

UNIVERSIDAD AUTÓNOMA METROPOLITANA

MAESTRÍA EN CIENCIAS (MATEMÁTICAS APLICADAS E INDUSTRIALES)

Propuesta de un modelo bayesiano para la descripción
de datos circulares longitudinales

PRESENTA:

EMILIANO GENEYRO SQUARZON

DIRECTOR DE TESIS:

GABRIEL NÚÑEZ ANTONIO

2016

Contenido

Introducción	3
1. Preliminares	7
1.1. Datos direccionales	7
1.1.1. Introducción	7
1.1.2. Estadística descriptiva	9
1.2. Modelos para datos circulares	14
1.3. El modelo Normal Proyectado	18
1.3.1. La normal proyectada $NP(\boldsymbol{\mu}, \boldsymbol{\Lambda} = \begin{pmatrix} 1 & 0 \\ 0 & \lambda \end{pmatrix})$	19
2. Estadística Bayesiana	24
2.1. Métodos de Monte Carlo vía cadenas de Markov	27
3. Modelo longitudinal para datos circulares	36
3.1. Introducción a los estudios longitudinales	36
3.2. Metodologías para el análisis de datos longitudinales	39
3.3. Modelos de efectos mixtos	43
3.4. Propuesta de modelo para datos longitudinales circulares	50
3.5. Análisis descriptivo del modelo	53
4. Inferencia bayesiana	63
4.1. Densidades condicionales completas	63
4.2. Implementación	72
5. Aplicación del modelo	76
Conclusiones	100
6. Anexo A: Programas	102

Introducción

En el ámbito estadístico y en diversas ramas del conocimiento, es de gran interés realizar estudios relacionados a mediciones de direcciones. Los datos direccionales son observaciones de vectores unitarios en un espacio q -dimensional. Debido a la complejidad en el análisis estadístico de este tipo de datos, el desarrollo de la teoría sufrió un relativo estancamiento. Sin embargo en los últimos años, debido al desarrollo computacional y la creación de software estadístico más eficiente, el desarrollo de métodos estadísticos para el análisis de este tipo de datos ha tenido un nuevo auge. Esta nueva oleada de aportaciones van desde métodos de graficación de este tipo de datos, herramientas para realizar un análisis descriptivo, hasta propuestas de modelos estadístico para explicar y realizar inferencia sobre observaciones de tipo direccional. A modo de ejemplo de estas aportaciones, el lector puede consultar Fisher (1993), Presnell et al. (1998), Mardia y Jupp (2000), D’Elia et al. (2001), Jammalamadaka y SenGupta (2001), Nuñez Antonio y Gutiérrez-Peña (2005, 2010 y 2014) y Song (2007).

Este trabajo se enfoca en los datos direccionales en el plano 2-dimensional, denominados datos circulares. Dichos vectores unitarios en el espacio 2-dimensional pueden ser representados por un ángulo, comunmente **denota** con θ . Así pues, se debe notar que el espacio muestral de este tipo de datos es el círculo unitario y debido a la topología específica de estos, es necesario utilizar métodos estadísticos específicos que consideren dicha estructura topológica del espacio muestral. Una de las características principales **],** por las cuales los métodos univariados y multivariados estándares pueden no ser apropiados, es la periodicidad que poseen los datos circulares **].**

En general, los modelos propuestos se pueden agrupar en tres categorías principales. El primero de ellos son los modelos “wrapped” o envueltos, donde

se destacan la distribución Normal envuelta, Cauchy envuelta y la Poisson envuelta. Estos son construidos a partir de envolver distribuciones escalares alrededor del círculo unitario. La segunda categoría son los modelos von Mises-Fisher, cuya distribución principal es la distribución von Mises-Fisher, y en particular para el caso de direcciones en el círculo unitario está distribuida es conocida como von Mises. La distribución von Mises es una de las más utilizadas en el análisis inferencial de este tipo de datos. Por último, se tienen distribuciones generadas por proyecciones, dentro de las cuales la más importante es la Normal proyectada. Esta distribución se genera a partir de la proyección radial de un vector Normal bivariado. Una característica importante de la Normal proyectada es su versatilidad ya que con ésta es posible modelar comportamientos simétricos, asimétricos, unimodales y multimodales (véase, por ejemplo, Nuñez Antonio, 2010).

A pesar de que el interés por el estudio de datos circulares se ha reactivado, el enfoque de datos longitudinales con variable de respuesta circular ha sido poco explorado. En el caso de variables de respuesta de tipo escalar existen diversos modelos propuestos, entre los que se destacan los modelos de efectos mixtos, los modelos de ecuaciones estimadoras generalizadas y los modelos de patrones de covarianzas.

La principal aportación en el contexto de datos longitudinales circulares es el modelo propuesto por Nuñez Antonio y Gutiérrez-Peña (2014), en el cual se proponen un modelo basado en la Normal proyectada. Los autores proponen para cada observación de cada individuo se establezca una distribución Normal proyectada y por lo tanto se pueda obtener un vector Normal bivariado, mediante la introducción de una variable latente. Además proponen que para cada componente del vector Normal bivariado generado una especificación formada por un modelo de efectos mixtos. La principal característica de esta propuesta es que permite no solo la introducción de covariables, sino también la separación de variabilidad entre sujetos y la variabilidad intrínseca de cada individuo.

En su artículo, Nuñez Antonio y Gutiérrez-Peña (2014) proponen que la variabilidad del error para el vector bivariado este determinada por la matriz identidad. Es decir, que el error de medición para cada componente del vector bivariado y para cada observación se distribuya como $N(0, 1)$. Los autores realizan inferencia bayesiana completa en diversos ejemplos, para lo

cual utilizan métodos de Monte Carlo vía cadenas de Markov.

Este trabajo de investigación pretende ampliar dicho modelo incluyendo un parámetro adicional de variabilidad en la segunda componente del vector Normal bivariado. El objetivo principal es obtener un modelo estadísticamente más flexible para analizar diversos patrones de comportamiento longitudinal, al reducir la restricción sobre la variabilidad de los errores de medición.

La estructura de este trabajo es la siguiente. En el Capítulo 1 presenta una introducción a los datos circulares y su modelación. El objetivo de dicho capítulo es que el lector se familiarice con las características principales de los datos angulares y sus diferencias con los datos de tipo escalar. También se busca establecer el tratamiento estadístico especial que requieren los datos circulares y dar un panorama general de las distintas herramientas necesarias para su análisis. Por último se presentan al lector, de manera general, los principales enfoques desarrollados para su modelación; haciendo especial hincapié en el desarrollo del modelo Normal proyectado.

Debido a que este trabajo se desarrolla desde el enfoque bayesiano de la estadística, el Capítulo 2 presenta una breve introducción a los conceptos básicos de la Estadística Bayesiana. Con ello se pretende que el lector comprenda, esencialmente, las ideas fundamentales de la Estadística Bayesiana que serán utilizadas durante el desarrollo de este trabajo. Por otro lado, debido a que realizar inferencia bayesiana se utilizan métodos de Monte Carlo vía cadenas de Markov, también se introducen estos conceptos. En lo referente a las cadenas de Markov, se esbozan las principales características de este tipo de procesos estocásticos y su relación con la distribución final de los parámetros, las cuales son el objetivo principal dentro del enfoque bayesiano. En este capítulo, también se presentan dos de los principales algoritmos de los métodos de Monte Carlo utilizados para realizar simulaciones y obtener muestras de las distribuciones finales para los parámetros de interés. Estos son el muestreo de Gibbs y el algoritmo de Metropolis-Hastings, ambos emplean la generación de cadenas de Markov cuya distribución estacional es la distribución objetivo. Estas cadenas deben cumplir con ciertas características para poder garantizar la convergencia hacia una distribución estacional y por lo tanto hacia la distribución final de la que se desea muestrear; por tal motivo se concluye con una explicación relacionada al diagnóstico empírico de la convergencia de las cadenas de Markov.

En el Capítulo 3 se exponen los aspectos principales de los datos longitudinales. Se muestran las distintas metodologías utilizadas para realizar estudios longitudinales, prestando un principal interés en el enfoque de los modelos de efectos mixtos propuesto por Laird y Ware (1982). Este tipo de modelos permite establecer una estructura de regresión donde la dependencia entre las observaciones de cada individuo se determina mediante de la inclusión de efectos aleatorios individuales. Una vez establecidas las características principales de los modelos de efectos mixtos, se procede a desarrollar un modelo para datos circulares cuya variable de respuesta sea de tipo angular y se realiza un análisis descriptivo del modelo .

Una vez establecido el modelos, el siguiente paso es realizar inferencia sobre los parámetros de interés. En el contexto bayesiano de la inferencia es común tener que implementar diversos métodos de simulación. Como ya introdujo, los métodos de Mote Carlo vía cadenas de Markov son uno de los más utilizados. Para poder implementar dichos métodos es necesario determinar las expresiones de las distintas distribuciones condicionales completas de los parámetros. Por lo tanto, en el Capítulo 4 se derivan todas estas distribuciones condicionales y se desarrolla la implementación del muestreo de Gibbs para el modelo propuesto. Por otro lado, además de la inferencia realizada a partir de los datos, es necesario poder determinar entre una serie de modelos distintos cuál es el más adecuado para implementar y por lo tanto se concluye el capítulo estableciendo una introducción a la comparación de modelos. En particular se da un panorama general de los criterios más utilizados para discernir entre dos modelos que aparentan explicar a los datos adecuadamente.

En el último capítulo, se realizan diversas aplicaciones del modelo. En los primeros ejemplos se aplican distintas especificaciones del modelos a datos simulados, a través de la especificación de los parámetros, con el objetivo de determinar el correcto funcionamiento y la adecuabilidad de su uso con datos longitudinales circulares. En el ejemplo final se aplica el modelo a una serie de datos reales relacionados con las direcciones de escape de 65 pulgas de mar y se realiza un análisis inferencial completo.

Capítulo 1

Preliminares

1.1. Datos direccionales

1.1.1. Introducción

En diversos campos del conocimiento se realizan estudios donde las observaciones son direcciones. Por ejemplo, existe interés en el análisis de direcciones de migraciones de aves, direcciones de viento, propagación de fisuras, orientación de las corrientes oceánicas, etc. A las observaciones que representan direcciones se les denomina *datos direccionales*. Existen diversas formas de obtener datos direccionales, dentro de las que destacan aquellas conocidas como mediciones de *brújula* y mediciones de *reloj*. Así, variables cuya naturaleza inicialmente no parece ser direccional pueden ser vistas como variables aleatorias de este tipo; por ejemplo, tiempos de arribo de personas a un hospital, horas de llegadas de aviones a una terminal específica, casos de influenza durante el año, etc.

Cabe resaltar que los datos direccionales presentan diferencias en la topología comparado con otros tipos de datos. Estos pertenecen a una curva cerrada, a diferencia de aquellos que forman parte de una línea recta. De lo anterior, además de su naturaleza periódica, es posible deducir que la teoría estadística necesaria para analizar datos direccionales deberá ser diferente que la utilizada para estudiar datos lineales.

Una de las primeras diferencias en el análisis surge con su representación gráfica. Los datos direccionales pueden ser representados como vectores unitarios en un espacio de dimensión q . En particular si los datos son vectores unitarios en \mathbb{R}^2 , se les denominan *datos circulares*; análogamente si dichos vectores unitarios pertenecen a \mathbb{R}^3 se dice que son *datos esféricos*. Este tipo de datos están asociados con ángulos en el espacio donde se encuentran representados y por lo tanto presentan una característica muy particular: periodicidad.

En el caso de los datos circulares, este tipo de datos se pueden representar de diversas formas; una de ellas es utilizando coordenadas cartesianas, donde el punto $(x, y) \in \mathbb{R}^2$ será un dato circular si $\|(x, y)\| = 1$. Otra forma es considerarlo como un número complejo unitario. Sin embargo, resulta más práctico expresar un dato circular en términos de un ángulo θ a través de coordenadas polares definidas como:

$$\mathbf{x} = (\cos \theta, \sin \theta)'$$

y para la cual se requiere una dirección inicial y sentido de rotación para evitar confusiones en la representación del dato.

De esta manera, la principal forma de visualizar datos direccionales en dos dimensiones es graficar puntos sobre la circunferencia de un círculo unitario. Como ejemplo, en la Figura 1.1 se muestran las direcciones de escape de 65 pulgas de mar.

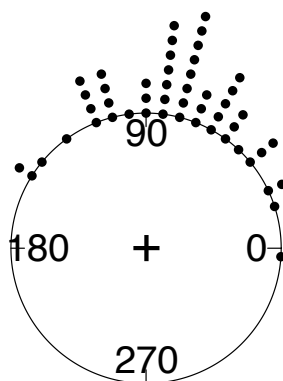


Figura 1.1: Representación gráfica de las direcciones de escape de 65 pulgas de mar.

En este tipo de gráfica los datos se presentan de manera desagrupada; sin embargo es posible representar los datos circulares agrupándolos de manera similar a los histogramas que se utilizan para los datos lineales. A una variante de este tipo de histograma circular se le denomina *diagrama de rosa*, el cual consiste en determinar sectores en lugar de barras y donde el área es proporcional a la frecuencia del grupo correspondiente. Para lograr esta relación entre el área y la frecuencia de los grupos se utilizan diferentes radios; una convención es que el radio sea la raíz cuadrada de la frecuencia del grupo seleccionado. La Figura 1.2 muestra el diagrama de rosa correspondiente a las direcciones de escape observadas para las 65 pulgas de mar.

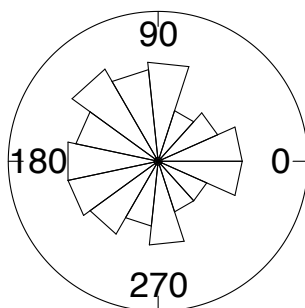


Figura 1.2: Diagrama de rosa para direcciones de las 65 pulgas de mar.

Estas representaciones gráficas de los datos circulares capturan con más detalle la naturaleza direccional de las observaciones. Tanto el diagrama circular crudo, como el diagrama de rosa permiten tener una referencia visual intuitiva de las direcciones medidas individualmente, así como de direccionales agrupadas.

1.1.2. Estadística descriptiva

Como ya se ha señalado, los datos direccionales poseen una topología diferente a los datos lineales al pertenecer a una curva cerrada. En particular, los datos circulares pueden ser representados como puntos sobre la circunferencia del círculo unitario donde el “inicio” de la curva coincide con el “final” de esta. Esta característica sugiere que las medidas descriptivas utilizadas

en los casos de datos lineales puedan no ser adecuadas, ya que para poder utilizar medidas descriptivas lineales estaríamos obligados a cortar el círculo en algún ángulo. El problema con esta posible solución es que los resultados obtenidos de las medidas tendrán una fuerte dependencia del ángulo de corte seleccionado.

Para notar esta dependencia, consideremos una muestra de dos ángulos formada por 20° y 340° . Si se corta el círculo en el ángulo 0° se tiene que el espacio muestral es $(0, 360)$ y que la media muestral será 180° (Figura 1.3), mientras que si el corte se realiza en 180° se obtiene 0° como la media muestral (Figura 1.4). Lo mismo sucede con la desviación estándar, en el primer caso se obtiene una desviación estándar de 160° y en el segundo caso es de 20° .

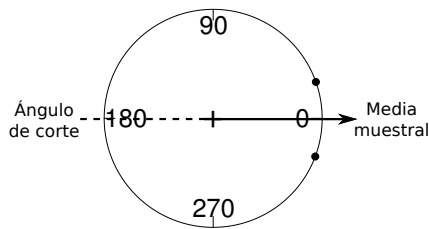


Figura 1.3: La media es 0° utilizando 180° como ángulo de corte.

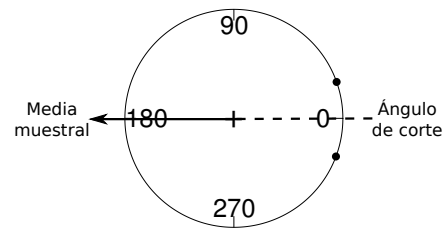


Figura 1.4: La media es 180° utilizando 0° como ángulo de corte.

Por lo anterior, es necesario introducir medidas adecuadas de localización y de dispersión que no sean dependientes del ángulo de inicio o de la dirección de giro seleccionada. A continuación se describen algunas de estas medidas.

Medidas de localización

Para describir datos o variables direccionales, si bien existen diversas medidas de localización, dos de las más importantes son la media y la mediana; por ello se definen a continuación.

Definición 1.1.1. Sean $\theta_1, \dots, \theta_n$ los ángulos asociados a los vectores unitarios $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$. Se define la dirección media $\bar{\theta}$ de $\theta_1, \dots, \theta_n$ como la dirección de la resultante $\mathbf{x}_1 + \mathbf{x}_2 + \dots + \mathbf{x}_n$, la cual coincide con el centro de masa $\bar{\mathbf{x}}$ de $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$.

Dado que $\frac{\sum_{i=1}^n \theta_i}{n}$ no está bien definida cuando nos referimos a ángulos, es importante resaltar que $\bar{\theta}$ no se refiere en ningún momento a dicha cantidad.

Sean $\theta_1, \dots, \theta_n$ los ángulos de los vectores unitarios $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$. Consideremos las coordenadas cartesianas de \mathbf{x}_i dadas por $(\cos \theta_i, \sin \theta_i)$, con $i = 1, \dots, n$. Entonces, las coordenadas del centro de masa de los vectores unitarios son:

$$\left(\frac{1}{n} \sum_{i=1}^n \cos \theta_i, \frac{1}{n} \sum_{i=1}^n \sin \theta_i \right)$$

Renombrando dichas coordenadas como

$$\bar{C} = \frac{1}{n} \sum_{i=1}^n \cos \theta_i \quad \text{y} \quad \bar{S} = \frac{1}{n} \sum_{i=1}^n \sin \theta_i$$

y dado que la dirección media coincide con la dirección del centro de masa, se tiene que

$$\bar{\theta} = \arg(\bar{C}, \bar{S})$$

o de manera equivalentemente $\bar{\theta}$ cumple con las ecuaciones:

$$\cos \bar{\theta} = \frac{\bar{C}}{\bar{R}}, \quad \sin \bar{\theta} = \frac{\bar{S}}{\bar{R}}$$

donde

$$\bar{R} = \sqrt{\bar{C}^2 + \bar{S}^2} \tag{1.1}$$

En este caso \bar{R} representa la longitud del centro de masa $\bar{\mathbf{x}}$. De las ecuaciones anteriores se puede observar que $\bar{\theta}$ no está definida cuando $\bar{R} = 0$. Ahora bien, cuando $\bar{R} > 0$ es posible dar una expresión explícita para $\bar{\theta}$ dada por:

$$\bar{\theta} = \begin{cases} \arctan\left(\frac{\bar{S}}{\bar{C}}\right) & \text{Si } \bar{C} \geq 0 \\ \arctan\left(\frac{\bar{S}}{\bar{C}}\right) + \pi & \text{Si } \bar{C} < 0 \end{cases}$$

Una propiedad importante de la media direccional es que esta es equi-variante bajo rotaciones, para un análisis de esta propiedad el lector puede dirigirse, por ejemplo a Jammalamadaka y SenGupta (2001).

Otra medida de localización importante para los datos circulares es la equivalente a la mediana utilizada en datos lineales.

Definición 1.1.2. Sean $\theta_1, \dots, \theta_n$ datos circulares representados por sus ángulos, se denomina mediana muestral, denotada por $\tilde{\theta}$, al ángulo α que cumple las siguientes condiciones:

1. la mitad de los datos estén contenidos dentro del arco $[\alpha, \alpha + \pi]$
2. La mayoría de los datos se encuentren más cerca de α que de $\alpha + \pi$

Para una revisión sobre la aplicación y el cálculo de la mediana para un conjunto de datos circulares, ver, por ejemplo, Mardia y Jupp (2000).

Medidas de concentración y de dispersión

Como se ha mencionado anteriormente, se necesita definir las medidas de localización de una manera distinta a las utilizadas en los datos lineales. Lo mismo sucede con las medidas de dispersión y de concentración. Tres ejemplos de estas medidas son: la longitud media resultante, la varianza direccional y una medida de dispersión de los datos con respecto de un ángulo dado.

Definición 1.1.3. *Se define la longitud media resultante \bar{R} de los datos circulares $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$, como la longitud del centro de masa $\bar{\mathbf{x}}$ dada por:*

$$\bar{R} = \sqrt{\bar{C}^2 + \bar{S}^2}.$$

Dado que los datos circulares $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ son vectores unitarios, entonces se tiene que $0 \leq \bar{R} \leq 1$. Además \bar{R} es una medida de concentración, ya que si las direcciones $\theta_1, \dots, \theta_n$ están muy agrupadas entonces \bar{R} será un valor aproximado a 1 y si dichas direcciones están muy dispersas entonces el valor será cercano a 0. Sin embargo, es necesario resaltar que si $\bar{R} \simeq 0$, esto no implica que las direcciones estén esparcidas de manera uniforme alrededor del círculo.

Definición 1.1.4. *Sean $\theta_1, \dots, \theta_n$ una muestra de ángulos, se define la varianza circular como:*

$$V = 1 - \bar{R}^2$$

donde $0 \leq V \leq 1$ debido a que $0 \leq \bar{R} \leq 1$.

Como se puede observar, variaciones mínimas ocurren cuando $V = 0$; de la misma manera, cuando los ángulos están distribuidos uniformemente alrededor del círculo la variación alcanza su límite superior que corresponde a $V = 1$. Algunos autores, p.ej. Batschelet (1981) y Jammalamadaka y Sen-Gupta (2001), se refieren a la cantidad $2(1 - \bar{R}^2)$ como la varianza circular. Sin embargo, para propósitos de este trabajo utilizaremos la definición dada

anteriormente.

Otra medida de utilidad es aquella que mide la dispersión de los ángulos $\theta_1, \dots, \theta_n$ con respecto a un ángulo α . Para poder dar una expresión de esta medida es necesario utilizar una distancia entre dos ángulos dados. Una forma útil de medir distancias entre dos ángulos β y ϕ es

$$1 - \cos(\beta - \phi)$$

Por lo tanto una medida de la dispersión promedio de los ángulos $\theta_1, \dots, \theta_n$ con respecto a un ángulo α esta dada por la cantidad

$$D(\alpha) = \frac{1}{n} \sum_{i=1}^n 1 - \cos(\theta_i - \alpha)$$

Una propiedad importante de $D(\alpha)$ es que esta alcanza su valor mínimo en $\bar{\theta}$. La demostración de este resultado puede consultarse en Mardia y Jupp(2000) y Nuñez-Antonio (2010). Esta propiedad es análoga al resultado para observaciones x_1, \dots, x_n en la recta, para las cuales se tiene que la cantidad

$$\frac{1}{n} \sum_{i=1}^n (x_i - a)^2$$

se minimiza cuando $a = \bar{x}$ y dicho valor mínimo es la varianza muestral.

1.2. Modelos para datos circulares

Debido a que los datos circulares presentan características especiales y diferentes a los datos definidos en la recta real, es necesario establecer distribuciones en el círculo unitario que describan estas propiedades. Una manera de especificar distribuciones en el círculo unitario es por medio de la definición de su función de distribución. De esta manera la distribución se puede considerar como la de un ángulo aleatorio θ y su función de distribución F se define como una función en la recta real que cumple:

$$F(x) = P(0 < \theta \leq x), \quad \text{donde } 0 \leq x \leq 2\pi$$

y además

$$F(x + 2\pi) - F(x) = 1, \quad , \text{ si } -\infty < x < \infty$$

Nótese que esta última condición indica que cualquier arco de longitud 2π en el círculo unitario tiene probabilidad 1.

Ahora bien, por definición se tiene que

$$F(0) = P(0 < \theta \leq 0) = 0 \quad \text{y} \quad F(2\pi) = 1$$

y si $\alpha \leq \beta \leq \alpha + 2\pi$, entonces

$$P(\alpha < \theta \leq \beta) = F(\beta) - F(\alpha)$$

Por otro lado, a diferencia de las funciones de distribución sobre la recta real, las funciones sobre el círculo unitario cumplen que:

$$\lim_{x \rightarrow \infty} F(x) = \infty, \quad \lim_{x \rightarrow -\infty} F(x) = -\infty$$

Para un enfoque más detallado sobre las propiedades de las funciones de distribución (así como la definición de la función característica y otras propiedades relacionadas) sobre el círculo unitario, se sugiere al lector consultar Mardia y Jupp (2000). A continuación se describen brevemente algunas de las distribuciones sobre el círculo más importantes.

Distribución Uniforme

La distribución uniforme es la más básicas de las distribuciones sobre el círculo unitario. Su función es:

$$f(\theta) = \frac{1}{2\pi}$$

y por lo tanto si $\alpha \leq \beta \leq \alpha + 2\pi$, se tiene que

$$P(\alpha < \theta \leq \beta) = \frac{\beta - \alpha}{2\pi}$$

Esta es la única distribución en el círculo unitario que es invariante bajo rotaciones y traslaciones. Debido a su simplicidad, generalmente es considerada como el modelo nulo.

Distribución cardioide

La distribución cardioide es generada realizando perturbaciones de la distribución uniforme con el uso de la función coseno. La densidad de una distribución cardioide $C(\mu, \rho)$, se define como:

$$f(\theta) = \frac{1}{2\pi} \{1 + 2\rho \cos(\theta - \mu)\}, \quad |\rho| < \frac{1}{2}$$

Si $\rho = 0$, entonces la distribución cardioide se reduce a una distribución uniforme. Ahora bien, cuando $0 < \rho$ se tiene que la longitud media resultante es ρ y la dirección media es μ . Además, en este caso, la distribución cardioide es simétrica y unimodal con moda en μ . El principal uso de esta distribución es como herramienta para aproximar distribuciones von Mises con valores pequeños de concentración.

Modelos wrapped

Dada una distribución definida en la recta real, es posible envolverla alrededor de la circunferencia del círculo. De tal manera que si X es una variable aleatoria sobre la recta real, entonces la variable aleatoria X_w de la distribución wrapped correspondiente esta dada por

$$X_w = X \pmod{2\pi}$$

O bien, debido a que es posible relacionar el círculo con los números complejos con módulo unitario, se puede expresar la función wrapped $X \mapsto X_w$ como

$$X_w \mapsto e^{2\pi X}$$

Si X tiene una función de distribución F , entonces la función de distribución F_w correspondiente a X_w está dada por

$$F_w(\theta) = \sum_{i=-\infty}^{\infty} \{F(\theta + 2\pi i) - F(2\pi i)\}, \quad 0 \leq \theta \leq 2\pi$$

Este tipo de modelos ha sido de los más utilizados para describir datos direccionales. Algunos modelos wrapped de mayor importancia son la distribución Normal wrapped, la distribución Cauchy wrapped y la distribución Poisson wrapped.

A modo de ejemplo, la distribución Normal wrapped $WN(\mu, \rho)$ se obtiene envolviendo una distribución $N(\mu, \sigma^2)$, donde

$$\rho = e^{-\sigma^2/2}$$

y por lo tanto la función de densidad de la normal wrapped $WN(\mu, \rho)$ es

$$\phi_w(\theta|\mu, \rho) = \frac{1}{\sqrt{-\log(\rho)}\sqrt{2\pi}} \sum_{i=-\infty}^{\infty} \exp\left\{\frac{-(\theta - \mu + 2\pi i)^2}{-2\log(\rho)}\right\}$$

La distribución $WN(\mu, \rho)$ es unimodal y simétrica con respecto a la moda μ . Además esta distribución se aproxima a una distribución uniforme cuando $\rho \rightarrow 0$ y se aproxima a un punto de la distribución en μ cuando $\rho \rightarrow 1$.

Modelos tipo von Mises-Fisher

En el contexto de la inferencia estadística en el círculo unitario, uno de los modelos más utilizados es el de von Mises-Fisher. Este tipo de modelos considera, a través de su construcción, la topología del espacio muestral. Dentro de este tipo de modelos se encuentran el modelo lattice, el modelo uniforme y el modelo von Mises-Fisher. Esta distribución es la más importante en el análisis de datos circulares y es análoga, sobre el círculo, a la distribución Normal en la recta real.

