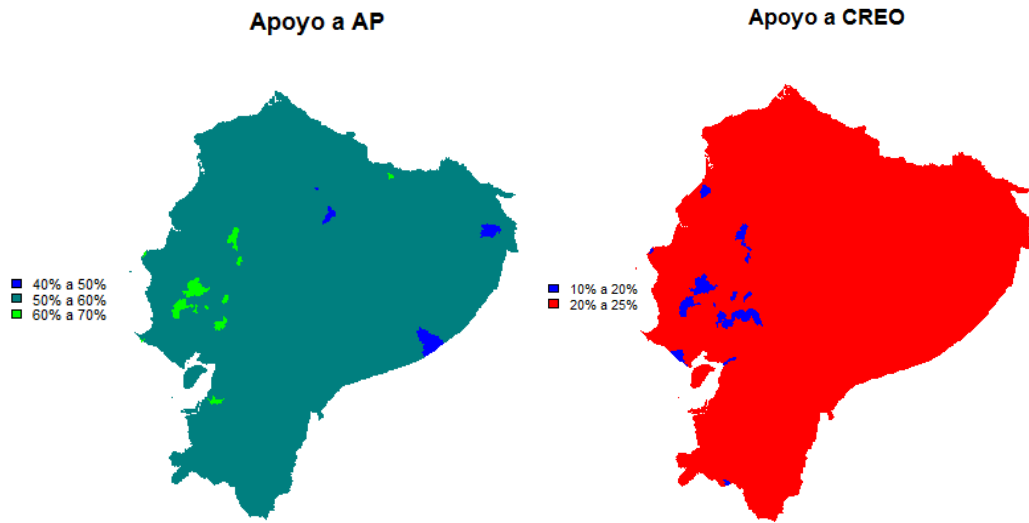


Figura 42: Proporciones de apoyo de cada parroquia utilizando el método multinomial.

## Referencias

- [1] Muñoz F. (2013). La Línea de Fuego Pensamiento Crítico: ELECCIONES FEBRERO 2013. Recuperado de <https://lalineadefuego.info/2013/03/01/elecciones-febrero-2013-un-analisis-por-francisco-munoz-jaramillo/>
- [2] Barredo D., Silva R., Martens Cheryl.(2015). El milagro ecuatoriano: pautas para entender el Ecuador del siglo XXI. <https://www.academia.edu>
- [3] Consejo Nacional Electoral del Ecuador CNE. Elecciones 2013 17 de Febrero (2017, Febrero 12) <http://resultados2013.cne.gob.ec>.
- [4] Seijas F. (2015). Los Exit Polls como Herramientas de Control y Confiabilidad de las Cifras Oficiales del Proceso Electoral. Recuperado de <http://gumilla.org/biblioteca/bases/biblo/texto/COM200613488-95.pdf>.
- [5] King, G. (1997). A Solution to the Ecological Inference Problem: Reconstructing Individual Behavior from Aggregate Data. Princeton University Press. Retrieved from <http://www.jstor.org/stable/j.ctt46n43p>
- [6] Morgan K., Ecological Inference from Goodman to King , Division of Humanities and Social Sciences California Institute of Technology, vol. 34, no. 3,p 101-123 (2001). ISSN 0161-5440.
- [7] King G., Rosen O., Tanner M., Ecological Inference New Methodological Strategies. Cambridge University Press, Cambridge, First Edition (2004) ISBN 0-521-83513-5
- [8] Balsa J. (2002) Ventajas y Limitaciones de la Metodología de Inferencia Ecológica propuesta por G. King. Aplicaciones al análisis del triunfo de Perón en las elecciones presidenciales argentinas de 1946. Cinta moebio 13: 33-57. [www.moebio.uchile.cl/13/balsa.htm](http://www.moebio.uchile.cl/13/balsa.htm)
- [9] Duncan, Otis Dudley y Beverly Davis (1953). An Alternative to Ecological Correlation, American Sociological Review, 18; p.p. 665-666.
- [10] Goodman, Leo. (1953). Ecological Regressions and the Behavior of Individuals, American Sociological Review, 18: 663-666.