A la distribución de von Mises-Fisher para datos circulares se le conoce como distribución von Mises. La función de densidad de un ángulo aleatorio θ que sigue una distribución von Mises, con media direccional μ y parámetro de concentración κ , se define como:

$$M(\theta|\mu, \kappa) = \frac{1}{2\pi I_0(\kappa)} e^{\kappa \cos(\theta - \mu)}$$

donde I_0 denota la función de Bessel modificada de primer tipo y orden cero, la cual está definida como:

$$I_0(\kappa) = \frac{1}{2\pi} \int_0^{2\pi} e^{\kappa \cos(\theta)} d\theta$$

La moda de la distribución se encuentra en $\theta = \mu$ y la distribución es unimodal y simétrica con respecto a la moda. Si se define la antimoda como $\mu + \pi$,

entonces se tiene que el cociente de la densidad evaluada en la moda, entre la densidad evaluada en la antimoda es $e^{2\kappa}$. Por lo tanto, se puede observar que mientras más grande sea el valor de κ , mayor será la concentración alrededor de la moda.

Cabe mencionar que existe una relación entre la distribución von Mises y otras distribuciones. Por ejemplo, cuando $\kappa = 0$ se obtiene la distribución uniforme. También es posible aproximar una distribución von Mises con parámetro de concentración pequeño, a través de una Cardioide $C(\mu, \kappa/2)$. De manera más general, cualquier distribución von Mises se puede aproximar por medio de la distribución Normal wrapped. Un análisis detallado de estas relaciones, puede ser consultado en Mardia y Jupp (2000).

1.3. El modelo Normal Proyectado

Para realizar análisis sobre datos circulares es de suma importancia poder generar distribuciones para este tipo de datos. Si bien existen diversas maneras de generar distribuciones sobre el círculo unitario (ver, por ejemplo, Mardia y Jupp, 2000), este trabajo se enfoca en el modelo Normal bivariado proyectado, el cual se describe a continuación.

Supóngase que se tiene un vector aleatorio bivariado \mathbf{Y} tal que $P(\mathbf{Y} = 0) = 0$, entonces es posible realizar una proyección radial de este en el círculo unitario. Dicha proyección consiste en dividir al vector \mathbf{Y} por su norma euclídeana; es decir, que la proyección de \mathbf{Y} en el círculo unitario es un vector unitario $\mathbf{X} = \|\mathbf{Y}\|^{-1}\mathbf{Y}$. De esta manera el vector \mathbf{X} es un vector direccional y puede ser representado por un solo ángulo Θ .

Lo anterior muestra que una forma de generar distribuciones circulares es mediante la proyección radial de distribuciones que están definidas en el plano. Un caso importante es aquel en el cual el vector \mathbf{Y} tiene una distribución normal bivariada $N_2(\boldsymbol{\mu}, \boldsymbol{\Lambda})$, con vector de medias $\boldsymbol{\mu} = E(\mathbf{Y})$ y matriz de precisión $\boldsymbol{\Lambda} = \text{var}(\mathbf{Y})^{-1}$. En este caso se dice que el vector \mathbf{X} tiene una distribución llamada *distribución normal proyectada*, la cual se denota por $NP(\boldsymbol{\mu}, \boldsymbol{\Lambda})$.

A continuación se presenta una expresión general de la distribución Normal proyectada $NP(\boldsymbol{\mu}, \boldsymbol{\Lambda})$ y se deriva el caso $\boldsymbol{\Lambda} = \begin{pmatrix} 1 & 0 \\ 0 & \lambda \end{pmatrix}$, el cual es relevante en los desarrollos posteriores de este trabajo de tesis.

Definición 1.3.1. *La función de densidad de probabilidad de una distribución Normal proyectada, para un ángulo aleatorio Θ , está dada por*

$$NP(\theta|\boldsymbol{\mu}, \boldsymbol{\Lambda}) = \frac{\varphi(\theta|\mathbf{0}, \boldsymbol{\Lambda}) + |\boldsymbol{\Lambda}|^{-1/2} D \Phi(D) \phi(|\boldsymbol{\Lambda}|^{-1/2}(\mathbf{u}'\boldsymbol{\Lambda}\mathbf{u})^{-1/2}\boldsymbol{\mu} * \mathbf{u})}{\mathbf{u}'\boldsymbol{\Lambda}\mathbf{u}} I_{(0,2\pi)}$$

donde $\varphi(\cdot|\mathbf{0}, \boldsymbol{\Lambda})$, denota la función de densidad de una $N_2(\cdot|\mathbf{0}, \boldsymbol{\Lambda})$, $\Phi(\cdot)$ y $\phi(\cdot)$ denotan las funciones de distribución y de densidad de una Normal estándar, respectivamente. Además se definen $\mathbf{u} = (\cos \theta, \sin \theta)'$, $\boldsymbol{\mu} * \mathbf{u} = \boldsymbol{\mu}'\mathbf{u}$ y

$$D = \frac{\boldsymbol{\mu}'\boldsymbol{\Lambda}^{-1}\mathbf{u}}{(\mathbf{u}'\boldsymbol{\Lambda}^{-1}\mathbf{u})^{1/2}}.$$

1.3.1. La normal proyectada $NP(\boldsymbol{\mu}, \boldsymbol{\Lambda} = \begin{pmatrix} 1 & 0 \\ 0 & \lambda \end{pmatrix})$.

Para poder obtener $NP(\boldsymbol{\mu}, \boldsymbol{\Lambda})$ para el caso cuando $\boldsymbol{\Lambda} = \begin{pmatrix} 1 & 0 \\ 0 & \lambda \end{pmatrix}$, es necesario definir la función de densidad de probabilidad de un vector bivariado \mathbf{Y} con una distribución $N_2(\boldsymbol{\mu}, \boldsymbol{\Lambda})$.

Definición 1.3.2. *Sea \mathbf{Y} vector bivariado con distribución $N_2(\boldsymbol{\mu}, \boldsymbol{\Lambda})$. Entonces, la función de densidad de probabilidad de \mathbf{Y} está dada por*

$$f(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Lambda}) = \frac{|\boldsymbol{\Lambda}|^{1/2}}{2\pi} \exp\left\{-\frac{1}{2}(\mathbf{y} - \boldsymbol{\mu})'\boldsymbol{\Lambda}(\mathbf{y} - \boldsymbol{\mu})\right\}.$$

De la definición (1.3.2) se concluye que cuando $\boldsymbol{\Lambda} = \begin{pmatrix} 1 & 0 \\ 0 & \lambda \end{pmatrix}$, debido a que $|\boldsymbol{\Lambda}| = \lambda$, entonces la densidad del vector \mathbf{Y} se expresa como

$$f(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Lambda}) = \frac{\lambda^{1/2}}{2\pi} \exp\left\{-\frac{1}{2}(\mathbf{y} - \boldsymbol{\mu})'\boldsymbol{\Lambda}(\mathbf{y} - \boldsymbol{\mu})\right\}. \quad (1.2)$$

Proposición 1.3.1. Sea \mathbf{Y} un vector aleatorio con distribución Normal bivariada con vector de medias $\boldsymbol{\mu} = (\mu_1, \mu_2)'$ y matriz de precisión $\boldsymbol{\Lambda} = \begin{pmatrix} 1 & 0 \\ 0 & \lambda \end{pmatrix}$ y sea la transformación a coordenadas polares definida como

$$\mathbf{y} = r(\cos\theta, \sin\theta)' = r\mathbf{v}'$$

donde $\theta \in (0, 2\pi]$ y $r \in \mathbb{R}^+$. Entonces, la función de densidad conjunta de R y Θ está dada por

$$f(r, \theta | \boldsymbol{\mu}, \boldsymbol{\Lambda}) = r C d^2 \exp\left(-\frac{1}{2} d^2 r^2 + d^2 b r\right),$$

donde $d^2 = \cos^2(\theta) + \lambda \sin^2(\theta)$,

$$b = \frac{\mu_1 \cos(\theta) + \lambda \mu_2 \sin(\theta)}{\cos^2(\theta) + \lambda \sin^2(\theta)}$$

$$\text{y } C = \frac{\lambda^{1/2} \exp\left\{-\frac{1}{2} (\mu_1^2 + \lambda \mu_2^2)\right\}}{2\pi d^2}.$$

DEMOSTRACIÓN. Utilizando teorema de cambio de variable y dado que el determinante de la transformación es

$$|J(r, \theta)| = \begin{vmatrix} \cos(\theta) & -r \sin(\theta) \\ \sin(\theta) & r \cos(\theta) \end{vmatrix} = r$$

se tiene lo siguiente

$$\begin{aligned} f(r, \theta | \boldsymbol{\mu}, \boldsymbol{\Lambda}) &= f(r \cos(\theta), r \sin(\theta) | \boldsymbol{\mu}, \boldsymbol{\Lambda}) |J(r, \theta)| \\ &= \frac{\lambda^{1/2}}{2\pi} \exp\left\{-\frac{1}{2}(r\mathbf{v} - \boldsymbol{\mu})' \boldsymbol{\Lambda} (r\mathbf{v} - \boldsymbol{\mu})\right\} r \\ &= \frac{r \lambda^{1/2}}{2\pi} \exp\left\{-\frac{1}{2}[(\mathbf{v}' \boldsymbol{\Lambda} \mathbf{v})r^2 - 2(\mathbf{v}' \boldsymbol{\Lambda} \boldsymbol{\mu})r + \boldsymbol{\mu}' \boldsymbol{\Lambda} \boldsymbol{\mu}]\right\} \\ &= \frac{r \lambda^{1/2}}{2\pi} \exp\left\{-\frac{1}{2} \boldsymbol{\mu}' \boldsymbol{\Lambda} \boldsymbol{\mu}\right\} \exp\left\{-\frac{1}{2}(\mathbf{v}' \boldsymbol{\Lambda} \mathbf{v})\right\} \left[r^2 - 2 \frac{\mathbf{v}' \boldsymbol{\Lambda} \boldsymbol{\mu}}{\mathbf{v}' \boldsymbol{\Lambda} \mathbf{v}} r\right] \\ &= \frac{r \lambda^{1/2}}{2\pi} \exp\left\{-\frac{1}{2} \boldsymbol{\mu}' \boldsymbol{\Lambda} \boldsymbol{\mu}\right\} \frac{1}{\mathbf{v}' \boldsymbol{\Lambda} \mathbf{v}} \times \\ &\quad (\mathbf{v}' \boldsymbol{\Lambda} \mathbf{v}) \exp\left\{-(1/2)(\mathbf{v}' \boldsymbol{\Lambda} \mathbf{v})r^2 + (\mathbf{v}' \boldsymbol{\Lambda} \boldsymbol{\mu}) \left(\frac{\mathbf{v}' \boldsymbol{\Lambda} \boldsymbol{\mu}}{\mathbf{v}' \boldsymbol{\Lambda} \mathbf{v}}\right) r\right\} \end{aligned}$$

Ahora bien, definiendo

$$d^2 = \mathbf{v}'\mathbf{\Lambda}\mathbf{v} = \cos^2(\theta) + \lambda \operatorname{sen}^2(\theta)$$

$$b = \frac{\mathbf{v}'\mathbf{\Lambda}\boldsymbol{\mu}}{\mathbf{v}'\mathbf{\Lambda}\mathbf{v}} = \frac{\mu_1 \cos(\theta) + \lambda \mu_2 \operatorname{sen}(\theta)}{\cos^2(\theta) + \lambda \operatorname{sen}^2(\theta)}$$

$$C = \frac{\lambda^{1/2} \exp\{-\frac{1}{2}\boldsymbol{\mu}'\mathbf{\Lambda}\boldsymbol{\mu}\}}{2\pi \mathbf{v}'\mathbf{\Lambda}\mathbf{v}} = \frac{\lambda^{1/2} \exp\{-\frac{1}{2}(\mu_1^2 + \lambda \mu_2^2)\}}{2\pi d^2}$$

se obtiene que

$$f(r, \theta | \boldsymbol{\mu}, \mathbf{\Lambda}) = r C d^2 \exp\left(-\frac{1}{2} d^2 r^2 + d^2 b r\right)$$

□

Tomando la definición dada para d^2 , b y C , en este momento es posible dar una expresión para la Normal proyectada para el caso cuando $\mathbf{\Lambda} = \begin{pmatrix} 1 & 0 \\ 0 & \lambda \end{pmatrix}$, la cual a partir de ahora se denotará por $NP(\theta | \boldsymbol{\mu}, \lambda)$

Proposición 1.3.2. Bajo las mismas condiciones de la Proposición (1.3.1), la función la densidad de probabilidad de la Normal proyectada $NP(\theta | \boldsymbol{\mu}, \lambda)$ es

$$NP(\theta | \boldsymbol{\mu}, \lambda) = \int_0^\infty f(r, \theta | \boldsymbol{\mu}, \lambda) dr$$

$$= C \left[1 + \frac{db \Phi(db)}{\phi(db)} \right] I_{(0, 2\pi]}(\theta)$$

donde $\Phi(\cdot)$ y $\phi(\cdot)$ denotan las funciones de distribución y de densidad de una Normal estándar, respectivamente.

DEMOSTRACIÓN. Por la Proposición (1.3.1), se tiene que

$$PN(\theta | \boldsymbol{\mu}, \lambda) = \int_0^\infty f(r, \theta | \boldsymbol{\mu}, \lambda) dr$$

$$= C \int_0^\infty r d^2 \exp\left\{-\frac{1}{2} d^2 r^2\right\} \exp\{d^2 b r\} dr.$$

Realizando integración por partes, para la cual se toman u y dv de la siguiente manera

$$u = \exp\{d^2 b r\} \quad y \quad dv = r d^2 \exp\{-\frac{1}{2}d^2 r^2\},$$

de donde

$$v = \int r d^2 \exp\{-\frac{1}{2}d^2 r^2\} dr = -\exp\{-\frac{1}{2}d^2 r^2\}.$$

y por lo tanto

$$\begin{aligned} NP(\theta|\boldsymbol{\mu}, \lambda) &= C \int_0^\infty r d^2 \exp\{-\frac{1}{2}d^2 r^2\} \exp\{d^2 b r\} dr \\ &= C \left(\left[-\exp\{d^2 b r\} \exp\{-\frac{1}{2}d^2 r^2\} \right]_0^\infty + \int_0^\infty \exp\{-\frac{1}{2}d^2 r^2\} \exp\{d^2 b r\} d^2 b dr \right) \\ &= C \left(1 + \int_0^\infty d^2 b \exp\{-(dr - db)^2/2 + d^2 b^2/2\} dr \right). \end{aligned}$$

Por otro lado, resolviendo la integral de la última igualdad se tiene que

$$\begin{aligned} PN(\theta|\boldsymbol{\mu}, \lambda) &= C \left(1 + \left[d^2 b \exp\{d^2 b^2/2\} \times \int_0^\infty \exp\{-(dr - db)^2/2\} dr \right] \right) \\ &= C \left(1 + \left[d^2 b \exp\{d^2 b^2/2\} \sqrt{2\pi} \times \int_{-db}^\infty \frac{1}{\sqrt{2\pi}} \exp\{-s^2/2\} ds \right] \right) \\ &= C \left(1 + \left[\frac{db \Phi(db)}{\phi(db)} \right] \right) \end{aligned}$$

□

Nótese que se han derivado tanto la función de densidad de probabilidad de la Normal proyectada $f(\theta|\boldsymbol{\mu}, \lambda)$, como la densidad conjunta $f(r, \theta|\boldsymbol{\mu}, \lambda)$. Entonces es posible obtener de manera particular la densidad condicional de R dado Θ denotada por $f(r|\theta, \boldsymbol{\mu}, \lambda)$. Esta densidad es importante dentro de los procedimientos de inferencia bayesiana que se realizan en las siguientes secciones de este trabajo, por lo que se deriva a continuación.

Proposición 1.3.3. Bajo las mismas condiciones de la Proposición (1.3.1), la función de densidad condicional de R dado Θ está dada por

$$f(r|\theta, \boldsymbol{\mu}, \lambda) = \frac{d^2 r \exp\left(-\frac{1}{2} d^2 [r^2 - 2 b r]\right)}{1 + \frac{db}{\phi(db)} \Phi(db)} I_{(0,\infty)}(r).$$

DEMOSTRACIÓN. Por la definición de probabilidad condicional y por las proposiciones (1.3.1) y (1.3.2) se obtiene

$$\begin{aligned} f(r|\theta, \boldsymbol{\mu}, \lambda) &= \frac{f(r, \theta|\boldsymbol{\mu}, \lambda)}{f(\theta|\boldsymbol{\mu}, \lambda)} \\ &= \frac{r C d^2 \exp\left(-\frac{1}{2} d^2 r^2 + d^2 b r\right)}{C \left(1 + \frac{db \Phi(db)}{\phi(db)}\right)} \\ &= \frac{d^2 r \exp\left(-\frac{1}{2} d^2 [r^2 - 2 b r]\right)}{1 + \frac{db}{\phi(db)} \Phi(db)} \end{aligned}$$

□

Capítulo 2

Estadística Bayesiana

Existe una extensa bibliografía en la cual se trata a profundidad los aspectos relacionados con la Estadística bayesiana; el lector puede referirse a Box y Tiao (1973), Bernardo (1981), Bernardo y Smith (2000), Press (2003), Gelman *et al* (2004), Gosh *et al*(2006), Koch (2007), Boldstad (2007), Hoff (2009)y Congdon (2014), entre otros. A continuación se presenta una breve introducción a la inferencia Bayesiana.

El objetivo principal de la inferencia estadística es aprender acerca de las cantidades desconocidas θ de un modelo $f(x|\theta)$, mediante la obtención de información muestral del proceso real asociado. Es decir, basados en datos se explora los valores de los parámetros planteados, qué valores de estos son más probables y cuáles pueden ser considerados como estimadores plausibles de los valores reales que son desconocidos. Para ello es necesario establecer un modelo $p(x|\theta)$. Para el desarrollo de esta tesis se utilizará el enfoque Bayesiano de la estadística.

Para realizar inferencia bayesiana se requiere tener, además del modelo, una distribución para θ . Esta estructura de una distribución de probabilidad para las cantidades desconocidas surge de la forma como la aproximación bayesiana interpreta la naturaleza de los parámetros.

En la metodología bayesiana, tanto los datos obtenidos como los parámetros son tratados como aleatorios y por lo tanto se describen a través de una distribución de probabilidad. En el caso de los parámetros, el conocimiento

inicial de estos (previo a la obtención de datos) es representado mediante lo que se conoce como *distribución inicial*, $p(\theta)$. Para la estadística bayesiana existe una incertidumbre acerca del valor real de los parámetros y por lo tanto son considerados como variables aleatorias por si mismos. Las afirmaciones de probabilidad que se realizan sobre los parámetros deben ser considerados como “grado de creencia” y por lo tanto la distribución inicial que se plantea tiene la característica de ser subjetiva. Es decir, que cada persona puede tener su propia distribución inicial, la cual contendrá los distintos pesos que dicha persona le asigne a los valores posibles de los parámetros para medir qué tan creíble es para el o ella cada valor de los parámetros antes de que se observen los datos.

Esta posibilidad de incluir la información previa (experiencia del observador o representación del conocimiento sobre el comportamiento de estas cantidades de interés) vía la distribución inicial, es una de las características principales del razonamiento Bayesiano. Sin embargo, su importancia no radica únicamente en introducir la información individual (subjetiva), si no en la posibilidad de interactuar con nueva información a partir de la obtención de una muestra de datos. Los datos obtenidos permiten actualizar la distribución inicial y de esta manera mejorar el conocimiento sobre los parámetros, lo cual se refleja en lo que se conoce como *distribución final*, $p(\theta|\mathbf{y})$. Dicha actualización se realiza por medio del Teorema de Bayes de la siguiente manera:

$$p(\theta|\mathbf{y}) \propto p(\mathbf{y}|\theta)p(\theta) \quad (2.1)$$

donde $p(\theta)$ es la *distribución inicial o a priori* de θ y $p(\mathbf{y}|\theta)$ es la verosimilitud, generalmente denotada como $L(\theta|y_1, \dots, y_n)$ ya que es vista como una función del parámetro θ .

Como se puede observar en (2.1) existe una constante de proporcionalidad, a la cual se le denomina *constante de normalización*. Esta constante está dada por $1/\int_{\Theta} p(\mathbf{y}|\theta)p(\theta)d\theta$, es decir la distribución predictiva de \mathbf{y} y la cual no depende de los parámetros.

La distribución final $p(\theta|\mathbf{y})$ es uno de los objetivos principales de la estadística Bayesiana. A partir de esta distribución se pueden construir medidas descriptivas que proporcionen resúmenes de dicha distribución. Por ejemplo

es posible calcular la media *a posteriori* o la varianza *a posteriori*, de la siguiente manera:

$$E(\theta|\mathbf{y}) = \int_{-\infty}^{\infty} \theta p(\theta|\mathbf{y})d\theta$$

$$\text{Var}(\theta|\mathbf{y}) = \int_{-\infty}^{\infty} (\theta - E(\theta|\mathbf{y}))^2 p(\theta|\mathbf{y})d\theta.$$

Sin embargo el uso de la distribución final puede ser aún mayor, es a partir de esta distribución que se realiza inferencia sobre los parámetros. Debido que a través del Teorema de Bayes se obtiene una distribución de probabilidad para los parámetros θ , que se ha actualizado por medio de los datos, entonces es posible utilizar esta distribución *a posteriori* para dar respuesta a problemas relacionados con la estimación, el contraste de hipótesis y la predicción. Es posible, por ejemplo, dar intervalos de credibilidad para el valor real del parámetro o para una observación “ futura ” de la variable de interés.

Cabe señalar que la distribución final tiene la propiedad de que puede ser recalculada cuando surgen nuevas observaciones, lo cual permite actualizar la información acerca de los parámetros sin tener que realizar el proceso metodológico desde el principio.

A manera de resumen el proceso bayesiano relacionado al análisis de datos puede ser idealizado de la siguiente manera:

- I Establecer un modelo probabilístico completo. Es decir, una distribución de probabilidad conjunta para todas las cantidad observables y no observables; así como una distribución inicial para los parámetros que resuma el conocimiento previo sobre los mismos
- II Cálculo e interpretación de la distribución final a través del Teorema de Bayes. Esto es la obtención de la distribución de probabilidad condicional de las cantidades desconocidas de interés dado los datos observados
- III Realizar inferencia sobre los parámetros a partir de la distribución final, la cual contiene toda la información disponible (la información a priori y la que aportan las observaciones muestrales) sobre las cantidades desconocidas.

Se debe mencionar que el proceso señalado en el punto (III) puede requerir del cálculo de integrales en varias dimensiones. Por este motivo, comúnmente es necesario el empleo de métodos de simulación estocástica.

2.1. Métodos de Monte Carlo vía cadenas de Markov

El desarrollo de la estadística bayesiana recae en la posibilidad de calcular integrales complejas de funciones multidimensionales. La importancia de la resolución de este tipo de integrales se debe a que son necesarias para obtener la constante de proporcionalidad de la distribución final de los parámetros de interés, obtener las distribuciones marginales o para calcular la esperanza y varianza a posteriori.

En muchos casos no es posible realizar una evaluación explícita de estas integrales, por lo que es necesario realizar integración numérica o emplear algún método de aproximación analítica. Ahora bien, la aplicación de este tipo de método se ve afectada por la dimensión del espacio paramétrico y por la complejidad de las expresiones. Sin embargo, se han desarrollado una serie de técnicas llamadas *Métodos de Monte Carlo vía cadenas de Markov (MCMC)*, que permiten obtener una muestra de la distribución final directamente aunque se desconozca la constante de normalización y con ello obtener estimaciones de las cantidades de interés.

Las cadenas de Markov son procesos que describen trayectorias donde las cantidades sucesivas son descritas probabilísticamente de acuerdo al valor de sus inmediatos predecesores. Así pues, la simulación de una distribución mediante los métodos de MCMC se realiza construyendo una cadena de Markov que converja a la distribución objetivo de la cual se pretende obtener una muestra. Para poder asegurar la convergencia de la cadena de Markov hacia la distribución deseada es necesario que cumpla con ser homogénea, aperiódica e irreducible. A continuación se muestran algunas propiedades de las cadenas de Markov y se detallan dos de los algoritmos más importantes de MCMC.

Cadenas De Markov

Una cadena de Markov es un tipo especial de proceso estocástico donde dado el estado actual, los estados pasado y futuro son independientes. Un proceso estocástico se define como una sucesión de variables aleatorias $\{X_t : t \in T\}$ para algún conjunto T . Al conjunto de todos los posibles estados se le denomina *espacio de estados* y T es el *conjunto de índices*. Si bien el conjunto de índices puede ser continuo, para el desarrollo de este trabajo se asumirá que $T = \{0, 1, 2, \dots\}$.

Definición 2.1.1. *Una **cadena de Markov** es un proceso estocástico $\{X_t\}$ tal que para $t \in \mathbb{N}$ y para cualquier sucesión de estados $x_0, x_1, x_2, \dots, x_t, x$ se tiene que:*

$$P(X_{t+1} = x | X_t = x_t, X_{t-1} = x_{t-1}, \dots, X_0 = x_0) = P(X_{t+1} = x | X_t = x_t)$$

A la probabilidad condicional $P(X_{t+1} = x_j | X_t = x_i)$ se le denomina **probabilidad de transición** al tiempo t del estado x_i al x_j . Como simplificación de la notación se denotara a los estados x_i y x_j simplemente como los estados i y j

Si bien las cadenas de Markov poseen diversas propiedades, enseguida se mencionan aquellas que son relevantes para entender el funcionamiento de los métodos de Monte Carlo vía cadenas de Markov. Por tal motivo únicamente se consideran cadenas de Markov homogéneas; es decir, aquellas cuya probabilidad de transición de un estado a otro no depende del tiempo en el que se encuentre. Formalmente esto es:

$$P(X_{t+1} = j | X_t = i) = P(X_1 = j | X_0 = i) = p_{ij}$$

y por lo tanto, para una cadena homogénea se tiene que la probabilidad de transición de n pasos entre dos estados i y j está dada por:

$$p_{ij}^n = P(X_{k+n} = j | X_k = i) = P(X_1 = j | X_0 = i)$$

Una propiedad importante de algunas cadenas de Markov se refiere a la posibilidad de pasar de cualquier estado actual a otro que pertenezca al espacio de estados. A lo anterior se le conoce como la comunicación entre estados y es deseable que exista comunicación de todos los posibles estados de la cadena. Esta propiedad se define como sigue.

Definición 2.1.2. Una cadena de Markov X_t es **irreducible** si para cualesquiera dos estados i y j existe un $n_{ij} > 0$ tal que

$$p_{ij}^{n_{ij}}(t) > 0$$

Por otro lado se tiene la propiedad de recurrencia, la cual indica que la cadena regresará al estado actual después de un cierto tiempo t .

Definición 2.1.3. La cadena irreducible X_t será **recurrente** si para todo estado i , $P(\min\{t > 0 : X_t = i | X_0 = i\} < \infty) = 1$

La siguiente propiedad es de gran importancia, puesto que es una condición necesaria para garantizar que una cadena irreducible tenga una distribución estacionaria.

Definición 2.1.4. Dada X_t irreducible y recurrente, se denominará **positiva recurrente** si para todo estado i se cumple que

$$E[\min\{t > 0 : X_t = i | X_0 = i\}] < \infty$$

Si bien para una cadena irreducible y positiva recurrente existe una distribución estacionaria, es importante garantizar que no sea posible regresar al estado actual a través de la transición a estados distintos al actual. De existir esta posibilidad de regresar al estado actual de esa forma, la cadena podría entrar en un ciclo que no le permitiera alcanzar una convergencia y esto haría imposible obtener resultados relacionados con la distribución objetivo. Por lo tanto, dicha característica se define como:

Definición 2.1.5. Una cadena de Markov X_t es **aperiódica** si para todo estado i e índice t se cumple que

$$m.c.d.\{n > 0 : p_{ii}^n > 0\} = 1$$

A continuación se caracterizan a las distribuciones estacionarias, las cuales son de gran utilidad en el contexto de la inferencia bayesiana ya que estas serán las distribuciones finales de los parámetros de interés.

Definición 2.1.6. Se dice que una distribución π es una distribución estacionaria de una cadena de Markov con probabilidad de transición p_{ij} si

$$\sum_{i=0}^{\infty} \pi(i) p_{ij} = \pi(j)$$

De esta manera, con las propiedades establecidas para una cadena de Markov, podemos enunciar el siguiente teorema importante y de gran utilidad en el desarrollo práctico de la estadística bayesiana.

Proposición 2.1.1. Si X es una cadena irreducible, aperiódica y positiva recurrente, entonces posee una única distribución estacionaria $\pi(\cdot)$. En ese caso se dirá que X es irreducible y ergódica, las cuales conforme cumplen:

- 1) $p_{ij}^n \rightarrow \pi(j)$, cuando $n \rightarrow \infty$ par toda i y j .
- 2) Si dada f una función real y $E_\pi\{|f(X)|\} < \infty$, entonces

$$P\left(\frac{\sum_{t=1}^n f(X_t)}{n} \rightarrow E_\pi[f(X)]\right) = 1$$

donde $E_\pi[f(X)] = \sum_i f(X_i)\pi(X_i)$, la esperanza de $f(X)$ con respecto a $\pi(\cdot)$.

Algoritmo Metropolis-Hastings

Uno de los principales algoritmos de MCMC es el algoritmo de Metropolis-Hastings, cuyo nombre se deriva del artículo de Metrópolis *et al*(1953) y su generalización propuesta por Hastings (1970).

La propuesta original del algoritmo fue planteado en el contexto de la mecánica estadística y se centra en el cálculo de propiedades de sustancias químicas, las cuales son consideradas compuestas por moléculas individuales que interactúan entre sí. A pesar del enfoque físico del algoritmo, este obtuvo cierto reconocimiento entre los estadísticos. Sin embargo, el método fue considerado como una manera de resolver integrales y no como una forma de simular observaciones de alguna distribución.

No fue hasta 1970, que Hastings generalizó y mejoró el algoritmo de Metropolis, que se le dió al método el sentido moderno de su aplicación. Hastings consideró al algoritmo de Metrópolis principalmente como una forma para obtener muestras de distribuciones de probabilidad con alta dimensionalidad. En su artículo, muestra que que el método de Metropolis involucra la matriz de transición de una cadena de Markov y presenta a la distribución objetivo en términos de la distribución estacionaria de la cadena.

A continuación se describe el algoritmo Metrópolis desde el enfoque bayesiano. Considérese una distribución final de los parámetros, $\pi(\theta|y)$, de la cual se desea obtener una muestra. Para construir una cadena de Markov θ_t , se elige una distribución de transición $q(\Theta_{t+1}|\Theta_t)$ que sea simétrica (esto es que $q(\theta_j|\theta_i) = q(\theta_i|\theta_j)$) con la cual se generan valores candidatos para los estados subsecuentes y se establece $\alpha(\theta^*, \theta)$ una probabilidad de aceptación del estado candidato θ^* como el siguiente estado de la cadena de Markov.

Por lo tanto, el algoritmo de Metrópolis-Hastings, asociado a la distribución objetivo $\pi(\theta|y)$ y la distribución de transición $q(\cdot|\cdot)$, está dado por:

Algoritmo de Metrópolis

Dado el valor θ_t ,

- 1) Generar $\theta^* \sim q(\theta^*|\theta_t)$
- 2) Actualizar la cadena para θ_{t+1} de la siguiente manera:

$$X_{t+1} = \begin{cases} \theta^* & \text{con probabilidad } \alpha(\theta^*, \theta_t) \\ \theta_t & \text{con probabilidad } 1 - \alpha(\theta^*, \theta_t) \end{cases}$$

donde

$$\alpha(\theta^*, \theta_t) = \min \left\{ \frac{\pi(\theta^*|y) q(\theta_t|\theta^*)}{\pi(\theta_t|y) q(\theta^*|\theta_t)}, 1 \right\}$$

En términos prácticos, la implementación del algoritmo para simular observaciones de $\pi(\theta|y)$ puede expresarse como:

- Paso 1) Establecer el un contador $t = 0$ y establecer un valor inicial θ_0
- Paso 2) Generar un valor θ^* de $q(\theta^*|\theta_t)$.
- Paso 3) Generar un valor u uniforme con $U(0, 1)$.
- Paso 4) Calcular $\alpha(\theta^*, \theta_t)$ donde

$$\alpha(\theta^*, \theta) = \min \left\{ \frac{\pi(\theta^*|y) q(\theta|\theta^*)}{\pi(\theta|y) q(\theta^*|\theta)}, 1 \right\}$$

Paso 5) Si $u < \alpha(\theta^*, \theta)$
 $\theta_{t+1} = \theta^*$
 En otro caso
 $\theta_{t+1} = \theta_t$

Paso 6) Actualizar el contador $t = t + 1$ y regresar al Paso 2 hasta alcanzar la convergencia.

Muestreo de Gibbs

El muestreo de Gibbs fue propuesto por Gelman y Gelman (1984) en el contexto del procesamiento de imágenes. El nombre se debe a que, en dicho contexto, la distribución objetivo de interés para Gelman y Gelman era la distribución de Gibbs. Este trabajo pasó prácticamente inadvertido para la comunidad estadística hasta que Gelfand y Smith (1990) señalaron que el muestreo de Gibbs podría servir para otras distribuciones posteriores.

Para detallar la idea principal del muestreo de Gibbs, supóngase que se tiene un vector de parámetros compuesto por d parámetros $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_d)'$; además las distribuciones condicionales completas $\pi_i(\theta_i | \boldsymbol{\theta}_{-i}, \mathbf{y})$, $i = 1, \dots, d$, son completamente conocidas y es posible simular de ellas. Entonces, el muestreo de Gibbs permite obtener una muestra de la densidad conjunta de los parámetros dados los datos $\pi(\boldsymbol{\theta} | \mathbf{y})$, a partir de las $\pi_i(\theta_i | \boldsymbol{\theta}_{-i}, \mathbf{y})$. Esto se logra creando una cadena de Markov multivariada $\boldsymbol{\theta}^{(t)} = (\theta_1^{(t)}, \theta_2^{(t)}, \dots, \theta_d^{(t)})'$ mediante la simulación secuencial de las condicionales completas para cada componente.

Por lo tanto el método de muestreo de Gibbs para generar la cadena de Markov $\boldsymbol{\theta}^{(t)} = (\theta_1^{(t)}, \theta_2^{(t)}, \dots, \theta_d^{(t)})'$ y con ello obtener una muestra de la densidad conjunta $\pi(\boldsymbol{\theta} | \mathbf{y})$, puede ser descrito por los siguientes pasos:

Muestreo de Gibbs

Dado $\boldsymbol{\theta}^{(t)} = (\theta_1^{(t)}, \theta_2^{(t)}, \dots, \theta_d^{(t)})'$,

Paso 1) Establecer el contador $t=0$

Paso 2) Generar $\boldsymbol{\theta}^{(t+1)} = (\theta_1^{(t+1)}, \theta_2^{(t+1)}, \dots, \theta_d^{(t+1)})'$ de la siguiente manera:

- Generar $\theta_1^{(t+1)}$ de $\pi_i(\theta_1|\theta_2^{(t)}, \dots, \theta_d^{(t)})$
- Generar $\theta_2^{(t+1)}$ de $\pi_i(\theta_2|\theta_1^{(t+1)}, \theta_3^{(t)}, \dots, \theta_d^{(t)})$
- Generar $\theta_3^{(t+1)}$ de $\pi_i(\theta_3|\theta_1^{(t+1)}, \theta_2^{(t+1)}, \theta_4^{(t)}, \dots, \theta_d^{(t)})$
- \vdots
- Generar $\theta_d^{(t+1)}$ de $\pi_i(\theta_d|\theta_1^{(t+1)}, \dots, \theta_{d-1}^{(t+1)})$

Paso 3) Actualizar el contador a $t = t + 1$ y regresar al Paso 2

Este método de muestreo utiliza densidades de una dimensión para poder simular de la densidad multivariada, las cuales son relativamente más fáciles de simular y es por esta razón que el muestreo de Gibbs sea una técnica muy utilizada para obtener muestras de distribuciones finales multivariadas. La característica principal es que la simulación de un parámetro utiliza los valores obtenidos en los pasos anteriores del ciclo para los parámetros que han sido simulados dentro del ciclo y mantiene los valores actuales para aquellos que aún no han sido generados.

Diagnostico de Convergencia

Un elemento generado de la cadena de Markov es considerado una observación de la distribución estacionaria, teóricamente cuando el número de iteraciones tiende a infinito. En la práctica no es posible cumplir con esa condición y en su lugar se toma el valor de la cadena después de una cantidad suficientemente grande de iteraciones, de tal forma que dicho valor se aproximado a una observación de la distribución objetivo. A partir de ese valor, los siguientes estados de la cadena continuarán siendo considerados provenientes de la distribución estacionaria. Otra complicación práctica es que, para obtener una muestra de la distribución estacionaria, las observaciones deben ser independientes entre si y los estados de una cadena de Markov tienen por su naturaleza dependencia entre sí. Estos dos aspectos forman parte fundamental del estudio de la convergencia de cualquier cadena simulada.

Existen principalmente dos formas para realizar el análisis de convergencia. El primero es un enfoque teórico, con el cual se busca medir distancias y establecer cotas sobre las funciones de distribución generadas por una cadena. En particular, se puede estudiar la variación total de la distancia entre la distribución de la cadena en la iteración i y la distribución límite. También es

posible analizar aspectos particulares derivados de la estructura probabilística de las cadenas. Si bien el enfoque teórico es importante para el estudio de la convergencia, en la práctica los resultados teóricos pueden ser complicados de aplicar.

El segundo enfoque es a través de métodos gráficos. En este enfoque se analizan las propiedades de las observaciones obtenidas de una cadena para establecer la aproximación de esta a la distribución límite. Uno de los principales inconvenientes de esta perspectiva es que no es posible garantizar la convergencia. A pesar de esta complicación, el análisis de la convergencia desde la perspectiva gráfica es de gran utilidad en la práctica. Por tal motivo, este trabajo se centra en el desarrollo de las herramientas de este último enfoque. El libro de Gamerman *et al.* (2006) es una referencia recomendada para los lectores interesados en una explicación más detallada de los distintos métodos y enfoques en el análisis de convergencia.

Un aspecto importante de los métodos de MCMC es establecer la cantidad de ciclos o iteraciones que se deben realizar para que la cadena generada se aproxime a la distribución estacionaria. A esta cantidad de iteraciones del algoritmo se le conoce como *período de calentamiento*. La obtención del periodo de calentamiento y por lo tanto la convergencia de la cadena, puede obtenerse a través de algunas verificaciones de convergencia basadas en técnicas gráficas puestas por Gelfand y Smith (1990).

Uno de estos métodos consiste en realizar n simulaciones paralelas de la cadena que se desea estudiar y graficarlas durante un mismo número de iteraciones posteriormente y graficar los histogramas de los n valores obtenidos en la k -ésima iteración. A continuación se grafican los histograma de las m iteraciones posteriores a k . Generalmente, es necesario utilizar valores para m entre 10 y 50. Entonces, se acepta la convergencia de la cadena si no existe diferencias notables entre los histogramas obtenidos. De esta manera el valor de k será considerado como el periodo de calentamiento.

Otra técnica gráfica requiere la simulación de una única cadena. En este caso si dicha cadena a partir de una iteración k pesenta, cualitativamente, el mismo comportamiento; entonces, esto es un indicador de convergencia y por lo tanto k representa el periodo de calentamiento. Otro enfoque utiliza los promedios ergódicos para los cuales el periodo de calentamiento se es-

tablecerá como aquella iteración a partir de la cual los promedios ergódicos presentan un comportamiento aproximadamente asintótico.

Las conclusiones sobre la convergencia de la cadena, realizadas utilizando estos métodos gráficos, pueden ser reafirmadas si se realiza dicho análisis utilizando distintos valores iniciales para la cadena y seleccionando el mayor valor de iteraciones requeridas para poder afirmar que la cadena ha convergido.

Se debe resaltar que estas técnicas gráficas de análisis deben utilizarse con precaución, ya que la convergencia puede no ser notada o falsamente determinada dependiendo de la escala que se utilice para realizar las gráficas. Además existe diversas cadenas que pueden exhibir todos los indicadores de convergencia sin haberla alcanzado realmente, a estas cadenas se les conocen como *cadenas metástasis*.

Capítulo 3

Modelo longitudinal para datos circulares

3.1. Introducción a los estudios longitudinales

En el ámbito de la investigación experimental, en diversas áreas del conocimiento, es común encontrarse con una colección de datos obtenidos a partir de la medición de una misma característica varias veces para cada individuo (unidad de investigación). A este tipo de datos se le denominan *medidas repetidas* y los estudios que los utilizan forman un caso particular de los estudios multivariados, en los cuales diferentes características son medidas para un mismo sujeto. Cuando las mediciones son realizadas repetidamente a lo largo del tiempo y el tiempo constituye la dimensión sobre la cual las mediciones son obtenidas, a estos datos se les denominan *datos longitudinales*. Es decir, un estudio longitudinal es aquel en el cual una variable de respuesta es medida varias veces a lo largo del tiempo para cada sujeto.

La inclusión del tiempo como dimensión de medición hace que las mediciones repetidas de un estudio longitudinal tengan un orden establecido y por lo tanto no permite que estas puedan ser presentadas con un orden distinto. Debido a esto los estudios longitudinales poseen características que los diferencian de los demás estudios de medidas repetidas.

Una de las características principales de los datos longitudinales es que

permiten distinguir distintos tipos de variabilidad, debido a la existencia de réplicas y al concepto de distancia incorporado en ellos. De acuerdo con Diggle, Liang y Zeger (2002) es posible distinguir 3 componentes de variabilidad: efectos aleatorios, correlación serial y errores de medición. La primera se deriva de la variabilidades entre sujetos; es decir, de la heterogeneidad entre los individuos. La segunda componente se refiere a la variabilidad intra-sujeto o específica del individuo, la cual se presenta al poder existir una variación en el tiempo de la mediciones de cada sujeto. Por último, el proceso de medición por si mismo puede agregar variabilidad al modelo y por lo tanto los errores de medición constituirían la tercer componente de variabilidad.

Esta posibilidad de distinguir entre diferentes fuentes de variabilidad hace que los modelos longitudinales sean ricos en estructura. Esto se debe a que los modelos longitudinales poseen replicas en dos direcciones, por un lado se tienen sujetos y por el otro mediciones repetidas a lo largo del tiempo para cada individuo. Es decir, podemos considerar a los estudios longitudinales como una mezcla entre los estudios transversales y las series de tiempo. De esta forma, los datos longitudinales presentan diversas ventajas.

En comparación con los estudios transversales, una ventaja es que los estudios longitudinales tienen mayor poder estadístico. Es decir, para alcanzar un nivel de poder estadístico similar son necesarios menos sujetos en un estudio longitudinal; esto se debe a que las observaciones repetidas para un mismo sujeto rara vez están perfectamente correlacionadas y por lo tanto aportan mayor información independiente que una sola observación obtenida para un individuo.

Por otro lado, los estudios longitudinales permiten al investigador separar los cambios a través del tiempo para cada individuo de las diferencias entre los sujetos, lo que permite evitar confundir los efectos de la cohorte con los cambios que suceden dentro del proceso de cada individuo.

Por ultimo, es posible obtener información acerca de cambios individuales y con ello utilizar los estimadores de las tendencias individuales para entender mejor la heterogeneidad de la población, así como los factores de cambio y crecimiento a nivel individual.

Ahora bien, los estudios longitudinales no están exentos de tener com-

plicaciones y retos durante su análisis. Los retos más importantes que un investigador debe resolver al enfrentarse a un conjunto de datos longitudinales son: la presencia de datos faltantes, la correlación entre las observaciones y los cambios a lo largo del tiempo de los predictores.

Los datos longitudinales son más susceptibles a presentar datos faltantes. No es inusual que dentro del estudio longitudinal, por razones fuera del control del investigador, se presenten procesos de observación que concluyan antes del tiempo final estipulado para el estudio o mediciones intermedias faltantes. Además existe la posibilidad de mediciones intermedias faltantes. Esto plantea un reto particular para el desarrollo metodológico.

Con respecto a la correlación, cabe notar que las observaciones en un estudio longitudinal, por definición, no son independientes y por lo tanto se debe considerar métodos estadísticos que contemplen esta dependencia en los datos para poder realizar un análisis correcto. Esto se debe a que es de esperarse que exista un alto grado de correlación entre datos recolectados en una misma unidad a través del tiempo, y que generalmente hay mas variabilidad en las mediciones entre sujetos que las propias de un sujeto dado. La correlación se debe incluir en el modelo para poder evaluar adecuadamente la información que poseen los datos acerca de la respuesta media de la población y sobre los efectos que tienen las covariables sobre la media. Si se trataran erróneamente las observaciones realizadas a cada individuo como independientes, se podría sobrestimar o subestimar la fortaleza de las conclusiones obtenidas.

Adicionalmente a los desafíos anteriores en este tipo de datos es posible que no solo la variable respuesta cambie a través del tiempo; sino también que los predictores se modifiquen a lo largo del tiempo. El tratamiento adecuado de las covariables cambiantes en el tiempo en el análisis permite realizar inferencia estadística más precisa acerca de las relaciones dinámicas.

Resolver las cuestiones anteriores inherentes a los datos longitudinales conlleva a tener que desarrollar enfoques de análisis especiales. Los modelos utilizados tendrán una mayor complejidad y el análisis presentará mayores complicaciones que en aquellos estudios que utilizan datos transversales. A continuación se presenta un resumen de los distintos enfoques generales clásicos utilizados para el tratamiento de este tipo de datos.

3.2. Metodologías para el análisis de datos longitudinales

Si bien los estudios longitudinales poseen características particulares que les otorgan diversas ventajas, también representan retos en el desarrollo metodológico para su correcto análisis. La existencia de correlación entre la observaciones, así como la inevitable presencia de datos faltantes hacen que el modelo planteado tenga una mayor complejidad que aquellos utilizados para otro tipo de datos. Dicha complejidad hizo que el desarrollo metodológico especial para los datos longitudinales fuera escaso y por lo tanto el análisis se restringiera al uso clásico de las herramientas ya existentes.

Sin embargo con el surgimiento de la computadora, aparecieron distintos softwares capaces de lidiar con los retos numéricos necesarios para el uso de modelos longitudinales mas sofisticados.

Existen diversos enfoques para el análisis de datos longitudinales cuya variable de respuesta es escalar. Dentro de estos enfoques se destacan: el uso de variables derivadas, el modelo de análisis de varianza (ANOVA) para medidas repetidas, los modelos de análisis de varianza multivariados (MANOVA), patrones de covarianza, Ecuaciones estimadoras generalizadas (GEE) y los modelos de efectos mixtos. En la actualidad las GEE y los modelos de efectos mixtos son los más utilizados para realizar análisis longitudinal.

A continuación se presentan las características principales de estos enfoques, así como las limitaciones que tienen alguno de estos en el contexto longitudinal. Debido a que este trabajo centra parte de su desarrollo en el uso del modelo de efectos mixtos, se hará más hincapie en la exposición teórica de esta metodología. Para una explicación detalla de los distintos enfoques utilizados para el análisis longitudinal, el lector puede referirse a Hedeker y Gibbons (2006), Diggle *et al.* (2002), Verbeke y Molenberghs (2000) , Fitzmaurice *et al.* (2004) y Fitzmaurice *et al.* (2009).

El enfoque de variables derivadas se refiere a la reducción de las observaciones repetidas en una variable resumen. Estrictamente hablando, el análisis realizado a través de variables derivadas no es un análisis longitudinal, puesto que solo hay una sola medida por individuo. Esta simplificación de las obser-

vaciones tiene varios inconvenientes. Dada la característica de las variables derivadas, obliga a que el análisis se centre en un solo aspecto de las observaciones a lo largo del tiempo; es decir, puede existir pérdida de información al remplazar varias observación por una sola variable escalar. Por otro lado, es posible que sujetos con diferencias importantes en sus observaciones a lo largo del tiempo generen el mismo valor para la variable resumen. Otra desventaja de este enfoque es que las covariables deben ser consideradas invariantes en el tiempo, ya que la variable resumen elimina el aspecto temporal de los datos.

Los modelos de ANOVA para medidas repetidas y los modelos basados en MANOVA son dos enfoques clásicos para el estudio de datos longitudinales. Para ambos modelos el objetivo principal es la comparación de las medias de los grupos y no proporcionan información sobre tendencias específicas de los individuos. Además, los puntos en el tiempo se asumen fijos a través de los sujetos y por lo tanto se asumen como variables de clasificación. Estos enfoques se basan en la estimación por mínimos cuadrados y por lo tanto se ven afectados por los datos faltantes y outliers. El modelo de ANOVA para medidas repetidas es esencialmente un modelo que contempla un intercepto aleatorio que asume que los individuos se pueden desviar de la media total de la variable respuesta por una constante, la cual es equivalente a través del tiempo. Lo anterior hace que el modelo suponga una matriz de varianzas y covarianzas con simetría compuesta; es decir, asume que las varianzas y covarianzas son constantes en el tiempo. Tal supuesto hace que este tipo de modelos no sea apropiado para datos longitudinales debido a que, en general, la varianza se incrementa a lo largo del tiempo y la correlación decae a medida que la separación en el tiempo aumenta.

Otro enfoque utilizado en el análisis longitudinal es el análisis multivariado de varianza para medidas repetidas o MANOVA para medidas repetidas. En este modelo las respuestas son trabajadas como un vector multivariado. La principal ventaja de este enfoque es que no realiza supuestos restrictivos referentes a la covarianza entre las observaciones de cada individuo. Esta característica permite que las correlaciones tengan cualquier patrón posible y que las varianzas puedan cambiar a lo largo del tiempo. Sin embargo, este enfoque tiene importantes limitaciones para su aplicación a los datos longitudinales. La principal desventaja es que al ser un modelo multivariado se requieren datos completos. Individuos con datos incompletos son removidos completamente del análisis, lo cual representa un manejo ineficiente de los

datos disponible. Además, la eliminación de datos incompletos puede significar pérdida de información y producir estimadores sesgados de los cambios a través del tiempo para la respuesta media. Por último, el modelo de MANOVA para medidas repetidas. Por último, la formulación del modelo no permite el uso de covariables que varíen a través del tiempo; es decir, obliga que las covariables utilizadas dentro de cada sujeto sean las mismas para todos los individuos.

El enfoque de patrón de covarianzas (CPM) es otra alternativa para el análisis de datos longitudinales. Consiste en la modelación directa de la estructura de varianzas y covarianzas de las observaciones repetidas y, en general, no tiene como objetivo distinguir entre la varianza interna de cada sujeto de la variabilidad existente entre los individuos. Es decir, este enfoque supone que la matriz de varianzas y covarianza es de una forma particular y no el resultado de haber incluido efectos aleatorios individuales. Al igual que el enfoque de MANOVA, los CPM admiten formas generales para dicha matriz y el tiempo es considerado de manera fija y de tipo categórico. Sin embargo, este enfoque sí permite sujetos que tengan datos incompletos a lo largo de los puntos fijados en el tiempo. Algunos ejemplos de este enfoque son las estructuras de simetría compuesta, estructuras autoregresivas de primer orden, estructuras Toeplitz o de bandas, estructuras de efectos aleatorios y formas sin estructuras.

Los modelos de ecuaciones estimadoras generalizadas (GEE), esencialmente, son una extensión de los modelos lineales generalizados para el caso de datos correlacionados. En consecuencia, esta clase de modelos se ha vuelto popular, en particular para el análisis de respuestas categóricas, aunque pueden ser utilizados para variables de respuestas continuas. Los modelos GEE se basan en una estimación de quasi-verosimilitud. Además estos son clasificados como modelos marginales y modelan la regresión de y sobre x y la dependencia intra-sujeto separadamente. El término “marginal” en este contexto indica que la modelación de respuesta media depende únicamente de las covariables de interés y esta no se ve afectada por los efectos aleatorios o respuestas anteriores.

Una característica fundamental de los modelos GEE es que la distribución conjunta de los vectores de respuesta de cada sujeto no necesitan ser especificadas. En su lugar, únicamente es necesario especificar la distribución

marginal del vector de respuesta para cada punto en el tiempo. Así, los GEE evitan el uso de distribuciones multivariadas solo asumiendo una forma funcional para la distribución marginal en cada punto del tiempo, lo cual hace que no se requieran evaluaciones numéricas complejas de la verosimilitud para modelos no lineales.

Una de las principales desventajas de las ecuaciones estimadoras generalizadas es el manejo de datos faltantes. Los datos faltantes son ignorables si estos son explicados por las covariables. Para los métodos GEE, los datos faltantes deben ser completamente aleatorios, lo que significa que la probabilidad de que una observación falte es independiente de los valores de las demás observaciones y de los valores de otras variables. Es difícil imaginar que una respuesta para un individuo que se hubiera obtenido después de su retiro fuera independiente de las respuestas observadas durante el estudio. Esta característica limita, en cierta medida, la aplicación de este enfoque para datos longitudinales incompletos.

Finalmente, el enfoque de modelos de regresión de efectos-mixtos generalizados o modelos de efectos mixtos, son los más utilizados para el análisis longitudinal. Este tipo de modelos son uno de los métodos más generales para el análisis de datos longitudinales que se pueden aplicar tanto a variables continuas con distribución normal, como a respuestas categóricas. Lo anterior se debe a que los modelos de efectos mixtos son sumamente robustos con respecto a los datos faltantes, y a ocasiones de medición irregularmente espaciadas; además pueden fácilmente manejar tanto covariables variantes e invariantes en el tiempo. Los modelos de efectos mixtos son considerados métodos de “verosimilitud completa” porque hacen uso completo de los datos disponibles para cada sujeto. La ventaja es que los datos faltantes son ignorables si las respuestas faltantes pueden explicarse ya sea por las covariables o por las respuestas disponibles de un individuo dado. Sin embargo, los métodos de verosimilitud completa tienen la desventaja de ser computacionalmente más complejos que aquellos métodos de pseudo-verosimilitud, tales como los GEE.

En este trabajo se busca desarrollar un modelo para el análisis longitudinal de datos circulares basado en la distribución normal proyectada, presentada anteriormente, y en el cual se incorporan modelos lineales de efectos mixtos. Por esta razón, a continuación se desarrollan algunos modelos de efectos mixtos

3.3. Modelos de efectos mixtos

Aunque el desarrollo inicial de los modelos de efectos mixtos (MRM) para datos jerárquicos o agrupados se remonta hasta el paradigma de ANOVA y al fundamental artículo de Harville (1977), su aplicación a los datos longitudinales fue resaltado hasta la principios de 1980 en un artículo de Laird y Ware (1982).

Una característica básica de los modelos de efectos mixtos es que estos incluyen efectos aleatorios individuales para poder explicar la influencia de los individuos en las repeticiones de sus observaciones. Además, estos efectos aleatorios introducen el grado de variación individual en la población. De esta forma, estos modelos pueden describir la dependencia de los datos y distinguir tendencia en el tiempo para diferentes individuos.

Como introducción a los modelos de efectos mixtos, considérese el modelo de regresión para la observación y del individuo i en la ocasión j , donde $i = 1, 2, \dots, N$ y $j = 1, 2, \dots, n_i$, dado por:

$$y_{ij} = \beta_0 + \beta_1 t_{ij} + \varepsilon_{ij} \quad (3.1)$$

La variable independiente t le da valores a los niveles de tiempo y puede representar tiempo en semanas, meses, años, etc. Como se puede observar, los subíndices permiten dar seguimiento a los detalles de los datos, como a quien corresponde una observación dada y en qué tiempo u ocasión fue realizada. Nótese que tanto y como t llevan los subíndices i y j , lo cual permite que ambas varíen entre individuos y entre ocasiones.