- [11] Boudon R. (1963). "Propiedades individuales y propiedades colectivas; un problema de análisis ecológico", en R. Boudon y P. Lazarsfeld, Metodología de las Ciencias Sociales. Volúmen II. Barcelona, Laia, 1974; p.p. 247-284.
- [12] Goldberger A. A course in econometrics, Harvard University Press, London (1991) p 58. ISBN 0-674-17544-1
- [13] Castela E., Villardón M.(2007). Inferencia ecológica para la caracterización de abstencionistas: Caso de Portugal. Departamento de Estadística, Universidad de Salamanca.
- [14] Plescia C., De Sio L.(2017). An evaluation of the performance and suitability of RxC methods for ecological inference with known true values. Springer. DOI: 10.1007/s11135-017-0481-z
- [15] Gelman A., Carlin J., Stern H., Dunson D., Vehtari A., Rubin D.(2014), Bayesian Data Analysis, Florida USA, CRC Press Taylor y Francis Group.
- [16] Lee P., (2012), Bayesian Statistics: An Introduction, The Atrium United Kingdom, WILEY, ISBN: 9781118332573
- [17] Proyecciones Poblacionales por Parroquia. (2017, Enero 15). <http://www.ecuadorencifras.gob.ec/proyecciones-poblacionales>
- [18] Metropolis, N. and Ulam, S. (1949). The Monte Carlo method. Journal of the American Statistical Association, 44(247):335–341. xvii, 452
- [19] Sheldon Ross (2010). A first course in probability, University of Southern California, Eighth Edition, Prentice Hall.
- [20] Stan Modeling Language User's Guide and Reference Manual Stan Development Team Stan Version 2.9.0. (2017,Abril 6) <http://mc-stan.org/documentation/>
- [21] Olivia L., Ryan T. M. y Michael K. (2015). eiPack: Ecological Inference and Higher - Dimension Data Managment. CRAN Repository.
- [22] Ori R., Wenxin J., Gary K. y Martin A. T. (2000). Bayesian and Frequentist Inference for Ecological Inference: the RxC Case.

- [23] Damodar N. Gujarati (2003). *Basic Econometrics*, United States Military Academy, West Point, McGraw - Hill.
- [24] Gary K., Ori R. y Martin A. T. (1999). Binomial-Beta Hierarchical Models for Ecological Inference. *Sociological Methods Research*, Vol. 28 No. 1, 61-90.
- [25] Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, M., and Teller, E. (1953). Equations of state calculations by fast computing machines. *Journal of Chemical Physics*, 21:1087–1092. 23, 24, 453
- [26] Gelman, A. and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, 7(4):457–472. 24, 454
- [27] Carpenter B., Gelman A., Hoffman M., Lee D., Goodrich B., Betancourt M., Brubaker M., Guo J., Li P. y Riddell A. (2017). Stan: A Probabilistic Programming Language. *Journal of Statistical Software*, 76(1), 1-15.
- [28] Betancourt M. y Girolami M. (2013). Hamiltonian Monte Carlo for Hierarchical Models. Cornell University Library. arXiv:1312.0906
- [29] Betancourt M. y Stein L. (2011). The Geometry of Hamiltonian Monte Carlo. Cornell University Library. arXiv:1112.4118
- [30] Pascual del Olmo V. (2011). Métodos avanzados de muestreo MCMC (Tesis de pregrado). Universidad Carlos III. Madrid, España.
- [31] Selvin H. (1958). Durkheim's Suicide and Problems of Empirical Research. *American Journal of Sociology*.63(6), 607-619.
- [32] Flaxman S., Wang Y. y Smola A. (2015). Who Supported Obama in 2012? Ecological Inference through Distribution Regression. KDD '15 Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 289-298. doi>10.1145/2783258.2783300.
- [33] Hastie T., Tibshirani R. y Friedman J. (2008) *The Elements of Statistical Learning*. Data Mining, Inference and Prediction. Second Edition. Springer Series.
- [34] Simon N., Friedman J. y Hastie T. (2013) A Blockwise Descent Algorithm for Group-penalized Multiresponse and Multinomial Regression.