En este modelo de regresión, los errores ε_{ij} se suponen normalmente distribuidos e independientes con media 0 y varianza común σ^2 . El supuesto de independencia es poco sensato en el contexto de los datos longitudinales, ya que las respuestas obtenidas y son observadas repetidamente para un mismo individuo y por lo tanto es de esperarse que los errores específicos de dicho individuo estén correlacionados. Además este modelo considera que el cambio a través del tiempo es el mismo para todos los sujetos, puesto que los parámetros β_0 (intercepto) y β_1 (pendiente) varían entre individuos. Estas condiciones hacen que este modelo sea poco razonable para el análisis de datos longitudinales.

Modelo de intercepto aleatorio

La primer extensión del modelo anterior consiste en permitir la influencia de cada individuo en sus mediciones repetidas. Por lo tanto este nuevo modelo está dado por

$$y_{ij} = \beta_0 + \beta_1 t_{ij} + b_{0i} + \varepsilon_{ij} \quad (3.2)$$

donde b_{0i} representa esa influencia del sujeto i sobre sus observaciones. En general, b_{0i} se desviara de 0 ya que es muy probable que los sujetos tengan una influencia positiva o negativa sobre sus datos longitudinales.

En este modelo los parámetros específicos para cada individuo son b_{0i} y los errores de medición ε_{ij} ; mientras que β_0 y β_1 son parámetros generales para todos los individuos y por lo tanto no incluyen varianciones entre los individuos. Por lo anterior, es posible distinguir las características específicas de cada individuo de los comportamiento generales para toda la población estudiada.

El nivel inicial para el i -ésimo individuo esta determinado por β_0 y b_{0i} . En este caso β_0 representa el nivel inicial para la población y b_{0i} es la contribución individual a su nivel inicial, la cual es invariante a través del tiempo. En consecuencia cada individuo posee su propio nivel inicial distinto al de los demás. Por el contrario, no existe distinción en la pendiente de las líneas de tendencias para cada individuo debido a que en todos los casos están determinada por β_1 , cuyo valor es el mismo para todos los individuos. Así, es posible pensar que las líneas de tendencia para cada sujeto son paralelas a la tendencia de la población determinada por β_0 y β_1 , y que la diferencia entre estas es b_{0i} . Por esta razón, a este modelo se le denomina *modelo de intercepto aleatorio*.

Para un mejor entendimiento de este modelo es posible representarlo jerárquicamente. Para esto, se divide en: Nivel 1 o intra-sujeto,

$$y_{ij} = v_{0i} + v_{1i} t_{ij} + \varepsilon_{ij} \quad (3.3)$$

y Nivel 2 o entre-sujetos,

$$v_{0i} = \beta_0 + b_{0i} \quad (3.4)$$

$$v_{1i} = \beta_1 \quad (3.5)$$

De esta manera la estructura de la línea de tendencia para el individuo i resulta más clara. En este caso el intercepto está determinado por v_{0i} y la pendiente está dada por v_{1i} . También es posible observar la composición tanto del intercepto como de la pendiente, donde se puede precisar que el intercepto es la única componente donde existe una contribución individual al comportamiento de la tendencia.

La representación jerárquica genera dos tipos de respuestas: intra-sujeto y entre-sujeto. Con esto es posible agregar covariables intra-sujeto que expliquen la variación de las respuestas en el Nivel 1 (y_{ij}) y, de la misma forma, agregar covariables entre-sujeto que expliquen la variación de las respuestas del Nivel 2 (el intercepto del individuo v_{0i} y la pendiente b_{1i}).

Puesto que los individuos en una muestra son pensados típicamente ser representativos de una población de individuos mayor, los efectos específicos de los individuos b_{0i} son considerados ser representativos de una distribución de efectos individuales en la población. Es decir, los b_{0i} son considerados como efectos aleatorios. La forma típica para esta distribución de efectos individuales es la distribución normal con media 0 y varianza σ_b^2 . Por otro lado, en el modelo de intercepto aleatorio los errores ε_{ij} se suponen normales distribuidos con media 0 y varianza σ^2 . Además los efectos individuales b_{0i} y los errores ε_{ij} se suponen independientes uno de otro.

Para este modelo se tiene que la varianza marginal de cada respuesta está dada por

$$\begin{aligned} \text{Var}(Y_{ij}) &= \text{Var}(\beta_0 + \beta_1 t_{ij} + b_{0i} + \varepsilon_{ij}) \\ &= \text{Var}(b_{0i} + \varepsilon_{ij}) \\ &= \text{Var}(b_{0i}) + \text{Var}(\varepsilon_{ij}) \\ &= \sigma_b^2 + \sigma^2 \end{aligned}$$

Por otro lado, la covarianza marginal entre dos respuestas de un mismo in-

dividuo, Y_{ij} y Y_{ik} esta dada por

$$\begin{aligned}
 Cov(Y_{ij}, Y_{ik}) &= Cov(\beta_0 + \beta_1 t_{ij} + b_{0i} + \varepsilon_{ij}, \beta_0 + \beta_1 t_{ik} + b_{0i} + \varepsilon_{ik}) \\
 &= Cov(b_{0i} + \varepsilon_{ij}, b_{0i} + \varepsilon_{ik}) \\
 &= Cov(b_{0i}, b_{0i}) \\
 &= Var(b_{0i}) \\
 &= \sigma_b^2
 \end{aligned}$$

Dado lo anterior, en el modelo de intercepto aleatorio la matriz de covarianzas de las medidas repetidas generada por el modelo de intercepto aleatorio es una matriz de simetría compuesta. Así pues, se observa que la inclusión de un efecto aleatorio para el intercepto induce correlación entre las medidas repetidas. Sin embargo, dicha correlación es constante para toda pareja de observaciones.

Por lo tanto, si bien el modelo de intercepto aleatorio es el ejemplo más simple de modelos de efectos mixtos, este no es completamente adecuado para realizar análisis de datos longitudinales. No obstante, es posible generalizar las ideas básicas de este modelo para disponer de un modelo más versátil para realizar análisis longitudinal.

Modelo intercepto y pendiente aleatorios

En el siguiente modelo se permite la variación aleatoria de coeficientes de regresión adicionales a los estipulados para el modelo anterior. Con la introducción de aleatoriedad a otros coeficientes se logra obtener un modelo más flexible, pero siendo este aún suficientemente parsimonioso. A continuación se detallan las características de este modelo, conocido como *modelo de intercepto y pendiente aleatoria*.

El modelo de intercepto y pendiente aleatoria es un modelo de efectos mixtos, en el cual el intercepto y la pendiente pueden variar aleatoriamente entre los individuos. Asumiendo que hay N individuos para los cuales se han recolectado n_i observaciones, con $i = 1, 2, \dots, N$; entonces para el i -ésimo sujeto en su j -ésima observación se tiene que:

$$Y_{ij} = \beta_0 + \beta_1 t_{ij} + b_{0i} + b_{1i} t_{ij} + \varepsilon_{ij} \quad (3.6)$$

o reagrupando de manera distinta se obtiene

$$Y_{ij} = (\beta_0 + b_{0i}) + (\beta_1 + b_{1i})t_{ij} + \varepsilon_{ij} \quad (3.7)$$

Se puede observar que con esta expresión cada sujeto varía tanto en sus niveles iniciales, dados por $\beta_0 + b_{0i}$, como en la tendencia de sus observaciones, determinada por la pendiente $\beta_1 + b_{1i}$. Lo anterior puede ser observado con mayor facilidad al reexpresar la ecuación anterior en su forma jerárquica, donde se tiene que el Nivel 1 esta dado por

$$y_{ij} = v_{0i} + v_{1i}t_{ij} + \varepsilon_{ij} \quad (3.8)$$

y el Nivel 2 se define como

$$v_{0i} = \beta_0 + b_{0i} \quad (3.9)$$

$$v_{1i} = \beta_1 + b_{1i} \quad (3.10)$$

Este este caso, el modelo tiene dos efectos aleatorios específicos para cada sujeto, entiéndase b_{0i} y b_{1i} , y posee dos parámetros relacionados con el intercepto y la pendiente general de la población, β_0 y β_1 respectivamente. Entonces es posible definir un vector formado por los efectos aleatorios del individuo y otro para los parámetros relacionados con la población, denominados efectos fijos.

De esta manera es posible utilizar notación de vectores y matrices para expresar el modelo como

$$\mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\varepsilon}_i \quad (3.11)$$

donde se tiene que

$$\mathbf{Y}_i = \begin{pmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{in_i} \end{pmatrix} \quad \text{y} \quad \mathbf{X}_i = \mathbf{Z}_i = \begin{pmatrix} 1 & t_{i1} \\ 1 & t_{i2} \\ \vdots & \vdots \\ 1 & t_{in_i} \end{pmatrix}$$

y los vectores $\boldsymbol{\beta}$ y \mathbf{b}_i definidos como

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} \quad \text{y} \quad \mathbf{b}_i = \begin{pmatrix} b_{0i} \\ b_{1i} \end{pmatrix}$$

Además, para este modelo se supone que el vector de efectos aleatorios se distribuye Normal bivariado con media 0 y matriz de covarianzas $\mathbf{\Omega}$. De la misma manera, los errores $\boldsymbol{\varepsilon}_i$ tienen una distribución normal multivariada con media 0 y una matriz de covarianzas R_i . Sin embargo, suele asumirse que la matriz R_i es la matriz diagonal $\sigma^2 I_{n_i}$. Con esta matriz se tiene que ε_{ij} y ε_{ik} no están correlacionados, con varianza igual y por lo tanto los errores ε_{ij} pueden ser pensados como una muestra de los errores de medición.

Modelo de efectos mixtos

La expresión matricial para el modelo de intercepto y pendiente aleatoria puede ser extendido, modificando únicamente las dimensiones de los vectores y matrices, para incorporar más covariables, efectos fijos y efectos aleatorios. Por esta razón dicha expresión es la representación general de cualquier modelo de efectos mixtos.

Un modelo de efectos mixtos, de gran interés para el desarrollo de este trabajo, es el modelo de efectos mixtos de dos etapas propuesto por Laird y Ware (1982). Este modelo incluye como caso especial a los modelos para datos longitudinales. En la primera etapa se introducen los parámetros poblacionales, efectos individuales y variaciones intra-sujeto; y en la segunda etapa son introducidas las variaciones entre individuos.

Para este modelo se considera a \mathbf{Y}_i como el vector de dimensión $n_i \times 1$ formado por las n_i observaciones correspondientes al i -ésimo individuo. La matriz \mathbf{X}_i es una matriz de diseño que contiene las covariables, la cual es de dimensión $n_i \times p$. Por otro lado, el vector $\boldsymbol{\beta}$, de dimensión $p \times 1$, representa los efectos fijos y el vector \mathbf{b}_i , $q \times 1$, representa los efectos aleatorios individuales del sujeto i . por último se considera la matriz \mathbf{Z}_i , la cual es una matriz de diseño que relaciona los efectos aleatorios individuales \mathbf{b}_i con las respuestas y_i . Así, la especificación del modelo esta dada por:

Etapas 1: Para cada individuo, i ,

$$\mathbf{Y}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{b}_i + \boldsymbol{\varepsilon}_i \quad (3.12)$$

donde $\boldsymbol{\varepsilon}_i$ esta distribuido como $N_{n_i}(0, R_i)$. La matriz R_i es una matriz de covarianzas positiva definida de $n_i \times n_i$. Esta matriz depende de i a través de

la dimensión, pero los parámetros desconocidos en R_i no tienen dependencia de i . En esta etapa, tanto β como \mathbf{b}_i son considerados fijos y el vector $\boldsymbol{\varepsilon}_i$ se asume ser independiente.

Etapla 2: Los vectores \mathbf{b}_i son distribuidos como $N_q(0, \boldsymbol{\Omega})$, independientes entre ellos y de los vectores $\boldsymbol{\varepsilon}_i$. Los parámetros de la población, β , son tratados como efectos fijos.

Con el modelo de efectos mixtos es posible distinguir la media condicional o media específica del individuo de \mathbf{Y}_i , dado \mathbf{b}_i , de la media marginal de \mathbf{Y}_i . La media condicional es:

$$E(\mathbf{Y}_i|\mathbf{b}_i) = \mathbf{X}_i\beta + \mathbf{Z}_i \mathbf{b}_i$$

y la media marginal del vector de observaciones, promediada sobre la distribución de los efectos aleatorios, es

$$\begin{aligned} E(\mathbf{Y}_i) &= E(E(\mathbf{Y}_i|\mathbf{b}_i)) \\ &= E(\mathbf{X}_i\beta + \mathbf{Z}_i \mathbf{b}_i) \\ &= \mathbf{X}_i\beta + \mathbf{Z}_i E(\mathbf{b}_i) \\ &= \mathbf{X}_i\beta \end{aligned}$$

De la manera similar, es posible distinguir entre la covarianza condicional y marginal. La covarianza de \mathbf{Y}_i , dados \mathbf{b}_i , es

$$Cov(\mathbf{Y}_i|\mathbf{b}_i) = Cov(\boldsymbol{\varepsilon}_i) = R_i$$

mientras que la covarianza marginal de \mathbf{Y}_i es

$$\begin{aligned} Cov(\mathbf{Y}_i) &= Cov(\mathbf{Z}_i \mathbf{b}_i) + Cov(\boldsymbol{\varepsilon}_i) \\ &= \mathbf{Z}_i Cov(\mathbf{b}_i) \mathbf{Z}_i' + Cov(\boldsymbol{\varepsilon}_i) \\ &= \mathbf{Z}_i \boldsymbol{\Omega} \mathbf{Z}_i' + R_i \end{aligned}$$

Incluso cuando $R_i = \sigma^2 I_{n_i}$, se tiene que

$$Cov(\mathbf{Y}_i) = \mathbf{Z}_i \boldsymbol{\Omega} \mathbf{Z}_i' + \sigma^2 I_{n_i}$$

es una matriz que no es diagonal. Es decir, $Cov()$ tendrá, en general, elementos distintos de 0 fuera de la diagonal y de este modo registra la correlación

entre las observaciones repetidas sobre los mismos individuos. En consecuencia la introducción de los efectos aleatorios, \mathbf{b}_i , permite inducir correlación entre las componentes del vector de observaciones \mathbf{Y}_i .

Ahora bien, otra propiedad del modelo de efectos mixtos es que la covarianza de \mathbf{Y}_i se define a partir de las matrices $\mathbf{\Omega}$ y de la matriz R_i . Es decir, que dicha covarianza esta descrita en términos de un conjunto de parámetros de covarianza. Así, el modelo de efectos mixtos permite realizar un análisis explícito de las fuentes de variación entre sujetos e intra-sujetos en las respuestas.

3.4. Propuesta de modelo para datos longitudinales circulares

Como ya se ha mencionado, el análisis de datos longitudinales ha sido abordado desde diversos enfoques. Existe bastante literatura dedicada a detallar las diferentes propuestas metodológicas más adecuadas para el tratamiento de datos longitudinales; sin embargo, el desarrollo metodológico para datos longitudinales se ha centrado en respuestas escalares y el caso de respuestas direccionales ha sido poco estudiado.

El establecimiento de los fundamentos teóricos para el análisis de datos longitudinales direccionales se ha visto limitado, debido en gran parte a la carencia de literatura que permita su desarrollo. Esto se debe, entre otras cosas, a que las distribuciones de probabilidad asociadas a este tipo de datos representan grandes dificultades en su manejo. Los datos direccionales ha sido tratados, básicamente con los enfoques de GEE (Artes *et al.* 2000; Artes y Jørgensen, 2000), a través de modelos lineales generalizados (Song, 2007; D'Elia *et al.* 2001). Los procedimientos de análisis poseen ciertas fallas que ocasionan problemas con la inferencia en configuraciones generales. Estos problemas incluyen complicaciones en el ajuste, comparación de modelos y predicción, así como también problemas de convergencia en los métodos iterativos utilizados.

Recientemente Nuñez Antonio y Gutiérrez-Peña (2014) propusieron un modelo bayesiano basado en la Normal Proyectada para el análisis de datos

circulares (PCL). El modelo considera covariables lineales y se basa en una versión de la distribución normal proyectada bivariada. Además cada componente del modelo esta especificada por un modelo de efectos mixtos bajo el supuesto de independencia condicional sobre cada una de las componentes.

Con el modelo PCL es posible describir distintos patrones longitudinales de datos circulares. Dentro de los patrones posibles se encuentra el de intercepto aleatorio, intercepto y pendiente aleatoria, similares a los obtenidos con los modelos de efectos mixtos estándares. Además es posible producir patrones de dependencia más generales y permite distinguir variaciones entre sujetos y variaciones intra-sujetos, descomponiendo la matriz de covarianza en dichas variaciones.

A continuación se construye un modelo para datos longitudinales con variable de respuesta direccional basado en el enfoque PCL. El objetivo de dicho modelo es extender el modelo original, propuesto por Nuñez Antonio y Gutiérrez-Peña (2014), a través de la inclusión de varianzas generales en los errores de la segunda componente. Lo cual permite describir una mayor diversidad de patrones de datos longitudinales.

Modelo PCL con la inclusión de λ

Supóngase que en cada una de las j ocasiones ($j = 1, \dots, n_i$, se toman mediciones de ángulos para el i -ésimo individuo ($i = 1, \dots, N$), formando un vector de respuestas $\theta_i = (\theta_{i1}, \dots, \theta_{in_i})$.

El primer paso en la construcción del modelo es considerar los datos aumentados, haciendo uso de variables latentes R_{ij} de tal manera que a partir de cada observación θ_{ij} se obtenga un vector bivariado Y_{ij} . Entonces los datos aumentados están dados por

$$Y_{ij} = \begin{pmatrix} Y_{ij}^I \\ Y_{ij}^{II} \end{pmatrix} = R_{ij} \begin{pmatrix} \cos\theta_{ij} \\ \text{sen}\theta_{ij} \end{pmatrix}$$

donde $\|Y_{ij}\| = R_{ij}$.

Se propone que el vector Y_{ij} se distribuya $N_2(\mu_{ij}, \Sigma)$, con $\Sigma = \begin{pmatrix} 1 & 0 \\ 0 & \lambda \end{pmatrix}^{-1}$

y para μ_{ij} se propone la siguiente estructura:

$$\mu_{ij} = \begin{pmatrix} \mu_{ij}^I \\ \mu_{ij}^{II} \end{pmatrix} = \begin{pmatrix} (x_{ij}^I)' \beta^I + (z_{ij}^I)' \mathbf{b}_i^I \\ (x_{ij}^{II})' \beta^{II} + (z_{ij}^{II})' \mathbf{b}_i^{II} \end{pmatrix}$$

Aquí, los vectores $\{\beta^I, \beta^{II}\}$ son coeficientes del modelo y $\{\mathbf{b}_i^I, \mathbf{b}_i^{II}\}$ representan los efectos aleatorios individuales, que se asumen distribuidos $N(0, \mathbf{\Omega}^{-1})$. Cabe destacar que, en general, cada componente de μ_{ij} puede depender de diferentes cantidades de covariables y de efectos aleatorios, por lo que las dimensiones de esos vectores podría ser distinta en cada entrada.

De esta manera se tiene que la especificación PCL con matriz $\mathbf{\Sigma}$, contiene en cada componente un modelo de efectos mixtos. Por lo tanto, definiendo $\sigma^2 = \frac{1}{\lambda}$ se tiene que

$$\begin{aligned} \mathbf{Y}_i^I &= \mathbf{X}_i^I \beta^I + Z^I \mathbf{b}_i^I + \boldsymbol{\varepsilon}_i^I \\ \mathbf{Y}_i^{II} &= \mathbf{X}_i^{II} \beta^{II} + Z^{II} \mathbf{b}_i^{II} + \boldsymbol{\varepsilon}_i^{II} \end{aligned}$$

donde

$$\begin{aligned} \boldsymbol{\varepsilon}_i^{II} &\sim N_{n_i}(0, I_{n_i}) \\ \boldsymbol{\varepsilon}_i^{II} &\sim N_{n_i}(0, \sigma^2 I_{n_i}) \end{aligned}$$

Nótese que \mathbf{Y}_i^I y \mathbf{Y}_i^{II} son vectores de dimensión n_i , para toda $i = 1, 2, \dots, N$.

Ahora bien, para utilizar el enfoque bayesiano es necesario definir distribuciones iniciales para β^k y $\mathbf{\Omega}^k$, $k \in \{I, II\}$. Por lo tanto, considerando cada componente por separado, la especificación jerárquica del modelo propuesto

PCL con matriz $\mathbf{\Sigma} = \begin{pmatrix} 1 & 0 \\ 0 & \lambda \end{pmatrix}^{-1}$ es

- **Nivel I:** Para cada individuo i

$$\begin{aligned} \mathbf{Y}_i^I | \beta^I, \{\mathbf{b}_i^I\} &\sim N_{n_i}(\mathbf{X}_i^I \beta^I + Z^I \mathbf{b}_i^I, I_{n_i}) \\ \mathbf{Y}_i^{II} | \beta^{II}, \{\mathbf{b}_i^{II}\} &\sim N_{n_i}(\mathbf{X}_i^{II} \beta^{II} + Z^{II} \mathbf{b}_i^{II}, \sigma^2 I_{n_i}) \end{aligned}$$

- **Nivel 2:** Dado $k \in \{I, II\}$, los vectores β^k y \mathbf{b}_i^k se consideran independientes con

$$\begin{aligned}\mathbf{b}_i^k | \Omega^k &\sim N_{q^k}(0, \Omega^k) \\ \beta^k &\sim N_{p^k}(0, A^k) \\ \lambda &\sim \text{Gamma}(\alpha, \gamma)\end{aligned}$$

- **Nivel 3:** Dado $k \in \{I, II\}$,

$$\Omega^k \sim \text{Wi}(v^k, B^k), \quad q^k \leq v^k$$

donde q^k es la dimensión del vector \mathbf{b}_i^k . Bajo esta parametrización de la distribución Wishart se tiene que $E(\Omega^k) = v^k (B^k)^{-1}$.

3.5. Análisis descriptivo del modelo

El modelo PCL introducido por Nuñez Antonio y Gutierrez-Peña permite modelar diversos comportamientos de dependencia longitudinal. Dentro de los patrones que pueden ser obtenidos se encuentran los comportamientos clásicos de los modelos de efectos mixtos de intercepto aleatorio e intercepto y pendiente aleatoria. Los autores destacan que con el modelo PCL es posible, además, obtener dependencias más generales.

Como ya se ha mencionado ese modelo utiliza la identidad como matriz de covarianzas y por lo tanto la matriz de covarianza para cada componente del vector de respuestas \mathbf{Y}_i esta dada por

$$\text{Var}(\mathbf{Y}_i) = \mathbf{Z}_i \Omega^{-1} \mathbf{Z}_i' + I_{n_i}$$

Entonces, patrones de covarianzas pueden ser generados solamente mediante la especificación de las matrices \mathbf{Z}_i y Ω , ya que en este caso la aportación de los errores de medición a la variabilidad total será constante para cualquier caso.

Ahora bien, el modelo propuesto en este trabajo permite que $\boldsymbol{\varepsilon}_i^{II}$ tenga una varianza general, $\sigma^2 I$. La introducción de este parámetro σ^2 permite

tener una mayor variedad de patrones para la matriz de covarianzas dado que para la segunda componente del vector \mathbf{Y}_i se tendría

$$\text{Var}(\mathbf{Y}_i) = \mathbf{Z}_i \boldsymbol{\Omega}^{-1} \mathbf{Z}_i' + \sigma^2 I_{n_i}$$

Es decir que, bajo una misma especificación de las matrices \mathbf{Z}_i y $\boldsymbol{\Omega}$, el modelo PCL con $\sigma^2 I_{n_i}$ es capaz de generar una estructura de patrones más general. Sin embargo, al utilizar una matriz diagonal las covarianzas de la matriz de covarianzas de la segunda componente no se ven alteradas por la inclusión de este parámetro.

Para ejemplificar lo anterior considérese un modelo con intercepto y pendiente aleatoria para la segunda componente y 3 observaciones. Entonces, omitiendo el superíndice para facilitar la lectura, la matriz \mathbf{Z}_i sería

$$\mathbf{Z}_i = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{pmatrix}$$

y definiendo la matriz $\boldsymbol{\Omega}^{-1}$ para dicha componente como

$$\boldsymbol{\Omega}^{-1} = \begin{pmatrix} \omega_1^2 & \omega_{12} \\ \omega_{21} & \omega_2^2 \end{pmatrix}$$

se tiene que la matriz de covarianzas del vector de respuestas para el individuo i , \mathbf{Y}_i , estaría dada por:

$$\text{Var}(\mathbf{Y}_i) = \begin{pmatrix} \omega_1^2 + 1 & \omega_1^2 + \omega_{12} & \omega_1^2 + 2\omega_{12} \\ \omega_1^2 + \omega_{12} & \omega_1^2 + 2\omega_{12} + \omega_2^2 + 1 & \omega_1^2 + 3\omega_{12} + 2\omega_2^2 \\ \omega_1^2 + 2\omega_{12} & \omega_1^2 + 3\omega_{12} + 2\omega_2^2 & \omega_1^2 + 4\omega_{12} + 4\omega_2^2 + 1 \end{pmatrix}$$

para el caso del modelo PCL con vector de errores $\boldsymbol{\varepsilon}_i \sim N(0, I_3)$ y

$$\text{Var}(\mathbf{Y}_i) = \begin{pmatrix} \omega_1^2 + \sigma^2 & \omega_1^2 + \omega_{12} & \omega_1^2 + 2\omega_{12} \\ \omega_1^2 + \omega_{12} & \omega_1^2 + 2\omega_{12} + \omega_2^2 + \sigma^2 & \omega_1^2 + 3\omega_{12} + 2\omega_2^2 \\ \omega_1^2 + 2\omega_{12} & \omega_1^2 + 3\omega_{12} + 2\omega_2^2 & \omega_1^2 + 4\omega_{12} + 4\omega_2^2 + \sigma^2 \end{pmatrix}$$

para el modelo con vector de errores $\boldsymbol{\varepsilon}_i \sim N(0, \sigma^2 I_3)$.

De esta manera, dadas las matrices $Var(\mathbf{Y}_i)$ para cada modelo, es posible observar que la estructura de las covarianzas son idénticas y que las varianzas se ven modificadas por la introducción de σ^2 . Además, σ^2 puede tomar cualquier valor no negativo y por lo tanto es posible generar estructuras más flexibles para $Var(\mathbf{Y}_i)$.

Por otro lado, en ambos modelos se tiene que la media marginal del vector de respuestas para el individuo i está dada por $\mathbf{X}_i \boldsymbol{\beta}$ y la media condicional es $\mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{b}_i$. Entonces, se concluye que la introducción del parámetro σ^2 únicamente influye en el comportamiento de la varianza del vector de respuestas \mathbf{Y}_i .

Es importante resaltar que la variabilidad existente en una muestra de vectores bivariados $(Y_{ij}^I, Y_{ij}^{II}), \dots, (Y_{i n_i}^I, Y_{i n_i}^{II})$ no puede ser trasladada directamente como valor de la varianza de los datos circulares generados por su proyección radial. Esto se debe a que para realizar la proyección radial de cada uno de estos vectores (Y_{i1}^I, Y_{i1}^{II}) es necesario calcular $\arctan(Y_{ij}^{II}/Y_{ij}^I)$. Sin embargo, es posible observar que la modificación de las varianzas de los $\boldsymbol{\varepsilon}_i$ producen comportamientos longitudinales con una mayor dispersión y, en algunos casos con patrones distintos al modelo PCL original. A continuación se presentan algunos ejemplos que muestran las diferencia en los comportamientos longitudinales que se pueden obtener con ambas versiones del modelo PCL.

Ejemplo 3.5.1. Se considera una muestra simulada de 40 individuos con 5 observaciones. Para obtener la muestra se tomaron las siguientes especificaciones del modelo.

- Caso 1:

$$\begin{aligned} \mathbf{Y}_i^I | \boldsymbol{\beta}^I &\sim N_5(\mathbf{X}_i^I \boldsymbol{\beta}^I,) \\ \mathbf{Y}_i^{II} | \boldsymbol{\beta}^{II}, \{\mathbf{b}_i^{II}\} &\sim N_5(\mathbf{X}_i^{II} \boldsymbol{\beta}^{II} + \mathbf{Z}_i^{II} \mathbf{b}_i^{II}, I) \end{aligned}$$

- Caso 2: Para $\lambda = 0.001, 1, 10$, se define $\sigma^2 = \frac{1}{\lambda}$ y

$$\begin{aligned} \mathbf{Y}_i^I | \beta^I &\sim N_5(\mathbf{X}_i^I \beta^I, \sigma^2 I) \\ \mathbf{Y}_i^{II} | \beta^{II}, \{\mathbf{b}_i^{II}\} &\sim N_5(\mathbf{X}_i^{II} \beta^{II} + \mathbf{Z}_i^{II} \mathbf{b}_i^{II}, \sigma^2 I) \end{aligned}$$

donde para ambos caso se utilizan

$$\beta^I = \begin{pmatrix} 100 \\ -4 \end{pmatrix}, \quad \beta^{II} = \begin{pmatrix} 200 \\ -10 \end{pmatrix}$$

$$\mathbf{b}_i^{II} | \Omega^{II} \sim N_2(0, \Omega^{II})$$

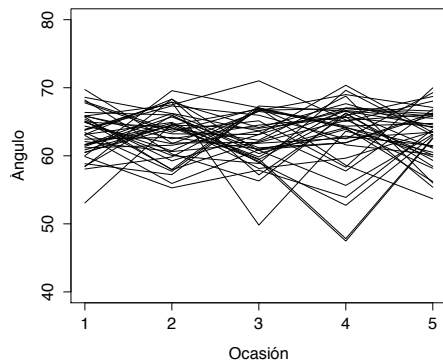
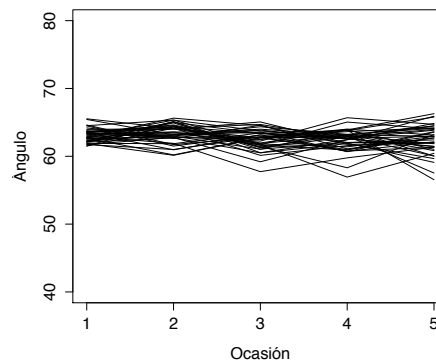
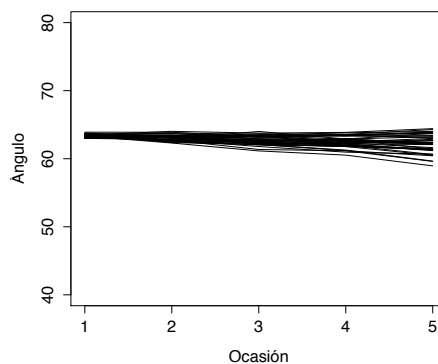
con

$$\Omega^{-1} = \begin{pmatrix} 0.001 & 0 \\ 0 & 5 \end{pmatrix}$$

Además, se especificaron las matrices \mathbf{X}_i^I , \mathbf{X}_i^{II} y Z_i de la siguiente manera

$$\mathbf{X}_i^I = \mathbf{X}_i^{II} = Z_i = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{pmatrix}$$

En las figuras 3.1, 3.2 y 3.3 se observan gráficamente los comportamientos para $\lambda = 0.001, 1, 10$. Se destaca que para $\lambda = 0.001$ el comportamiento longitudinal presenta mayor dispersión en comparación al obtenido con $\lambda = 1$. Por otro lado, en el caso de $\lambda = 10$ el comportamiento obtenido es muy similar al que se obtiene con el modelo PCL original. Así, esencialmente se obtiene dos patrones distintos para este ejemplo, uno para $\lambda = 0.001$ y otro para los casos de $\lambda = 1$ y $\lambda = 10$. En los últimos dos casos la diferencia en el patrón longitudinal es escasa, por lo que se considera que ambos generan el mismo comportamiento general.

Figura 3.1: $\lambda = 0.001$ Figura 3.2: $\lambda = 0.01$ Figura 3.3: $\lambda = 1$

En el siguiente ejemplo se agrega un modelo de efectos mixtos a la primera componente y se utiliza la misma especificación para la segunda componente. La única modificación entre los dos casos será la inclusión de los distintos valores de λ .

Ejemplo 3.5.2. Se considera una muestra simulada de 40 individuos con 5 observaciones. Para obtener la muestra se toman las siguientes especificaciones del modelo.

- Caso 1:

$$\begin{aligned}\mathbf{Y}_i^I | \beta^I, \{\mathbf{b}_i^I\} &\sim N_5(\mathbf{X}_i^I \beta^I + \mathbf{Z}_i^I \mathbf{b}_i^I, I) \\ \mathbf{Y}_i^{II} | \beta^{II}, \{\mathbf{b}_i^{II}\} &\sim N_5(\mathbf{X}_i^{II} \beta^{II} + \mathbf{Z}_i^{II} \mathbf{b}_i^{II}, I)\end{aligned}$$

- Caso 2: Para $\lambda = 0.001, 1, 10$, se define $\sigma^2 = \frac{1}{\lambda}$ y

$$\begin{aligned}\mathbf{Y}_i^I | \beta^I, \{\mathbf{b}_i^I\} &\sim N_5(\mathbf{X}_i^I \beta^I + \mathbf{Z}_i^I \mathbf{b}_i^I, I) \\ \mathbf{Y}_i^{II} | \beta^{II}, \{\mathbf{b}_i^{II}\} &\sim N_5(\mathbf{X}_i^{II} \beta^{II} + \mathbf{Z}_i^{II} \mathbf{b}_i^{II}, \sigma^2 I)\end{aligned}$$

donde para ambos caso y para $k \in \{I, II\}$, se utilizan

$$\begin{aligned}\beta^I &= \begin{pmatrix} 100 \\ -4 \end{pmatrix}, \quad \beta^{II} = \begin{pmatrix} 200 \\ -10 \end{pmatrix} \\ \mathbf{b}_i^k | \boldsymbol{\Omega}^k &\sim N_2(0, \boldsymbol{\Omega}^k)\end{aligned}$$

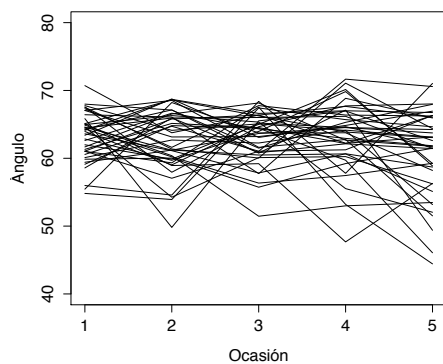
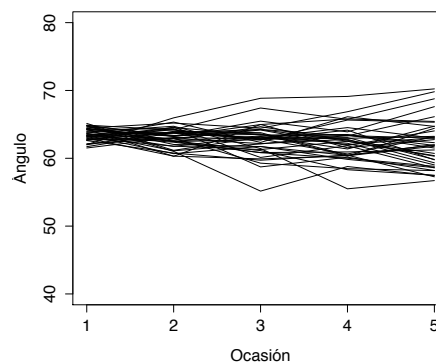
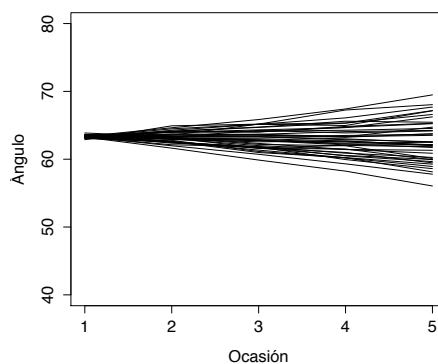
y con

$$(\boldsymbol{\Omega}^k)^{-1} = \begin{pmatrix} 0.001 & 0 \\ 0 & 5 \end{pmatrix}$$

Además se especificaron las matrices \mathbf{X}_i^I , \mathbf{X}_i^{II} y Z_i de la siguiente manera

$$\mathbf{X}_i^I = \mathbf{X}_i^{II} = Z_i^k = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{pmatrix}$$

Se puede notar en la Figura 3.4 que para el caso $\lambda = 0.001$ la modificación a la segunda componente permite tener aún más dispersión y generar patrones diferentes a los del Ejemplo 3.5.1. Los otros dos casos nuevamente son similares entre sí, sin poder distinguir una diferencia considerable en la dispersión de ambas.

Figura 3.4: $\lambda = 0.001$ Figura 3.5: $\lambda = 0.01$ Figura 3.6: $\lambda = 1$

Por último, en el siguiente ejemplo se utilizan matrices \mathbf{X}_i y \mathbf{Z}_i diferentes entre sí, para generar patrones longitudinales distintos a los que ya se han obtenido y poder observar las diferencias que se generan al modificar las varianzas de $\boldsymbol{\varepsilon}_i$.

Ejemplo 3.5.3. Se considera una muestra simulada de 40 individuos con 5 observaciones. Para obtener la muestra se tomaron las siguientes especificaciones del modelo.

- Caso 1:

$$\begin{aligned} \mathbf{Y}_i^I | \beta^I, \{\mathbf{b}_i^I\} &\sim N_5(\mathbf{X}_i^I \beta^I + \mathbf{Z}_i^I \mathbf{b}_i^I, I) \\ \mathbf{Y}_i^{II} | \beta^{II}, \{\mathbf{b}_i^{II}\} &\sim N_5(\mathbf{X}_i^{II} \beta^{II} + \mathbf{Z}_i^{II} \mathbf{b}_i^{II}, I) \end{aligned}$$

- Caso 2: Para $\lambda = 0.001, 1, 10$, se define $\sigma^2 = \frac{1}{\lambda}$ y

$$\begin{aligned} \mathbf{Y}_i^I | \beta^I, \{\mathbf{b}_i^I\} &\sim N_5(\mathbf{X}_i^I \beta^I + \mathbf{Z}_i^I \mathbf{b}_i^I, I) \\ \mathbf{Y}_i^{II} | \beta^{II}, \{\mathbf{b}_i^{II}\} &\sim N_5(\mathbf{X}_i^{II} \beta^{II} + \mathbf{Z}_i^{II} \mathbf{b}_i^{II}, \sigma^2 I) \end{aligned}$$

donde para ambos caso y para $k \in \{I, II\}$, se utilizan

$$\beta^I = \begin{pmatrix} 1 \\ -4 \end{pmatrix}, \quad \beta^{II} = \begin{pmatrix} 2 \\ -10 \end{pmatrix}$$

$$\mathbf{b}_i^k | \boldsymbol{\Omega}^k \sim N_2(0, \boldsymbol{\Omega}^k)$$

y con

$$(\boldsymbol{\Omega}^k)^{-1} = \begin{pmatrix} 0.001 & 0 \\ 0 & 5 \end{pmatrix}$$

y además,

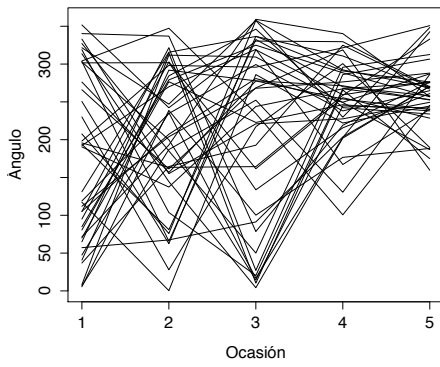
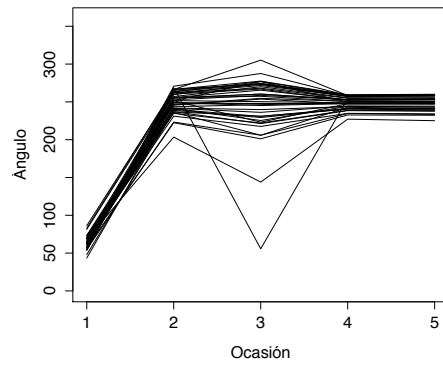
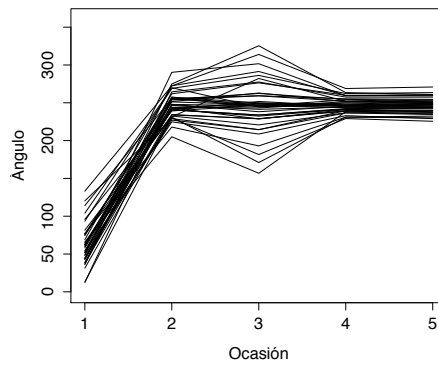
$$\mathbf{X}_i^I = \mathbf{X}_i^{II} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 1 \\ 1 & 4 \\ 1 & 5 \end{pmatrix} \quad \text{y} \quad \mathbf{Z}_i^k = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{pmatrix}$$

Las figuras 3.7, 3.8 y 3.9 muestran la diferencia entre los patrones longitudinales para estas especificaciones. En ellas es posible ver que el patrón generado por el uso del valor $\lambda = 0.001$ tiene, en términos generales, una variabilidad creciente al final de las observaciones. Por otro lado, se puede

observar que en las estructuras obtenidas a partir de los valores $\lambda = 1$ y $\lambda = 10$ es posible determinar características más detalladas de la tendencia creciente de estos. Además es posible notar un comportamiento general paralelo a partir de la segunda ocasión, presentando fluctuaciones en la tercera ocasión y una estabilización del comportamiento después de esta observación.

Con la especificación utilizada se puede apreciar la diferencia en la dispersión de las respuestas repetidas entre ambos casos simulados. Al utilizar el valor de 10 para λ se obtiene una mayor concentración de las líneas que representan las observaciones de cada individuo, lo cual está de acuerdo con la explicación dada previamente con respecto a la modificación a la variabilidad que induce el valor de $\sigma^2 = \frac{1}{\lambda}$.

Una comparación entre los ejemplos desarrollados nos permite observar que la especificación de los vectores y matriz que integran el modelo, otorgan los rasgos característicos del comportamiento longitudinal. Por otro lado la especificación de las componentes \mathbf{Y}_i y de los errores $\boldsymbol{\varepsilon}_i$ detallan las particularidades de cada modelo, al permitir una mayor variabilidad en las observaciones.

Figura 3.7: $\lambda = 0.001$ Figura 3.8: $\lambda = 10$ Figura 3.9: $\lambda = 1$

Capítulo 4

Inferencia bayesiana

Una vez establecido el modelo propuesto y realizado un análisis descriptivo, el siguiente objetivo es realizar inferencia sobre los parámetros de interés. Por lo tanto, debido a la necesidad de implementar métodos MCMC, a continuación se obtienen las condicionales completas para el modelo propuesto.

4.1. Densidades condicionales completas

Proposición 4.1.1. Sea $\mathbf{D}_N = \{\mathbf{Y}_1, \dots, \mathbf{Y}_N\} = \{(\mathbf{Y}_1^I, \mathbf{Y}_1^{II}), \dots, (\mathbf{Y}_N^I, \mathbf{Y}_N^{II})\}$ una muestra aleatoria del modelo

$$\begin{aligned}\mathbf{Y}_i^I | \beta^I, \{\mathbf{b}_i^I\} &\sim N_n(\mathbf{X}_i^I \beta^I + Z^I \mathbf{b}_i^I, I_n) \\ \mathbf{Y}_i^{II} | \beta^{II}, \{\mathbf{b}_i^{II}\} &\sim N_n(\mathbf{X}_i^{II} \beta^{II} + Z^{II} \mathbf{b}_i^{II}, \frac{1}{\lambda} I_n)\end{aligned}$$

con la especificación para cada componente $k \in \{I, II\}$

$$\begin{aligned}f(\boldsymbol{\beta}^k | \mathbf{A}^k) &= N_p(\boldsymbol{\beta}^k | 0, \mathbf{A}^k) \\ f(\mathbf{b}_i^k | \boldsymbol{\Omega}^k) &= N_q(\mathbf{b}_i^k | 0, \boldsymbol{\Omega}^k), \text{ para cada } i = 1, \dots, N \\ f(\lambda | \alpha, \gamma) &= \text{Gamma}(\lambda | \alpha, \gamma) \\ f(\boldsymbol{\Omega}^k | \mathbf{B}^k) &= \text{Wishart}(\boldsymbol{\Omega}^k | v^k, \mathbf{B}^k), \text{ con } E(\boldsymbol{\Omega}^k) = v^k (\mathbf{B}^k)^{-1}\end{aligned}$$

donde \mathbf{A}^k , $\boldsymbol{\Omega}^k$, α, γ y \mathbf{B}^k son hiperparámetros de las densidades correspondientes. Entonces, omitiendo los superíndices para simplificar la notación, la distribución final conjunta para los parámetros en cada componente son:

$$\begin{aligned}
f(\boldsymbol{\beta}, \mathbf{b}_i, \boldsymbol{\Omega} | \mathbf{D}_n) &\propto |\boldsymbol{\Omega}|^{(v-q-1+N)/2} \\
&\times \exp\left\{-\frac{1}{2} \left[\sum_{i=1}^N (\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta} - \mathbf{Z} \mathbf{b}_i)' (\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta} - \mathbf{Z} \mathbf{b}_i) \right. \right. \\
&\quad \left. \left. + \boldsymbol{\beta}' \mathbf{A} \boldsymbol{\beta} + \sum_{i=1}^N \mathbf{b}_i' \boldsymbol{\Omega} \mathbf{b}_i + \text{tr}(\mathbf{B} \boldsymbol{\Omega}) \right] \right\}
\end{aligned}$$

para la primer componente y

$$\begin{aligned}
f(\boldsymbol{\beta}, \{\mathbf{b}_i\}_{i=1}^N, \boldsymbol{\Omega}, \lambda | \mathbf{D}_n) &\propto \lambda^{(nN/2+\alpha)-1} \times |\boldsymbol{\Omega}|^{(v+q-1+N)/2} \\
&\times \exp\left\{-\frac{1}{2} \left[\lambda \sum_{i=1}^N (\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta} - \mathbf{Z} \mathbf{b}_i)' (\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta} - \mathbf{Z} \mathbf{b}_i) \right. \right. \\
&\quad \left. \left. + \boldsymbol{\beta}' \mathbf{A} \boldsymbol{\beta} + \sum_{i=1}^N \mathbf{b}_i' \boldsymbol{\Omega} \mathbf{b}_i + \text{tr}(\mathbf{B} \boldsymbol{\Omega}) \right] \right\}
\end{aligned}$$

para la segunda componente.

DEMOSTRACIÓN. La verosimilitud para la primer componente está dada por

$$\begin{aligned}
f(\mathbf{Y}_1, \dots, \mathbf{Y}_N | \boldsymbol{\beta}, \{\mathbf{b}_i\}_{i=1}^N) &= \prod_{i=1}^N f(\mathbf{Y}_i | \boldsymbol{\beta}, \mathbf{b}_i) \\
&\propto \prod_{i=1}^N \exp\left(-\frac{1}{2} (\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta} - \mathbf{Z} \mathbf{b}_i)' (\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta} - \mathbf{Z} \mathbf{b}_i)\right) \\
&= \exp\left(-\frac{1}{2} \sum_{i=1}^N (\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta} - \mathbf{Z} \mathbf{b}_i)' (\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta} - \mathbf{Z} \mathbf{b}_i)\right)
\end{aligned}$$

Por otro lado, la distribución inicial para la primer componente está dada por:

$$\begin{aligned}
f(\boldsymbol{\beta}, \{\mathbf{b}_i\}_{i=1}^N, \boldsymbol{\Omega}) &= f(\boldsymbol{\beta} | \mathbf{A}) \times \prod_{i=1}^N f(\mathbf{b}_i | \boldsymbol{\Omega}) \times f(\boldsymbol{\Omega} | v, \mathbf{B}) \\
&\propto |\mathbf{A}|^{1/2} \times \exp\left\{-\frac{1}{2} \boldsymbol{\beta}' \mathbf{A} \boldsymbol{\beta}\right\} \times |\boldsymbol{\Omega}|^{N/2} \times \exp\left\{-\frac{1}{2} \sum_{i=1}^N \mathbf{b}_i' \boldsymbol{\Omega} \mathbf{b}_i\right\} \\
&\times |\boldsymbol{\Omega}|^{(v-q-1)/2} \times \exp\left\{-\frac{1}{2} \text{tr}(\mathbf{B} \boldsymbol{\Omega})\right\} \\
&\propto |\boldsymbol{\Omega}|^{(v-q-1+N)/2} \exp\left\{-\frac{1}{2} \left[\boldsymbol{\beta}' \mathbf{A} \boldsymbol{\beta} + \sum_{i=1}^N \mathbf{b}_i' \boldsymbol{\Omega} \mathbf{b}_i + \text{tr}(\mathbf{B} \boldsymbol{\Omega}) \right] \right\}
\end{aligned}$$

Por lo tanto la distribución final de los parámetros asociados a la primera componente resulta

$$\begin{aligned}
f(\boldsymbol{\beta}, \{\mathbf{b}_i\}_{i=1}^N, \boldsymbol{\Omega} | \mathbf{D}_n) &= f(\mathbf{Y}_1, \dots, \mathbf{Y}_N | \boldsymbol{\beta}, \{\mathbf{b}_i\}_{i=1}^N) \times f(\boldsymbol{\beta}, \{\mathbf{b}_i\}_{i=1}^N, \boldsymbol{\Omega}) \\
&\propto \exp \left\{ -\frac{1}{2} \sum_{i=1}^N (\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta} - \mathbf{Z} \mathbf{b}_i)' (\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta} - \mathbf{Z} \mathbf{b}_i) \right\} \times |\boldsymbol{\Omega}|^{(v-q-1+N)/2} \\
&\quad \times \exp \left\{ -\frac{1}{2} \left[\boldsymbol{\beta}' \mathbf{A} \boldsymbol{\beta} + \sum_{i=1}^N \mathbf{b}_i' \boldsymbol{\Omega} \mathbf{b}_i + \text{tr}(\mathbf{B} \boldsymbol{\Omega}) \right] \right\} \\
&= |\boldsymbol{\Omega}|^{(v-q-1+N)/2} \exp \left\{ -\frac{1}{2} \left[\sum_{i=1}^N (\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta} - \mathbf{Z} \mathbf{b}_i)' (\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta} - \mathbf{Z} \mathbf{b}_i) \right. \right. \\
&\quad \left. \left. + \boldsymbol{\beta}' \mathbf{A} \boldsymbol{\beta} + \sum_{i=1}^N \mathbf{b}_i' \boldsymbol{\Omega} \mathbf{b}_i + \text{tr}(\mathbf{B} \boldsymbol{\Omega}) \right] \right\}
\end{aligned}$$

Para la segunda componente, la verosimilitud es

$$\begin{aligned}
f(\mathbf{Y}_1, \dots, \mathbf{Y}_N | \boldsymbol{\beta}, \{\mathbf{b}_i\}_{i=1}^N) &= \prod_{i=1}^N f(\mathbf{Y}_i | \boldsymbol{\beta}, \mathbf{b}_i) \\
&\propto \prod_{i=1}^N |\lambda I_n|^{1/2} \exp \left(-\frac{1}{2} (\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta} - \mathbf{Z} \mathbf{b}_i)' \lambda I_n (\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta} - \mathbf{Z} \mathbf{b}_i) \right) \\
&= |\lambda I_n|^{N/2} \exp \left(-\frac{\lambda}{2} \sum_{i=1}^N (\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta} - \mathbf{Z} \mathbf{b}_i)' (\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta} - \mathbf{Z} \mathbf{b}_i) \right) \\
&= \lambda^{nN/2} \exp \left(-\frac{\lambda}{2} \sum_{i=1}^N (\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta} - \mathbf{Z} \mathbf{b}_i)' (\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta} - \mathbf{Z} \mathbf{b}_i) \right)
\end{aligned}$$

y la distribución inicial de la segunda componente, para la especificación dada, es

$$\begin{aligned}
f(\boldsymbol{\beta}, \{\mathbf{b}_i\}_{i=1}^N, \boldsymbol{\Omega}, \lambda) &= f(\boldsymbol{\beta} | \mathbf{A}) \times \prod_{i=1}^N f(\mathbf{b}_i | \boldsymbol{\Omega}) \times f(\boldsymbol{\Omega} | v, \mathbf{B}) \times f(\lambda | \alpha, \gamma) \\
&\propto |\mathbf{A}|^{1/2} \times \exp \left\{ -\frac{1}{2} \boldsymbol{\beta}' \mathbf{A} \boldsymbol{\beta} \right\} \times |\boldsymbol{\Omega}|^{N/2} \times \exp \left\{ -\frac{1}{2} \sum_{i=1}^N \mathbf{b}_i' \boldsymbol{\Omega} \mathbf{b}_i \right\} \\
&\quad \times |\boldsymbol{\Omega}|^{(v-q-1)/2} \times \exp \left\{ -\frac{1}{2} \text{tr}(\mathbf{B} \boldsymbol{\Omega}) \right\} \times \lambda^{(\alpha-1)} \exp \{ -\gamma \lambda \} \\
&\propto \lambda^{(\alpha-1)} |\boldsymbol{\Omega}|^{(v-q-1+N)/2} \exp \left\{ -\frac{1}{2} \left[\boldsymbol{\beta}' \mathbf{A} \boldsymbol{\beta} + \sum_{i=1}^N \mathbf{b}_i' \boldsymbol{\Omega} \mathbf{b}_i \right. \right. \\
&\quad \left. \left. + \text{tr}(\mathbf{B} \boldsymbol{\Omega}) + 2\gamma \lambda \right] \right\}
\end{aligned}$$

Entonces, la distribución final de los parámetros asociados a la segunda componente esta determinada por

$$\begin{aligned}
f(\boldsymbol{\beta}, \{\mathbf{b}_i\}_{i=1}^N, \boldsymbol{\Omega}, \lambda | \mathbf{D}_n) &= f(\mathbf{Y}_1, \dots, \mathbf{Y}_N | \boldsymbol{\beta}, \{\mathbf{b}_i\}_{i=1}^N) \times f(\boldsymbol{\beta}, \{\mathbf{b}_i\}_{i=1}^N, \boldsymbol{\Omega}, \lambda) \\
&\propto \lambda^{nN/2} \exp \left\{ -\frac{\lambda}{2} \sum_{i=1}^N (\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta} - \mathbf{Z} \mathbf{b}_i)' (\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta} - \mathbf{Z} \mathbf{b}_i) \right\} \\
&\quad \times \lambda^{(\alpha-1)} |\boldsymbol{\Omega}|^{(v-q-1+N)/2} \times \exp \left\{ -\frac{1}{2} \left[\boldsymbol{\beta}' \mathbf{A} \boldsymbol{\beta} + \sum_{i=1}^N \mathbf{b}_i' \boldsymbol{\Omega} \mathbf{b}_i + \text{tr}(\mathbf{B} \boldsymbol{\Omega}) + 2\gamma \lambda \right] \right\} \\
&= \lambda^{(nN/2+\alpha)-1} |\boldsymbol{\Omega}|^{(v-q-1+N)/2} \exp \left\{ -\frac{1}{2} \left[\lambda \sum_{i=1}^N (\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta} - \mathbf{Z} \mathbf{b}_i)' (\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta} - \mathbf{Z} \mathbf{b}_i) \right. \right. \\
&\quad \left. \left. + \boldsymbol{\beta}' \mathbf{A} \boldsymbol{\beta} + \sum_{i=1}^N \mathbf{b}_i' \boldsymbol{\Omega} \mathbf{b}_i + \text{tr}(\mathbf{B} \boldsymbol{\Omega}) + 2\gamma \lambda \right] \right\}
\end{aligned}$$

□

Una vez establecidas las distribuciones finales de cada componente del vector \mathbf{Y}_i es posible obtener las densidades condicionales finales completas. A continuación se detallan por separado estas condicionales completas para cada componente.

Debido a que la estructura del modelo, los resultados para la primera componente son un caso particular de del análisis de la segunda componente cuando $\lambda = 1$, en seguida se describen las condicionales completas de dicha componente omitiendo la demostración.

Proposición 4.1.2. Sea $\mathbf{D}_n = \{\mathbf{Y}_1, \dots, \mathbf{Y}_N\} = \{(r_{11}, \theta_{11}), \dots, (r_{Nn}, \theta_{Nn})\}$ un conjunto de observaciones del modelo

$$\begin{aligned}
\mathbf{Y}_i^I | \beta^I, \{\mathbf{b}_i^I\} &\sim N_n(\mathbf{X}_i^I \beta^I + \mathbf{Z}^I \mathbf{b}_i^I, I_n) \\
\mathbf{Y}_i^{II} | \beta^{II}, \{\mathbf{b}_i^{II}\} &\sim N_n(\mathbf{X}_i^{II} \beta^{II} + \mathbf{Z}^{II} \mathbf{b}_i^{II}, \frac{1}{\lambda} I_n)
\end{aligned}$$

para toda $i = 1, \dots, N$ y con la misma especificación de la Proposición 4.1.1. Entonces, omitiendo el superíndice, las densidades condicionales completas de la primera componente resultan ser:

1. $f(\boldsymbol{\beta} | \{\mathbf{b}_i\}_{i=1}^N, \boldsymbol{\Omega}, \mathbf{D}_n) = N_p(\boldsymbol{\beta} | \mathbf{C}^{-1} \sum_{i=1}^N (\mathbf{X}_i)' \mathbf{e}_i, \mathbf{C})$
2. $f(\mathbf{b}_i | \boldsymbol{\beta}, \boldsymbol{\Omega}, \mathbf{D}_n) = N_q(\mathbf{b}_i | \mathbf{D}_i^{-1} \mathbf{Z}_i' \mathbf{E}_i, \mathbf{D}_i) \quad \forall i = 1, \dots, N$

$$3. f(\boldsymbol{\Omega}|\beta, \{\mathbf{b}_i\}, \boldsymbol{\Lambda}, \mathbf{D}_n) = Wi(v + N, \mathbf{B} + \sum_{i=1}^N \mathbf{b}_i \mathbf{b}_i')$$

donde

$$\begin{aligned} \mathbf{C} &= \sum_{i=1}^N \mathbf{X}_i' \mathbf{X}_i + \mathbf{A} \\ \mathbf{e}_i &= \mathbf{Y}_i - \mathbf{Z}_i \mathbf{b}_i \\ \mathbf{E}_i &= \mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta} \\ \mathbf{D}_i &= \mathbf{Z}_i' \mathbf{Z}_i + \boldsymbol{\Omega} \end{aligned}$$

Proposición 4.1.3. Sea $\mathbf{D}_n = \{\mathbf{Y}_1, \dots, \mathbf{Y}_N\} = \{(r_{11}, \theta_{11}), \dots, (r_{Nn}, \theta_{Nn})\}$ un conjunto de observaciones del modelo, bajo las mismas condiciones de la Proposición 4.1.2. Entonces, omitiendo el superíndice, las densidades condicionales completas de la segunda componente resultan ser:

1. $f(\boldsymbol{\beta}|\{\mathbf{b}_i\}_{i=1}^N, \boldsymbol{\Omega}, \lambda, \mathbf{D}_n) = N_p(\boldsymbol{\beta} | \lambda \mathbf{C}^{-1} \sum_{i=1}^N (\mathbf{X}_i)' \mathbf{e}_i, \mathbf{C})$
2. $f(b_i|\beta, \boldsymbol{\Omega}, \lambda, \mathbf{D}_n) = N_q(\lambda \mathbf{D}_i^{-1} \mathbf{Z}_i' \mathbf{E}_i, \mathbf{D}_i) \quad \forall i = 1, \dots, N$
3. $f(\boldsymbol{\Omega}|\beta, \{\mathbf{b}_i\}, \lambda, \mathbf{D}_n) = Wi(v + N, \mathbf{B} + \sum_{i=1}^N \mathbf{b}_i \mathbf{b}_i')$
4. $f(\lambda|\beta, \{\mathbf{b}_i\}, \boldsymbol{\Omega}, \mathbf{D}_n) = Gamma\left(\frac{nN}{2} + \alpha, \frac{1}{2}[\sum_{i=1}^N \tilde{\mathbf{E}}_i' \tilde{\mathbf{E}}_i + 2\gamma]\right)$

donde

$$\begin{aligned} \mathbf{C} &= \lambda \sum_{i=1}^N \mathbf{X}_i' \mathbf{X}_i + \mathbf{A} \\ \mathbf{e}_i &= \mathbf{Y}_i - \mathbf{Z}_i \mathbf{b}_i \\ \mathbf{E}_i &= \mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta} \\ \mathbf{D}_i &= \lambda \mathbf{Z}_i' \mathbf{Z}_i + \boldsymbol{\Omega} \\ \tilde{\mathbf{E}}_i &= \mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta} - \mathbf{Z}_i \mathbf{b}_i. \end{aligned}$$

DEMOSTRACIÓN. De la Proposición 4.1.1 se tiene que la distribución final conjunta asociada a la segunda componente esta dada por

$$\begin{aligned} f(\boldsymbol{\beta}, \{\mathbf{b}_i\}_{i=1}^N, \boldsymbol{\Omega} | \mathbf{D}_n) &\propto \lambda^{(nN/2+\alpha)-1} \times |\boldsymbol{\Omega}|^{(v+q-1+N)/2} \\ &\times \exp\left\{-\frac{1}{2} \left[\lambda \sum_{i=1}^N (\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta} - \mathbf{Z} \mathbf{b}_i)' (\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta} - \mathbf{Z} \mathbf{b}_i) \right. \right. \\ &\left. \left. + \boldsymbol{\beta}' \mathbf{A} \boldsymbol{\beta} + \sum_{i=1}^N \mathbf{b}_i' \boldsymbol{\Omega} \mathbf{b}_i + \text{tr}(\mathbf{B} \boldsymbol{\Omega}) \right] \right\}. \end{aligned}$$

1. A partir de distribución final se tiene que la densidad condicional completa para $\boldsymbol{\beta}$ está dada por

$$\begin{aligned} f(\boldsymbol{\beta} | \{\mathbf{b}_i\}_{i=1}^N, \boldsymbol{\Omega}, \lambda, \mathbf{D}_n) &\propto \exp\left\{-\frac{1}{2} \left[\lambda \sum_{i=1}^N (\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta} - \mathbf{Z} \mathbf{b}_i)' \right. \right. \\ &\left. \left. (\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta} - \mathbf{Z} \mathbf{b}_i) + \boldsymbol{\beta}' \mathbf{A} \boldsymbol{\beta} \right] \right\} \end{aligned}$$

Analizando el exponente de la expresión anterior y definiendo $e_i = \mathbf{Y}_i - \mathbf{Z}_i \mathbf{b}_i$ se tiene

$$\begin{aligned} &\lambda \sum_{i=1}^N (e_i - \mathbf{X}_i \boldsymbol{\beta})' (e_i - \mathbf{X}_i \boldsymbol{\beta}) + \boldsymbol{\beta}' \mathbf{A} \boldsymbol{\beta} \\ &= \lambda \left[\sum_{i=1}^N (e_i' e_i - e_i' \mathbf{X}_i \boldsymbol{\beta} - (\mathbf{X}_i \boldsymbol{\beta})' e_i + (\mathbf{X}_i \boldsymbol{\beta})' (\mathbf{X}_i \boldsymbol{\beta})) \right] + \boldsymbol{\beta}' \mathbf{A} \boldsymbol{\beta} \\ &= \lambda \sum_{i=1}^N e_i' e_i - 2 \lambda \sum_{i=1}^N e_i' \mathbf{X}_i \boldsymbol{\beta} - \lambda \sum_{i=1}^N (\boldsymbol{\beta}' \mathbf{X}_i' \mathbf{X}_i \boldsymbol{\beta}) + \boldsymbol{\beta}' \mathbf{A} \boldsymbol{\beta} \\ &\propto -2 \lambda \sum_{i=1}^N e_i' \mathbf{X}_i \boldsymbol{\beta} - \lambda \sum_{i=1}^N (\boldsymbol{\beta}' \mathbf{X}_i' \mathbf{X}_i \boldsymbol{\beta}) + \boldsymbol{\beta}' \mathbf{A} \boldsymbol{\beta} \\ &= -2 \lambda \sum_{i=1}^N e_i' \mathbf{X}_i \boldsymbol{\beta} - \boldsymbol{\beta}' (\lambda \sum_{i=1}^N \mathbf{X}_i' \mathbf{X}_i + \mathbf{A}) \boldsymbol{\beta}. \end{aligned}$$

Ahora, nótese que

$$\begin{aligned}
& (\boldsymbol{\beta} - \mathbf{C}^{-1} \boldsymbol{\mu}_\beta)' \mathbf{C} (\boldsymbol{\beta} - \mathbf{C}^{-1} \boldsymbol{\mu}_\beta) \\
&= \boldsymbol{\beta}' \mathbf{C} \boldsymbol{\beta} - \boldsymbol{\beta}' \mathbf{C} \mathbf{C}^{-1} \boldsymbol{\mu}_\beta - (\mathbf{C}^{-1} \boldsymbol{\mu}_\beta)' \mathbf{C} \boldsymbol{\beta} + (\mathbf{C}^{-1} \boldsymbol{\mu}_\beta)' (\mathbf{C}^{-1} \boldsymbol{\mu}_\beta) \\
&\propto \boldsymbol{\beta}' \mathbf{C} \boldsymbol{\beta} - \boldsymbol{\beta}' \boldsymbol{\mu}_\beta - \boldsymbol{\mu}'_\beta (\mathbf{C}^{-1})' \mathbf{C}^{-1} \boldsymbol{\beta} \\
&= \boldsymbol{\beta}' \mathbf{C} \boldsymbol{\beta} - \boldsymbol{\beta}' \boldsymbol{\mu}_\beta - \boldsymbol{\mu}'_\beta \boldsymbol{\beta} \\
&= \boldsymbol{\beta}' \mathbf{C} \boldsymbol{\beta} - 2 \boldsymbol{\mu}'_\beta \boldsymbol{\beta}.
\end{aligned}$$

Entonces, tomando $\boldsymbol{\mu}_\beta = (\lambda \sum_{i=1}^N e_i' \mathbf{X}_i)'$ y $\mathbf{C} = \lambda \sum_{i=1}^N \mathbf{X}_i' \mathbf{X}_i + \mathbf{A}$, se tiene que

$$f(\boldsymbol{\beta} | \{\mathbf{b}_i\}_{i=1}^N, \boldsymbol{\Omega}, \lambda, \mathbf{D}_n) \propto \exp\left\{-\frac{1}{2} (\boldsymbol{\beta} - \mathbf{C}^{-1} \boldsymbol{\mu}_\beta)' \mathbf{C} (\boldsymbol{\beta} - \mathbf{C}^{-1} \boldsymbol{\mu}_\beta)\right\}$$

y por lo tanto

$$\begin{aligned}
f(\boldsymbol{\beta} | \{\mathbf{b}_i\}_{i=1}^N, \boldsymbol{\Omega}, \lambda, \mathbf{D}_n) &= N_p(\boldsymbol{\beta} | \mathbf{C}^{-1} \boldsymbol{\mu}_\beta, \mathbf{C}) \\
&= N_p(\boldsymbol{\beta} | \lambda \mathbf{C}^{-1} \sum_{i=1}^N (\mathbf{X}_i)' e_i, \mathbf{C}).
\end{aligned}$$

2. Dada la distribución final y definiendo $\mathbf{E}_i = \mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta}$, para cada $i = 1, \dots, N$ se tiene que la densidad condicional completa para \mathbf{b}_i es

$$\begin{aligned}
f(\mathbf{b}_i | \boldsymbol{\beta}, \boldsymbol{\Omega}, \lambda, \mathbf{D}_n) &\propto \exp\left\{-\frac{1}{2} \left[\lambda \sum_{i=1}^N (\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta} - \mathbf{Z}_i \mathbf{b}_i)' (\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta} - \mathbf{Z}_i \mathbf{b}_i) \right. \right. \\
&\quad \left. \left. + \mathbf{b}_i' \boldsymbol{\Omega} \mathbf{b}_i \right] \right\} \quad \forall i = 1, \dots, N.
\end{aligned}$$

realizando un análisis similar al efectuado anteriormente sobre el exponente, se obtiene que

$$\begin{aligned}
& \lambda (\mathbf{E}_i - \mathbf{Z}_i \mathbf{b}_i)' (\mathbf{E}_i - \mathbf{Z}_i \mathbf{b}_i) + \mathbf{b}_i' \boldsymbol{\Omega} \mathbf{b}_i \\
&= \lambda \mathbf{E}_i' \mathbf{E}_i - \lambda \mathbf{E}_i' \mathbf{Z}_i \mathbf{b}_i - \lambda (\mathbf{Z}_i \mathbf{b}_i)' \mathbf{E}_i + \lambda (\mathbf{Z}_i \mathbf{b}_i)' (\mathbf{Z}_i \mathbf{b}_i) + \mathbf{b}_i' \boldsymbol{\Omega} \mathbf{b}_i \\
&= \lambda \mathbf{E}_i' \mathbf{E}_i - 2 \lambda \mathbf{E}_i' \mathbf{Z}_i \mathbf{b}_i - \lambda (\mathbf{b}_i' \mathbf{Z}_i' \mathbf{Z}_i \mathbf{b}_i) + \mathbf{b}_i' \boldsymbol{\Omega} \mathbf{b}_i \\
&\propto -2 \lambda \mathbf{E}_i' \mathbf{Z}_i \mathbf{b}_i - \lambda (\mathbf{b}_i' \mathbf{Z}_i' \mathbf{Z}_i \mathbf{b}_i) + \mathbf{b}_i' \boldsymbol{\Omega} \mathbf{b}_i \\
&= 2 \lambda \mathbf{E}_i' \mathbf{Z}_i \mathbf{b}_i - \mathbf{b}_i' (\lambda \mathbf{Z}_i' \mathbf{Z}_i + \boldsymbol{\Omega}) \mathbf{b}_i.
\end{aligned}$$

Entonces, utilizando las ideas de la demostración anterior y definiendo $\mu_i = \lambda \mathbf{Z}'_i \mathbf{E}_i$ y $\mathbf{D}_i = \lambda \mathbf{Z}'_i \mathbf{Z}_i + \mathbf{\Omega}$, se tiene que

$$f(\mathbf{b}_i | \boldsymbol{\beta}, \mathbf{\Omega}, \lambda, \mathbf{D}_n) \propto \exp\left\{-\frac{1}{2}(\mathbf{b}_i - \mathbf{D}_i^{-1} \mu_i)' \mathbf{D}_i (\mathbf{b}_i - \mathbf{D}_i^{-1} \mu_i)\right\}$$

y por lo tanto

$$\begin{aligned} f(\mathbf{b}_i | \boldsymbol{\beta}, \mathbf{\Omega}, \lambda, \mathbf{D}_n) &= N_q(\mathbf{b}_i | \mathbf{D}_i^{-1} \mu_i, \mathbf{D}_i). \\ &= N_q(\mathbf{b}_i | \lambda \mathbf{D}_i^{-1} \mathbf{Z}'_i \mathbf{E}_i, \mathbf{D}_i). \end{aligned}$$

3. La densidad condicional completa de $\mathbf{\Omega}$ está dada por

$$\begin{aligned} f(\mathbf{\Omega} | \boldsymbol{\beta}, \{\mathbf{b}_i\}, \mathbf{\Lambda}, \mathbf{D}_n) &= |\mathbf{\Omega}|^{(v-q-1+N)/2} \exp\left\{-\frac{1}{2} \left[\sum_{i=1}^N \mathbf{b}'_i \mathbf{\Omega} \mathbf{b}_i + \text{tr}(\mathbf{B} \mathbf{\Omega}) \right]\right\} \\ &= |\mathbf{\Omega}|^{((v+N)-q-1)/2} \exp\left\{-\frac{1}{2} \left[\sum_{i=1}^N \text{tr}(\mathbf{\Omega} \mathbf{b}_i \mathbf{b}'_i) + \text{tr}(\mathbf{B} \mathbf{\Omega}) \right]\right\} \\ &= |\mathbf{\Omega}|^{((v+N)-q-1)/2} \exp\left\{-\frac{1}{2} \left[\text{tr} \left(\sum_{i=1}^N \mathbf{\Omega} \mathbf{b}_i \mathbf{b}'_i + \mathbf{B} \mathbf{\Omega} \right) \right]\right\} \\ &= |\mathbf{\Omega}|^{((v+N)-q-1)/2} \exp\left\{-\frac{1}{2} \text{tr} \left(\left[\sum_{i=1}^N \mathbf{b}_i \mathbf{b}'_i + \mathbf{B} \right] \mathbf{\Omega} \right)\right\}, \end{aligned}$$

es decir,

$$f(\mathbf{\Omega} | \boldsymbol{\beta}, \{\mathbf{b}_i\}, \mathbf{\Lambda}, \mathbf{D}_n) = Wi(v + N, \mathbf{B} + \sum_{i=1}^N \mathbf{b}_i \mathbf{b}'_i)$$

4. Para la densidad condicional de λ , definiendo $\tilde{\mathbf{E}}_i = \mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta} - \mathbf{Z}_i \mathbf{b}_i$, se tiene que

$$\begin{aligned} f(\lambda | \boldsymbol{\beta}, \{\mathbf{b}_i\}, \mathbf{\Omega}, \mathbf{D}_n) &= \lambda^{(nN/2+\alpha)-1} \exp\left\{-\frac{\lambda}{2} \sum_{i=1}^N (\tilde{\mathbf{E}}_i)' (\tilde{\mathbf{E}}_i) + \gamma \lambda\right\} \\ &= \lambda^{(nN/2+\alpha)-1} \exp\left\{-\frac{1}{2} \left[\lambda \sum_{i=1}^N (\tilde{\mathbf{E}}_i)' (\tilde{\mathbf{E}}_i) + 2\gamma \lambda \right]\right\} \\ &= \lambda^{(nN/2+\alpha)-1} \exp\left\{-\lambda \left[\frac{1}{2} \sum_{i=1}^N (\tilde{\mathbf{E}}_i)' (\tilde{\mathbf{E}}_i) + \gamma \right]\right\} \end{aligned}$$

Así, se obtiene que

$$f(\lambda|\beta, \{\mathbf{b}_i\}, \boldsymbol{\Omega}, \mathbf{D}_n) = \text{Gamma} \left(\frac{nN}{2} + \alpha, \frac{1}{2} \left[\sum_{i=1}^N \tilde{\mathbf{E}}_i' \tilde{\mathbf{E}}_i + 2\gamma \right] \right)$$

□

Por último, resta encontrar la densidad condicional completa para R relaciona ambas componentes. Sin embargo, la derivación de esta densidad se realizó de manera general en la Proposición 1.3.3 y por lo tanto para este modelo se tiene que:

$$f(r_{ij}|\theta_{ij}, \boldsymbol{\mu}_{ij}, \lambda) = \frac{d^2 r \exp(-\frac{1}{2} d^2 [r_{ij}^2 - 2 b r_{ij}])}{1 + \frac{db}{\phi(db)} \Phi(db)}$$

donde

$$\boldsymbol{\mu}_{ij} = \begin{pmatrix} \mu_{ij}^I \\ \mu_{ij}^{II} \end{pmatrix} = \begin{pmatrix} (x_{ij}^I)' \beta^I + (z_{ij}^I)' \mathbf{b}_i^I \\ (x_{ij}^{II})' \beta^{II} + (z_{ij}^{II})' \mathbf{b}_i^{II} \end{pmatrix}.$$

Una vez que se tienen especificadas todas las condicionales completas asociadas al modelo longitudinal, si se es capaz de simular de estas, se estaría en condiciones de obtener muestras de la distribución final de los parámetros del modelo a través de un muestreo de Gibbs (definido en la sección). La simulación de las densidades $f(\boldsymbol{\beta}|\{\mathbf{b}_i\}_{i=1}^N, \boldsymbol{\Omega}, \lambda, \mathbf{D}_n)$, $f(b_i|\beta, \boldsymbol{\Omega}, \lambda, \mathbf{D}_n)$, $f(\boldsymbol{\Omega}|\beta, \{\mathbf{b}_i\}, \lambda, \mathbf{D}_n)$ y de $f(\lambda|\beta, \{\mathbf{b}_i\}, \boldsymbol{\Omega}, \mathbf{D}_n)$ se realizan de manera directa. Por otro lado, para obtener simulaciones de la densidad condicional $f(r|\theta_{ij}, \boldsymbol{\mu}_{ij}, \lambda)$ es necesario implementar un algoritmo de Metropolis-Hastings.

De esta manera es posible realizar simulaciones de las distribuciones conjuntas $f(\boldsymbol{\beta}, \{\mathbf{b}_i\}_{i=1}^N, \boldsymbol{\Omega}|\mathbf{D}_n)$ y $f(\boldsymbol{\beta}, \{\mathbf{b}_i\}_{i=1}^N, \boldsymbol{\Omega}, \lambda|\mathbf{D}_n)$ correspondientes a la primera y segunda componente del modelo, respectivamente. La implementación del muestreo Gibbs para el modelo propuesto es el siguiente:

- Para la primera componente, muestrear:
 1. $\boldsymbol{\beta}$ de $f(\boldsymbol{\beta}|\{\mathbf{b}_i\}_{i=1}^N, \boldsymbol{\Omega}, \mathbf{D}_n)$
 2. \mathbf{b}_i de $f(b_i|\beta, \boldsymbol{\Omega}, \mathbf{D}_n)$
 3. $\boldsymbol{\Omega}$ de $f(\boldsymbol{\Omega}|\beta, \{\mathbf{b}_i\}, \mathbf{D}_n)$

- para la segunda componente, muestrear:
 1. β de $f(\beta|\{\mathbf{b}_i\}_{i=1}^N, \Omega, \lambda, \mathbf{D}_n)$
 2. \mathbf{b}_i de $f(b_i|\beta, \Omega, \lambda, \mathbf{D}_n)$
 3. Ω de $f(\Omega|\beta, \{\mathbf{b}_i\}, \lambda, \mathbf{D}_n)$
 4. λ de $f(\lambda|\beta, \{\mathbf{b}_i\}, \Omega, \mathbf{D}_n)$
- Muestrear R_{ij} de $f(r|\theta_{ij}, \mu_{ij}, \lambda) \forall i = 1, \dots, N$ y $j = 1, \dots, n$, a través del algoritmo de Metropolis-Hastings.
- Realizar los pasos anteriores utilizando los valores más recientes de los parámetros.
- Repetir el ciclo hasta lograr la convergencia y obtener el tamaño de muestra deseado.

4.2. Implementación

La implementación de los algoritmos de MCMC requieren ciertas consideraciones para su funcionamiento adecuado. Para el uso del algoritmo con el modelo propuesto, dentro de los aspectos que se deben tomar en cuenta se encuentra la determinación del periodo de calentamiento necesario para asegurar la convergencia a la distribución estacionaria. También se requiere especificar el número de iteraciones necesarios en el algoritmo de Metrópolis-Hastings utilizado. De la misma manera, debido a que los valores generados por una cadena de Markov tiene cierta dependencia de los estados anteriores, es necesario establecer el intervalo de muestreo, una vez alcanzada la convergencia, de tal manera que los elementos obtenidos sean lo menos dependientes entre sí. Por último, se debe realizar un constante análisis de la convergencia, ya que este último permite afirmar que los valores obtenidos provengan de la distribución objetivo planteada.

Además de lo anterior es importante realizar una aplicación eficiente del algoritmo para reducir el tiempo necesario para obtener la muestra deseada de la distribución final de los parámetros del modelo. En este sentido, el algoritmo anterior puede ser poco útil, debido a que puede requerir una gran cantidad de iteraciones. Existen diversas alternativas propuestas en la

literatura para mejorar la eficiencia del muestreo de Gibbs. Una de ellas, propuesta por Chib y Carlin (1999), es realizar la actualización de algunos de los parámetros en un solo bloque a través de la simulación indirecta de la densidad conjunta de aquellos parámetros incluidos en el grupo. A esta modificación se le conoce como *simulación por bloque*.

La aplicación de la simulación por bloques en este trabajo, consiste en simular en grupo los parámetros $\boldsymbol{\beta}$ y \mathbf{b}_i para toda $i = 1, \dots, N$ en cada componente del modelo. Esto se realiza simulando de la distribución conjunta de $\boldsymbol{\beta}$ y \mathbf{b}_i , a través de las densidades $f(\boldsymbol{\beta}, \mathbf{b}_i | \boldsymbol{\Omega}, \mathbf{D}_n)$ y $f(\boldsymbol{\beta}, \mathbf{b}_i | \boldsymbol{\Omega}, \lambda, \mathbf{D}_n)$. Lo anterior es posible ya que:

$$f(\boldsymbol{\beta}, \mathbf{b}_i | \cdot) = f(\boldsymbol{\beta} | \boldsymbol{\Omega}, \lambda, \mathbf{D}_n) f(\mathbf{b}_i | \boldsymbol{\beta}, \boldsymbol{\Omega}, \lambda, \mathbf{D}_n)$$

En la siguiente proposición se muestra la expresión de $f(\boldsymbol{\beta} | \boldsymbol{\Omega}, \mathbf{D}_n)$ y $f(\boldsymbol{\beta} | \boldsymbol{\Omega}, \lambda, \mathbf{D}_n)$ para la primera y segunda componente del modelo propuesto.

Proposición 4.2.1. Sea $\mathbf{D}_n = \{\mathbf{Y}_1, \dots, \mathbf{Y}_N\} = \{(r_{11}, \theta_{11}), \dots, (r_{Nn}, \theta_{Nn})\}$ un conjunto de observaciones del modelo

$$\begin{aligned} \mathbf{Y}_i^I | \beta^I, \{\mathbf{b}_i^I\} &\sim N_n(\mathbf{X}_i^I \beta^I + Z^I \mathbf{b}_i^I, I_n) \\ \mathbf{Y}_i^{II} | \beta^{II}, \{\mathbf{b}_i^{II}\} &\sim N_n(\mathbf{X}_i^{II} \beta^{II} + Z^{II} \mathbf{b}_i^{II}, \frac{1}{\lambda} I_n) \end{aligned}$$

para toda $i = 1, \dots, N$ y con la misma especificación de la Proposición 4.1.1. Entonces, se tiene que

$$f(\boldsymbol{\beta} | \boldsymbol{\Omega}, \mathbf{D}_n) = N_p(\boldsymbol{\beta} | \boldsymbol{\mu}_\beta, \boldsymbol{\Lambda}_\beta)$$

para la primer componente y

$$f(\boldsymbol{\beta} | \boldsymbol{\Omega}, \lambda, \mathbf{D}_n) = N_p(\boldsymbol{\beta} | \boldsymbol{\mu}_\beta, \boldsymbol{\Lambda}_\beta)$$

para la segunda componente. Donde para ambos casos:

$$\begin{aligned} \boldsymbol{\mu}_\beta &= \boldsymbol{\Lambda}_\beta^{-1} \left(\sum_{i=1}^N \mathbf{X}_i' \mathbf{V}_i^{-1} \mathbf{Y}_i \right) \\ \boldsymbol{\Lambda}_\beta &= \left(\mathbf{A} + \sum_{i=1}^N \mathbf{X}_i' \mathbf{V}_i^{-1} \mathbf{X}_i \right) \end{aligned}$$

Además, se tiene que $\mathbf{V}_i = \mathbf{Z}_i \boldsymbol{\Omega}^{-1} \mathbf{Z}_i' + \mathbf{I}$ para la primer componente y $\mathbf{V}_i = \mathbf{Z}_i \boldsymbol{\Omega}^{-1} \mathbf{Z}_i' + (1/\lambda) \mathbf{I}$ para la segunda componente del modelo.

DEMOSTRACIÓN. Dada la estructura Gaussiana condicional del modelo se tiene que

$$\begin{aligned} \mathbf{Y}_i^I | \boldsymbol{\beta}^I, \Omega &\sim N_n(\mathbf{X}_i^I \boldsymbol{\beta}^I, \mathbf{Z}_i \Omega^{-1} \mathbf{Z}_i' + \mathbf{I}_n) \\ \mathbf{Y}_i^{II} | \boldsymbol{\beta}^{II}, \Omega, \lambda &\sim N_n(\mathbf{X}_i^{II} \boldsymbol{\beta}^{II}, \mathbf{Z}_i \Omega^{-1} \mathbf{Z}_i' + (1/\lambda)\mathbf{I}_n) \end{aligned}$$

donde la varianza \mathbf{V}_i está dada por $\mathbf{V}_i = \mathbf{Z}_i \Omega^{-1} \mathbf{Z}_i' + \mathbf{I}$ para la primer componente y $\mathbf{V}_i = \mathbf{Z}_i \Omega^{-1} \mathbf{Z}_i' + (1/\lambda)\mathbf{I}$ para la segunda componente del modelo. Por otro lado, si se considera una distribución inicial $N_p(\boldsymbol{\beta} | 0, A)$ para el parámetro $\boldsymbol{\beta}$ de cada componente y omitiendo los superíndices, entonces

$$\begin{aligned} f(\boldsymbol{\beta}, \Omega, | \mathbf{D}_n) &\propto \exp\left\{-\frac{1}{2} \left[\sum_{i=1}^N (\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta})' \mathbf{V}_i^{-1} (\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta}) + \boldsymbol{\beta}' \mathbf{A} \boldsymbol{\beta} \right]\right\} \\ &\propto \exp\left\{-\frac{1}{2} \left[-2 \sum_{i=1}^N (\mathbf{Y}_i' \mathbf{V}_i^{-1} \mathbf{X}_i \boldsymbol{\beta}) \right. \right. \\ &\quad \left. \left. + \boldsymbol{\beta}' \left(A + \sum_{i=1}^N \mathbf{X}_i' \mathbf{V}_i^{-1} \mathbf{X}_i \right) \boldsymbol{\beta} \right]\right\} \\ &\propto \exp\left\{-\frac{1}{2} (\boldsymbol{\beta} - \boldsymbol{\mu}_\beta)' \boldsymbol{\Lambda}_\beta (\boldsymbol{\beta} - \boldsymbol{\mu}_\beta)\right\} \end{aligned}$$

donde

$$\boldsymbol{\mu}_\beta = \boldsymbol{\Lambda}_\beta^{-1} \sum_{i=1}^N (\mathbf{X}_i' \mathbf{V}_i^{-1} \mathbf{Y}_i) \quad \text{y} \quad \boldsymbol{\Lambda}_\beta = A + \sum_{i=1}^N \mathbf{X}_i' \mathbf{V}_i^{-1} \mathbf{X}_i$$

y por lo tanto

$$f(\boldsymbol{\beta}, \Omega, | \mathbf{D}_n) = N_p(\boldsymbol{\beta} | \boldsymbol{\mu}_\beta, \boldsymbol{\Lambda}_\beta)$$

□

Así, la implementación del muestreo de Gibbs por bloques para el modelo propuesto resulta:

- Para la primera componente, muestrear:
 1. $\boldsymbol{\beta}$ y $\{\mathbf{b}_i\}_{i=1}^N$ de $f(\boldsymbol{\beta}, \{\mathbf{b}_i\}_{i=1}^N | \Omega, \mathbf{D}_n)$ de la siguiente manera:
 - Obtener $\boldsymbol{\beta}$ de $f(\boldsymbol{\beta} | \Omega, \mathbf{D}_n)$

- Obtener \mathbf{b}_i de $f(\mathbf{b}_i|\beta, \Omega, \mathbf{D}_n)$ para toda $i = 1, \dots, N$
- 2. Ω de $f(\Omega|\beta, \{\mathbf{b}_i\}, \mathbf{D}_n)$
- Para la segunda componente, muestrear:
 1. β y $\{\mathbf{b}_i\}_{i=1}^N$ de $f(\beta, \{\mathbf{b}_i\}_{i=1}^N|\Omega, \mathbf{D}_n)$ de la siguiente manera:
 - Obtener β de $f(\beta|\Omega, \lambda, \mathbf{D}_n)$
 - Obtener \mathbf{b}_i de $f(\mathbf{b}_i|\beta, \Omega, \lambda, \mathbf{D}_n)$ para toda $i = 1, \dots, N$
 2. Ω de $f(\Omega|\beta, \{\mathbf{b}_i\}, \lambda, \mathbf{D}_n)$
 3. λ de $f(\lambda|\beta, \{\mathbf{b}_i\}, \Omega, \mathbf{D}_n)$
- Muestrear R_{ij} de $f(r|\theta_{ij}, \mu_{ij}, \lambda)$ para toda $i = 1, \dots, N$ y $j = 1, \dots, n$, a través del algoritmo de Metropolis-Hastings.
- Realizar los pasos anteriores utilizando los valores más recientes de los parámetros.
- Repetir el ciclo hasta lograr la convergencia y obtener el tamaño de muestra deseado.

Capítulo 5

Aplicación del modelo

Como ya se ha mencionado, el análisis de datos circulares es de gran importancia en diversas áreas del conocimiento. En particular, los datos longitudinales cuya variable de respuesta es de tipo circular permiten al investigador realizar estudios de su comportamiento, la relación que poseen las observaciones con diversas variables explicativas y además reflejan una dependencia entre observaciones intrínsecamente relacionadas con el tiempo en el cual son obtenidas. Por tal motivo, a pesar de su estructura compleja, este tipo de datos permiten realizar análisis más detallados de los datos y centrar el estudio en los objetivos principales de la investigación.

A continuación se aplica el modelo desarrollado en este trabajo a datos reales. Con el desarrollo de esta implementación se expone el funcionamiento adecuado del modelo para recuperar los parámetros de interés y se establece la importancia de la inclusión del parámetro λ al modelo propuesto por Nuñez Antonio y Gutiérrez-Peña (2014). Además se muestra la capacidad del modelo para distinguir la variabilidad entre cada individuo. Esa variabilidad se representa a través de efectos aleatorios determinados por los vectores \mathbf{b}_i , los cuales son controlados por la matriz $\mathbf{\Omega}$.

A través del análisis de los resultados obtenidos de la implementación del modelo, se busca mostrar que la inclusión del parámetro de variabilidad λ permite explicar de una mejor manera los datos circulares longitudinales. Cabe resaltar que la especificación del modelo utilizada para el desarrollo de los ejemplos considera efectos aleatorios unidimensionales; sin embargo, es posible generalizar su aplicación a mayores dimensiones.

Ejemplo 5.0.1. En este ejemplo se consideran 60 individuos, donde para cada uno de ellos se tienen 5 observaciones de datos circulares representados por sus ángulos. Estos datos fueron generados mediante:

$$\begin{aligned} \mathbf{Y}_i^I | \beta^I &\sim N_5(\mathbf{X}_i^I \beta^I, \mathbf{I}) \\ \mathbf{Y}_i^{II} | \beta^{II}, \{\mathbf{b}_i^{II}\} &\sim N_5(\mathbf{X}_i^{II} \beta^{II} + \mathbf{Z}_i^{II} \mathbf{b}_i^{II}, (1/\lambda) \mathbf{I}) \end{aligned}$$

con la siguiente especificación de los parámetros

$$\beta^I = \begin{pmatrix} 1 \\ -4 \\ 10 \\ 0.5 \end{pmatrix}, \quad \beta^{II} = \begin{pmatrix} 2 \\ 3 \\ 10 \\ 0.2 \end{pmatrix}, \quad \Omega^{-1} = 5 \quad \text{y} \quad \lambda = 0.5$$

Además se consideraron dos covariables X y W , para las cuales los valores fueron generados de una Poisson(5) y una Bernoulli(0.4), respectivamente. Además \mathbf{Z} fue establecido como el vector de 1. Por lo tanto las matrices \mathbf{X}_i^I , \mathbf{X}_i^{II} y el vector \mathbf{Z} se definen de la siguiente manera

$$\mathbf{X}_i^I = \mathbf{X}_i^{II} = \begin{pmatrix} 1 & x_{i1} & w_{i1} & 0 \\ 1 & x_{i2} & w_{i2} & 1 \\ 1 & x_{i3} & w_{i3} & 2 \\ 1 & x_{i5} & w_{i4} & 3 \\ 1 & x_{i5} & w_{i5} & 4 \end{pmatrix} \quad \text{y} \quad \mathbf{Z}^{II} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$$

A partir de los vectores (y_{ij}^I, y_{ij}^{II}) , se obtienen los ángulos θ_{ij} realizando

$$\theta_{ij} = \arctan(y_{ij}^{II}/y_{ij}^I) \text{ módulo}(2\pi)$$

Esta especificación permite imitar la estructura del modelo que se utilizará en el siguiente ejemplo para analizar datos reales. Los datos generados con estas especificación poseen un comportamiento similar al que se registra con los datos obtenidos para las pulgas de mar. De esta manera se muestra que el uso del modelo utilizando intercepto aleatorio en la segunda componente

es adecuado para realizar el análisis de datos longitudinales circulares con estructuras como el que presentan los datos reales que se analizan en este trabajo.

Con los datos generados, considerados como observados, se realizan diversas implementaciones del muestreo de Gibbs por bloques para poder determinar el número de iteraciones que conforman el periodo de calentamiento de las cadenas y de esta forma establecer la convergencia de los promedios ergódicos. La convergencia de estos promedios, constituyen la primer herramienta para poder afirmar que la muestra obtenida a partir de la cadena proviene, aproximadamente, de la distribución final del modelo propuesto.

Las figuras 5.1 y 5.2 muestran los promedios ergódicos de las cadenas de markov generadas para cada parámetro, al realizar 1500000 iteraciones del algoritmo de Gibbs. Se puede notar que a partir de la iteración 700000, los promedios ergódicos de cada cadena cualitativamente se estabilizan y por lo tanto a este número de iteraciones se le considera como el periodo de calentamiento del MCMC.

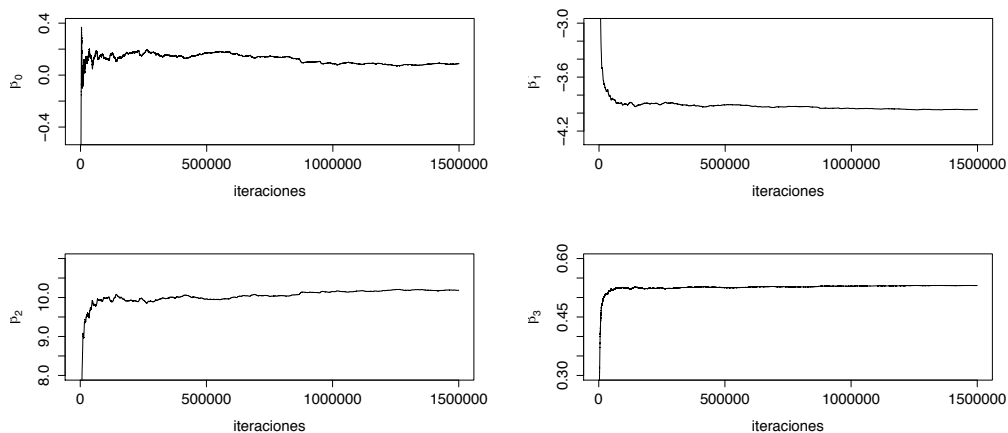


Figura 5.1: Promedios ergódicos para la componente I .

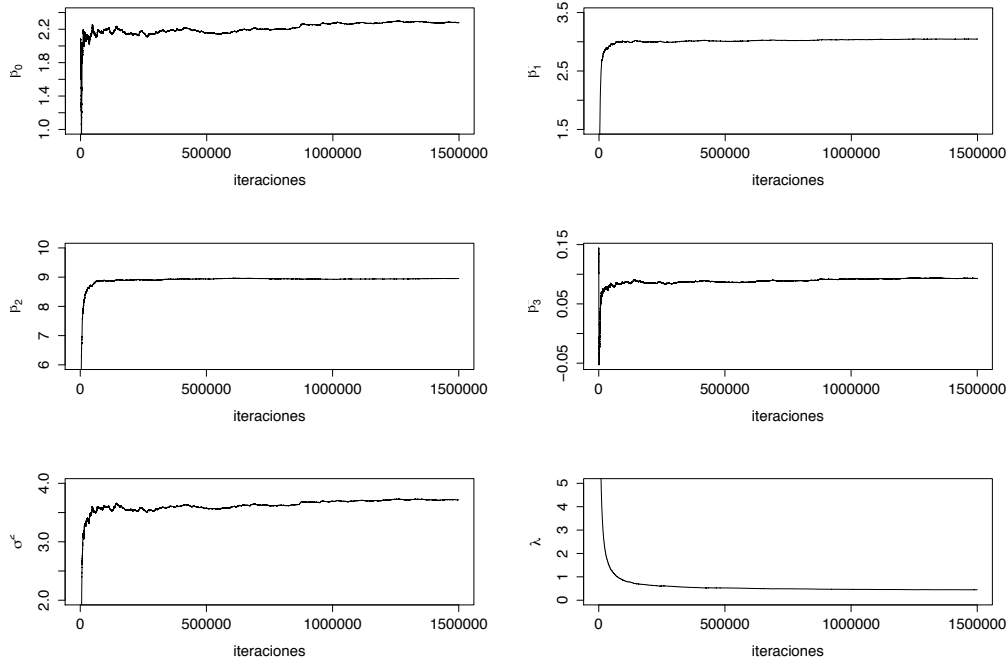


Figura 5.2: Promedios ergódicos para la componente II .

Si bien los promedios ergódicos muestran la convergencia aproximada de la cadena de Markov hacia la distribución final del modelo, se debe notar que por construcción existe una dependencia entre un valor de la cadena con los estados anteriores y posteriores de la cadena. Así pues, para que una muestra de valores obtenidos después del periodo de calentamiento pueda ser considerado como una muestra aproximada de la distribución final, es necesario asegurar que estos valores estén poco autocorrelacionados entre sí.

Para lograr una aproximación a lo anterior se utiliza la función de autocorrelación que muestra la correlación existente entre un estado y cada uno de los posteriores. A esta cantidad de elementos descartados entre un valor tomado de la cadena y el siguiente, de tal manera que representen elementos independientes, se le conoce como el salto o lag requerido para muestrear de la cadena.

Entonces, analizando las autocorrelaciones de cada cadena de Markov ge-

nerada para los parámetros, podemos notar que existe una gran dependencia de un estado de la cadena. Por lo tanto es necesario descartar 400 iteraciones entre un elemento considerado como parte de la muestra y el siguiente elemento. En la Figura 5.3 y Figura 5.4 se observan las autocorrelaciones una vez que se han descartado 400 iteraciones entre cada elemento considerado como parte de la muestra.

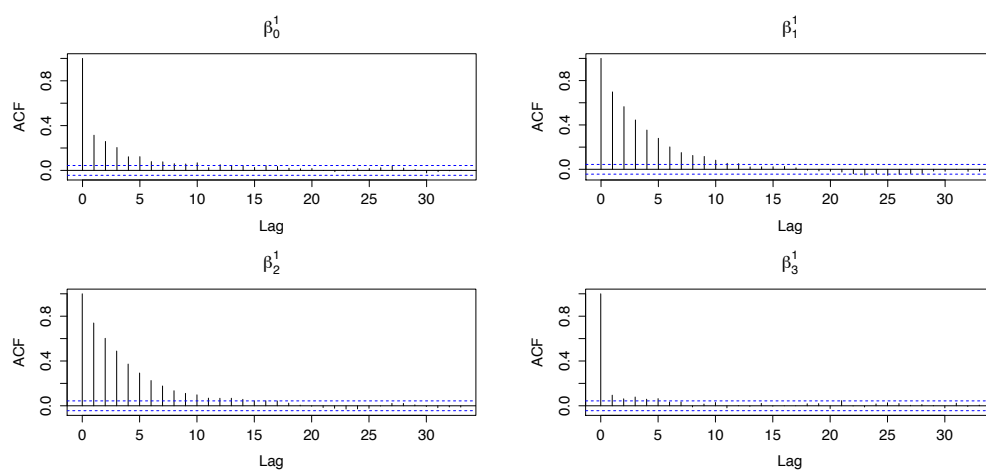


Figura 5.3: Autocorrelaciones para la componente I .

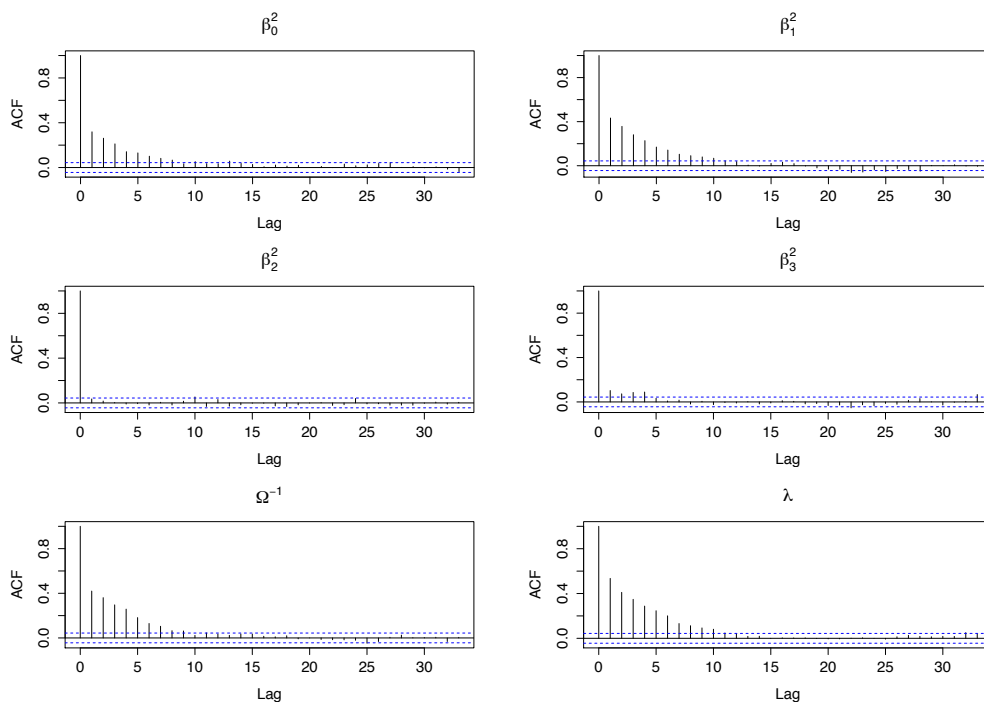


Figura 5.4: Autocorrelaciones para la componente *II*.

Con lo anterior se está en condiciones de obtener muestras de las distribuciones finales de los parámetros de interés y partir de éstas se puede realizar inferencia sobre los parámetros. Se obtuvieron 2000 muestras de los parámetros involucrados, realizando un muestreo de Gibbs con un periodo de calentamiento de 700000 y un salto o lag de 400 iteraciones.

A partir del análisis de los histogramas (Figura 5.5 y Figura 5.6) y de los intervalos de credibilidad, se observa que el modelo propuesto recupera los valores reales de los parámetros. Esto debido a que dichos valores son incluidos en los intervalos de credibilidad al 95 % obtenidos en el Cuadro 5.1.

Se destaca que, en lo referente al parámetro λ , la muestra obtenida permite obtener resultados que son congruentes con el valor real utilizado en la generación de los datos. Esto es, que dado que el intervalo de credibilidad para λ es (0.082897, 0.785565) se concluye que la inclusión de dicha variabilidad es necesaria. Además se descarta la posibilidad de que se pudiera utilizar el

modelo original propuesto por Nuñez-Antonio y Gutiérrez-Peña (2014) para analizar correctamente los datos observados, que fueron generados con $\lambda \neq 1$, lo cual está en concordancia con lo que se esperaría antes de la simulación.

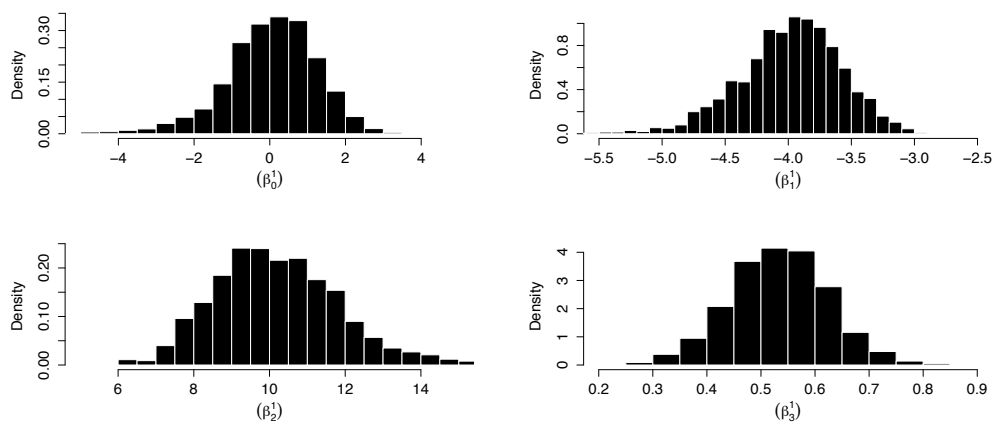
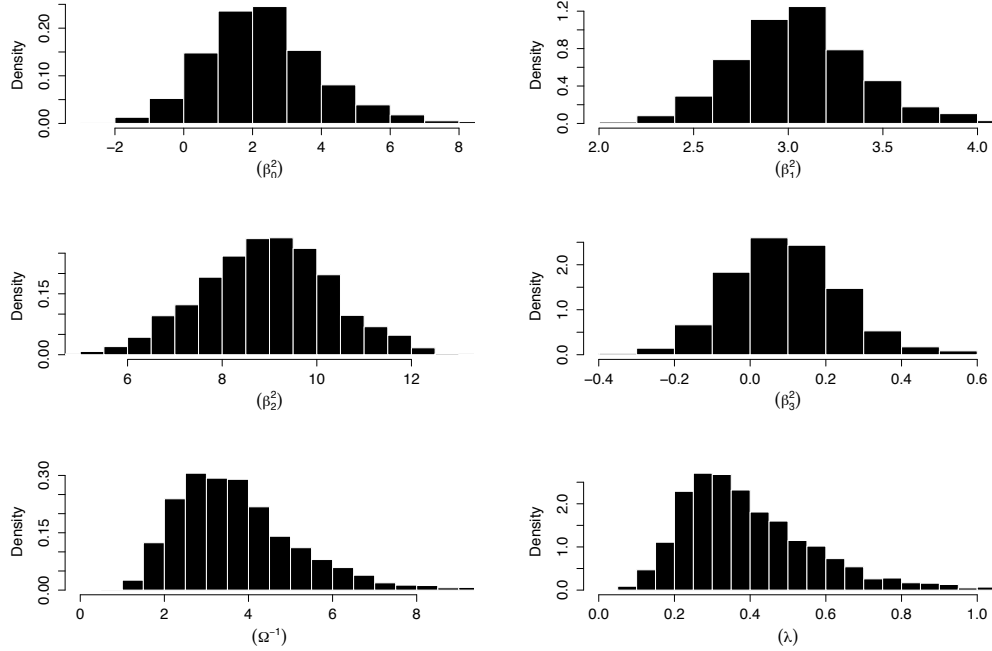


Figura 5.5: Histogramas para la componente I .

Figura 5.6: Histogramas para la componente I .

Cuadro 5.1: Intervalos de credibilidad al 95 % para los parámetros

	Componente I	Componente II
β_0	(-2.291028 , 2.366298)	(-0.9075461 , 5.6517368)
β_1	(-4.739325 , -3.205836)	(2.389183 , 3.731780)
β_2	(6.938841 , 13.712870)	(6.160511 , 11.707837)
β_3	(0.351763 , 0.712738)	(-0.194942 , 0.392268)
Ω^{-1}		(1.477255 , 6.946804)
λ		(0.082897 , 0.785565)

Ejemplo 5.0.2. En este ejemplo se simularon nuevamente 60 individuos con 5 observaciones para cada uno, implementando la misma especificación del modelo utilizado en el Ejemplo 5.0.1. Para esta implementación se estableció $\lambda = 5$, con el objetivo de determinar el correcto funcionamiento del modelo para valores de dicho parámetro mayores a 1.

Para el desarrollo de este ejemplo se considera un periodo de calentamiento de 700000 iteraciones, ya que cada una de las cadenas se estabilizas a partir de dicha iteración. Lo anterior se observa en las gráficas de los promedios ergódicos de las figuras 5.7 y 5.8.

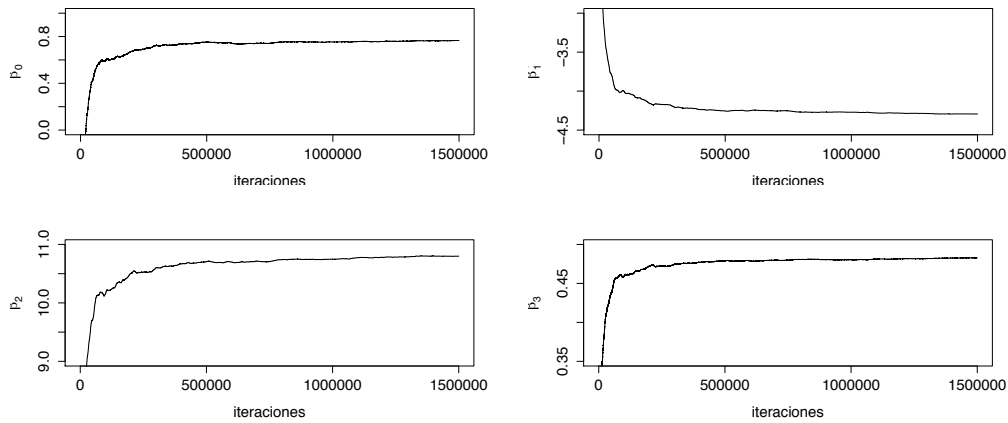
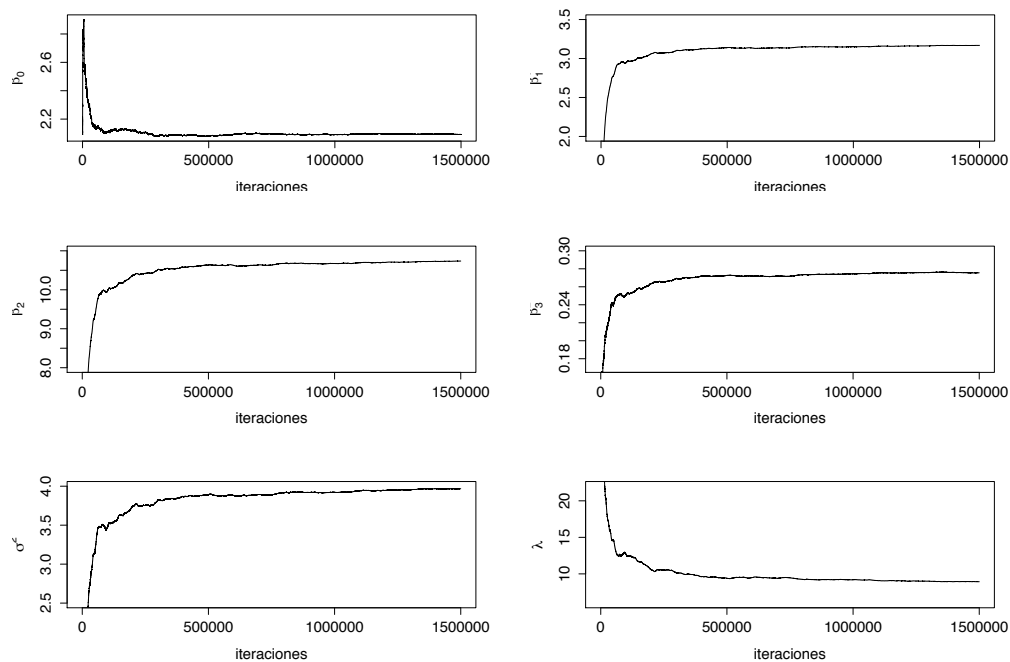
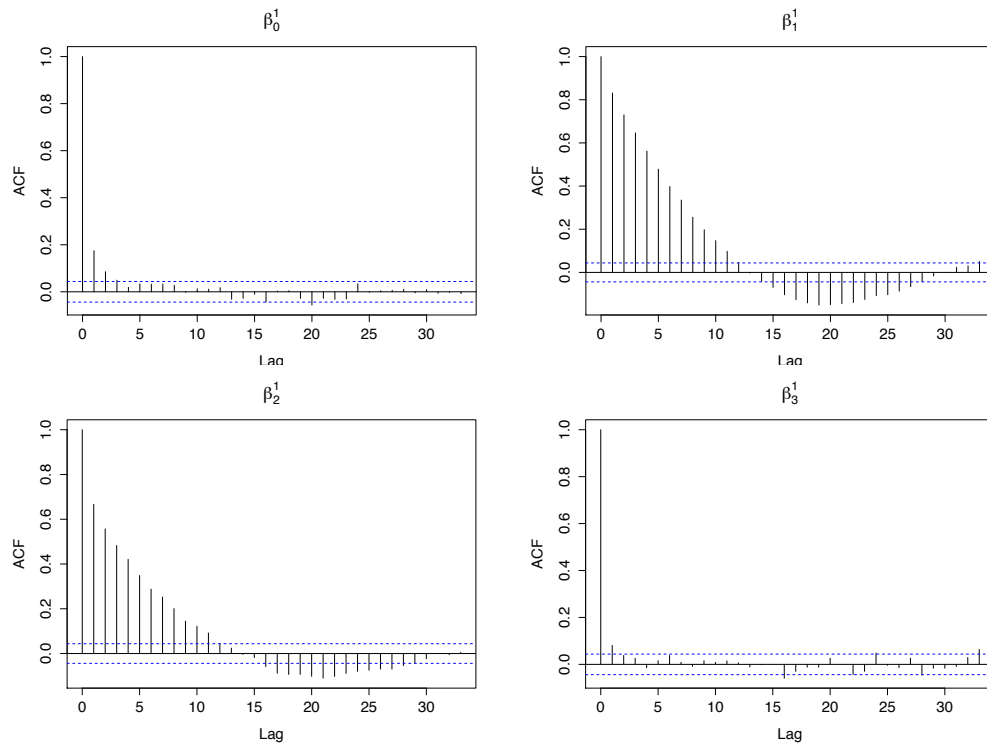


Figura 5.7: Promedios ergódicos para la componente I .

Figura 5.8: Promedios ergódicos para la componente I .

Una vez descartado el periodo de calentamiento se analiza la autocorrelación que poseen los elementos de cada cadena de Markov. Para disminuir la autocorrelación se contemplo un lag de 400 iteraciones. Las graficas de las Figura 5.9 y Figura 5.10 muestra las autocorrelaciones obtenidas una vez que se han descartado 400 iteraciones entre cada elemento considerado parte de la muestra.

Figura 5.9: Autocorrelaciones para la componente I .

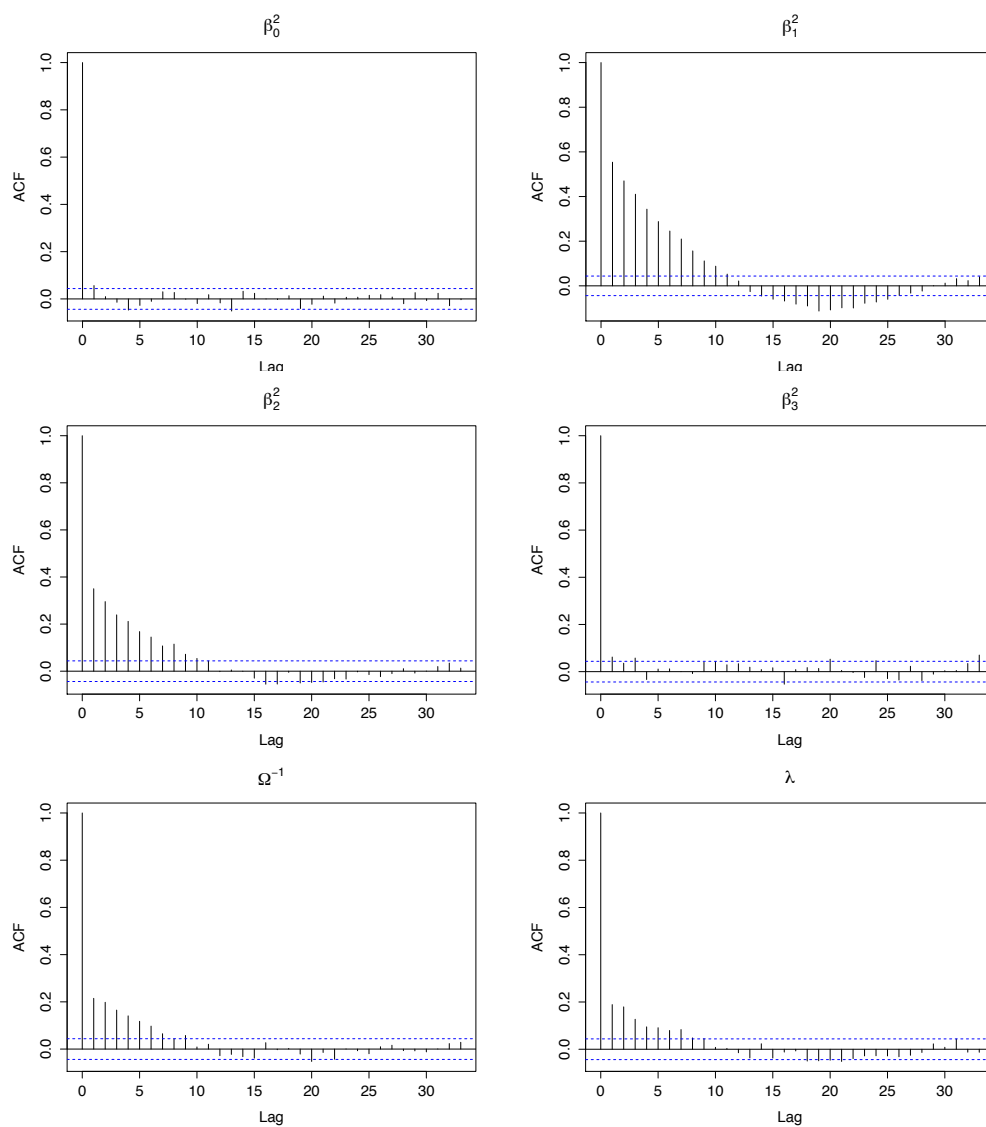


Figura 5.10: Autocorrelaciones para la componente I .

Por lo tanto, se obtuvieron 2000 muestras de los parámetros involucrados, realizando un muestreo de Gibbs con un periodo de calentamiento de 700000 y un salto o lag de 400 iteraciones. A partir de los resultados obtenidos se graficaron los histogramas, los cuales se muestran en Figura 5.11 y Figura 5.12.

Como se puede notar, a partir de los histogramas correspondientes y de los intervalos de credibilidad obtenidos (Cuadro 5.2), la implementación del modelo recupera adecuadamente los valores reales de los parámetros para datos generados con $\lambda = 5$. De tal forma que se puede afirmar que el modelo funciona adecuadamente para datos con características de poca variabilidad del error.

De manera similar al Ejemplo 5.0.1, se concluye que para el parámetro λ los resultados obtenidos son congruentes con el valor real utilizado. Esto se debe a que el valor de λ utilizado para la generación de los datos se encuentra incluido en el intervalo de credibilidad al 95 %, (1.126706, 23.283487). Además, dado que el intervalo de credibilidad no contiene el valor de $\lambda = 1$ se determina la inclusión de dicho parámetro de variabilidad es necesario para describir estos datos.

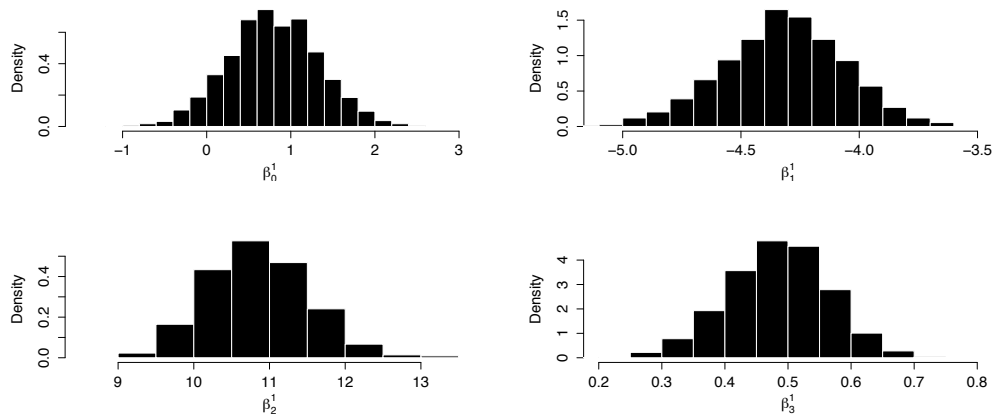


Figura 5.11: Histogramas para la componente I .

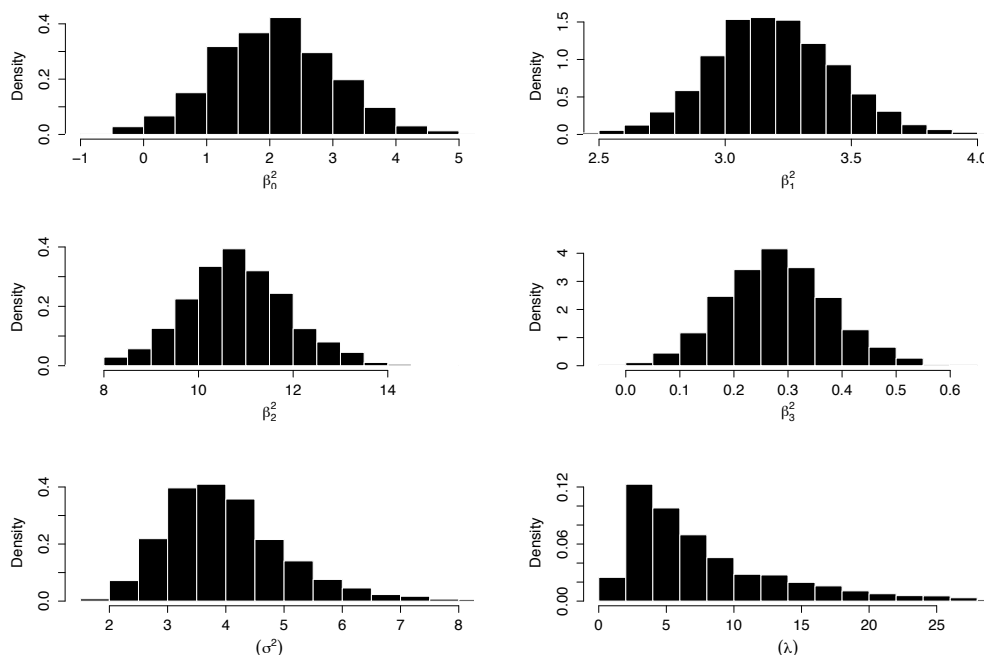


Figura 5.12: Histogramas para la componente I .

Cuadro 5.2: Intervalos de credibilidad al 95 % para los parámetros

	Componente I	Componente II
β_0	(-0.2177825 , 1.8784187)	(0.084423 , 3.895800)
β_1	(-4.832642 , -3.796154)	(2.699984 , 3.673158)
β_2	(9.671041 , 12.242702)	(8.713227 , 13.078182)
β_3	(0.335151 , 0.645957)	(0.101841 , 0.492889)
Ω^{-1}		(2.11253 , 6.16521)
λ		(1.126706 , 23.283487)

Con los dos ejemplos anteriores se muestra el correcto funcionamiento del modelo y por tal motivo que es adecuado para realizar inferencia sobre datos circulares longitudinales. Además, como ya se demostró, el modelo es capaz de distinguir adelas diversas fuentes de variabilidad intra e inter sujeto con la inclusión. Con su implementación es posible determinar, con certeza, si la inclusión del parámetro λ es necesaria para explicar los datos longitudinales

analizados. En los dos ejemplo anteriores se mostró que si $\lambda \neq 1$, no sería adecuado utilizar el modelo original propuesto por Núñez Antonio y Gutiérrez-Peña (2014) para realizar inferencia sobre esos datos.

Por último se observar al comparar los resultados en los dos ejemplos anteriores que, bajo las mismas condiciones de caleantamiento y lag, las cadenas de Markov para valores de $\lambda < 1$ convergen más rápido y poseen una menor autocorrelación que aquellas generadas para valores de $1 < \lambda$. Lo anterior se debe a que valores de λ menores a 1 representan una mayor variabilidad y por lo tanto la cadena es capaz de moverse con mayor facilidad a través del espacio muestral.

Ahora bien, una vez establecido el correcto funcionamiento del modelo y la importancia del la inclusión del parámetro λ , a continuación se implementa el modelo para analizar datos reales.

Ejemplo 5.0.3. En este ejemplo se consideró un conjunto de datos relacionado a la dirección de escape de 65 pulgas de mar (*Talitrus saltator*) y cuyos datos fueron tomados de Song (2007). En ese estudio longitudinal las 65 pulgas de mar fueron liberadas secuencialmente en 5 ocasiones y se registraron cada una de las direcciones de escape. Además se registraron covariables relacionadas con mediciones de la simetría ocular (*Eye*), la dirección azimutal del sol (*Sun*) y la velocidad del viento dividida en: offshore, longshore-east y lonshore-west (*OS, LSE* y *LSW*). En la Figura 5.13 se muestran las 65 series, cada una de estas conformadas por las 5 observaciones de las direcciones de escape.

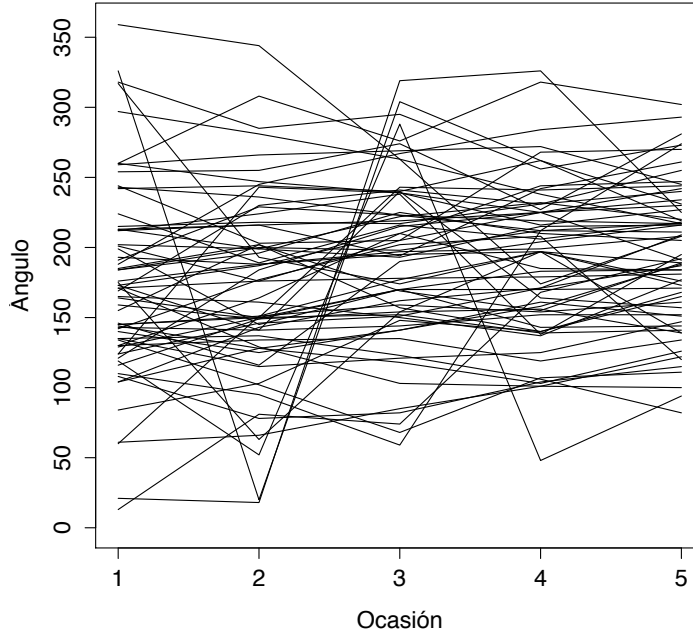


Figura 5.13: Gráfica de las direcciones de escape de 65 *Talitrus saltator*

Para el desarrollo de este ejemplo se considero el modelo longitudinal circular con la inclusión del parametro λ y con

$$\begin{aligned}\mu_{ij}^I &= \beta_0^I + \beta_1^I Sun + \beta_2^I Eye + \beta_3^I OS + \beta_4^I LSW + \beta_5^I LSE + \beta_6^I Tiempo \\ \mu_{ij}^{II} &= \beta_0^{II} + \beta_1^{II} Sun + \beta_2^{II} Eye + \beta_3^{II} OS + \beta_4^{II} LSW + \beta_5^{II} LSE \\ &\quad + \beta_6^{II} Tiempo + b_{0i}\end{aligned}$$

Para la especificación de la distribución inicial se usaron $\mathbf{A}^I = \mathbf{0}$, $\mathbf{A}^{II} = \mathbf{0}$, $v^{II} = 2$, $\mathbf{B}^{II} = 0.001$, $\alpha = 0.01$ y $\gamma = 0.01$. Además se implemento el algoritmo de Gibbs con valores iniciales $b = 0$, $\Omega = 1$, $\lambda = 1$ y $\mathbf{R} = \mathbf{1}$. En las Figuras 5.14, 5.15 y 5.16 se muestran los promedios ergódicos de las primeras 50000. En estas gráficas se puede notar que las cadenas presentan un comportamiento cualitativamente asintótico a partir de la iteración 30000 y por tal motivo se contempló este número de iteraciones como el periodo de calentamiento.

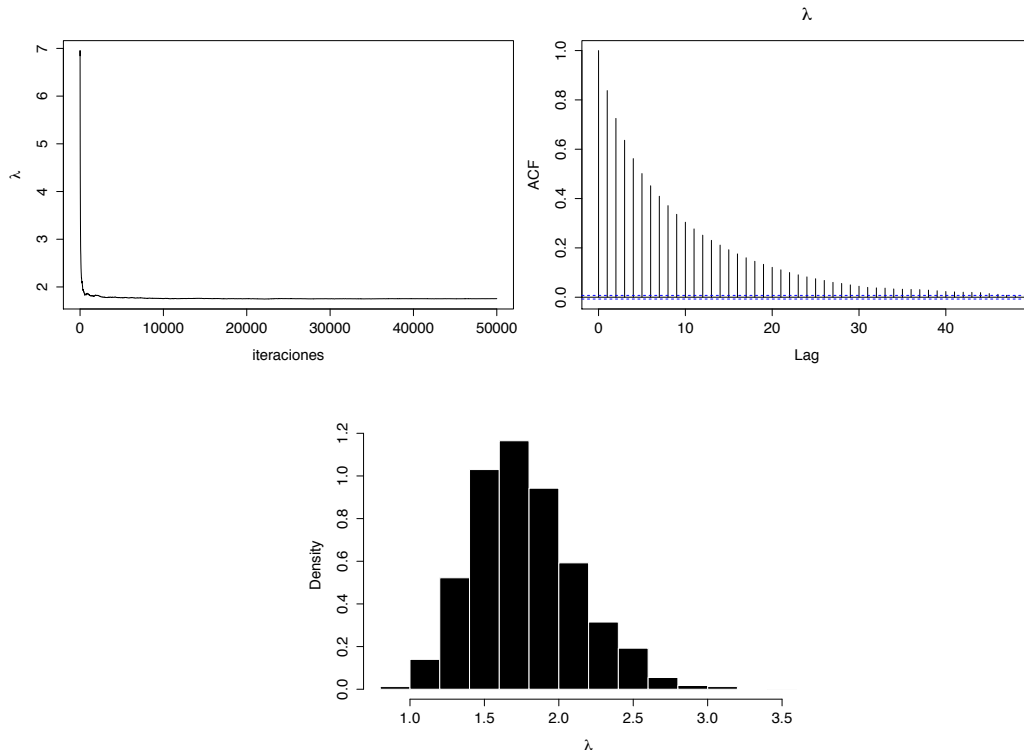


Figura 5.14: Promedio ergódico, autocorrelación e histograma del parámetro λ .

Una vez descartadas las iteraciones correspondientes al periodo de calentamiento, se analiza la autocorrelación entre los diferentes estados de la cadena. Para ellos se graficó la función de auto correlación para cada parámetro, las cuales se presentan en las Figuras 5.14, 5.17 y 5.18. En dichas gráficas se puede observar que los parámetros β_3^1 , Ω^{-1} y λ poseen la mayor autocorrelación entre sus estados por lo que se establece un salto o lag, entre un estado seleccionado y el siguiente, de 35 iteraciones. Con lo anterior se busca obtener una muestra que esté menos correlacionada.

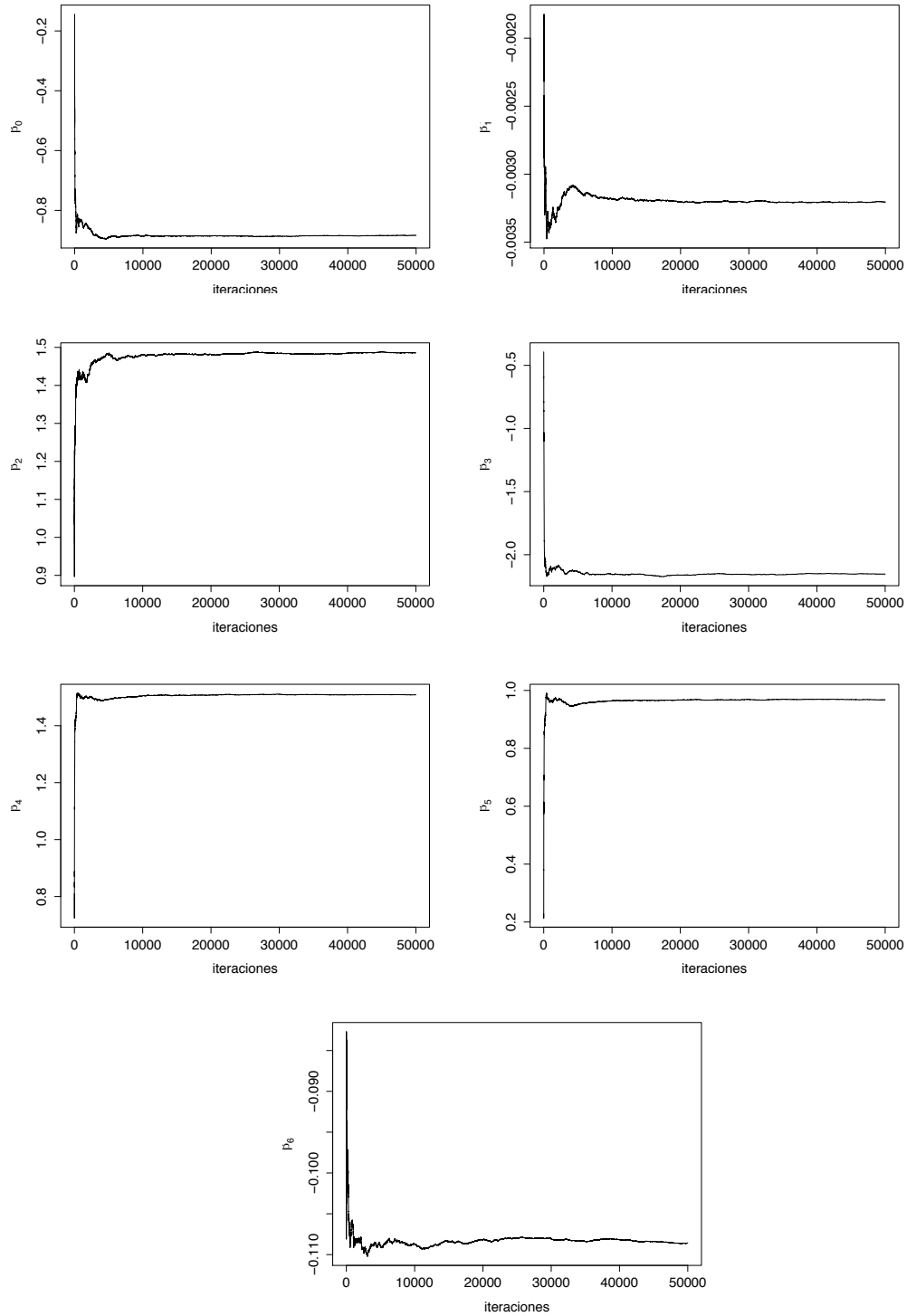
Por lo tanto, para obtener una muestra de la distribución final de los parámetros se realizaron un total de 100,000 iteraciones, considerando un periodo de calentamiento de 30,000, un lag o salto de 35 y donde en cada paso se realizó el algoritmo de Metrópolis-Hastings con 5 iteraciones.

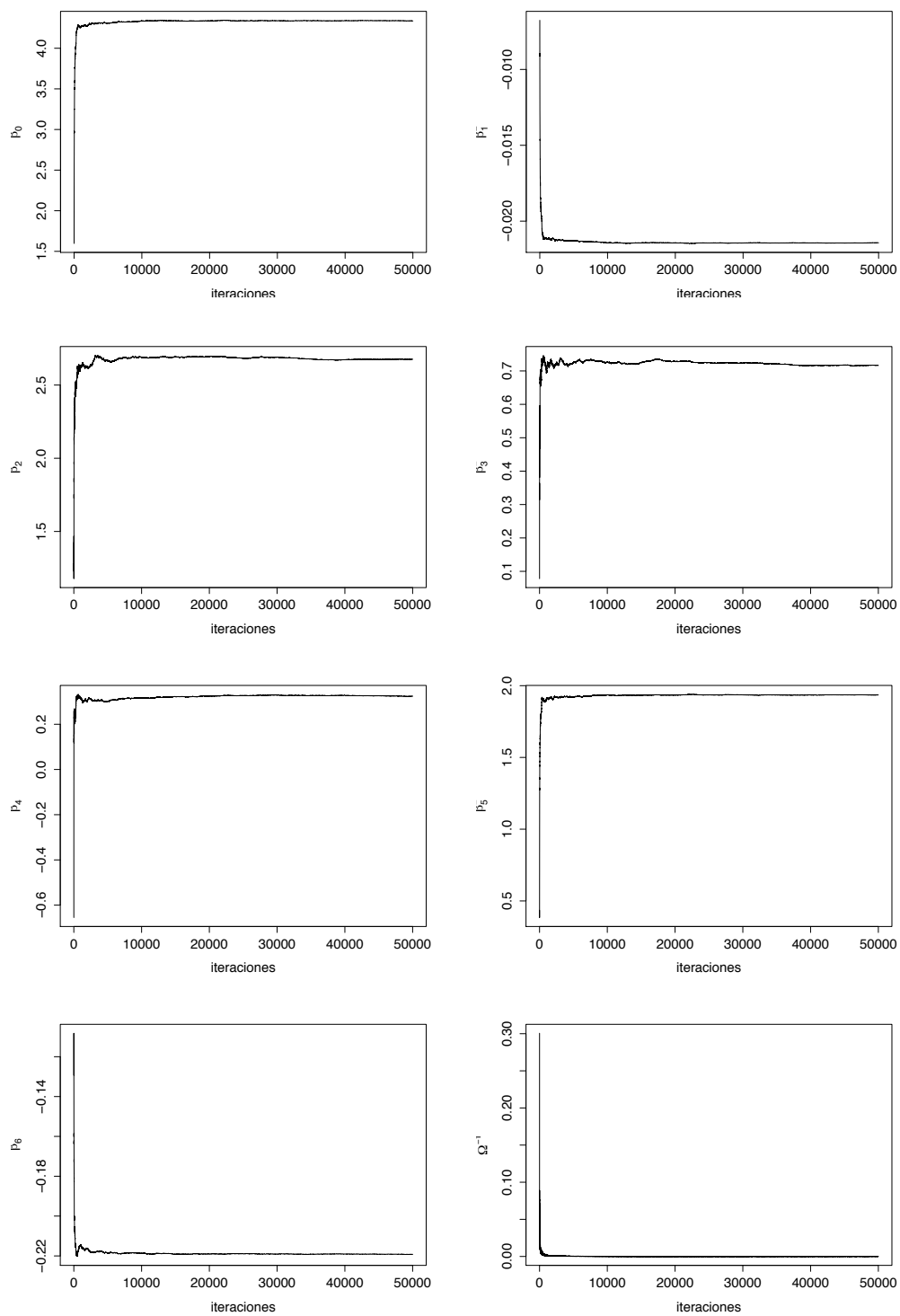
Las Figuras 5.14, 5.19 y 5.20 muestran las distribuciones finales para los parámetros de las componentes I y II del modelo, respectivamente. Así mismo, el Cuadro 5.3 contiene los intervalos de credibilidad para todos los parámetros involucrados en cada componente. A partir del análisis de estos resultados se concluye que las variables $\{\text{Sun}\}$ y $\{\text{Eye,Os,LWS}\}$ no son relevantes para μ^I y μ^{II} , respectivamente. También se concluye que, dado que el parámetro Ω^{-1} es distinto de 0, existe una variación significativa entre los individuos y por lo tanto dicha variabilidad debe ser explicada a través de la inclusión del intercepto aleatorio. Estos resultados obtenidos concuerdan, esencialmente, con aquellos obtenidos en análisis realizados a estos datos anteriormente (ver D'Elia *et al*,2001; Song,2007; Núñez Antonio y Gutiérrez-Peña,2014).

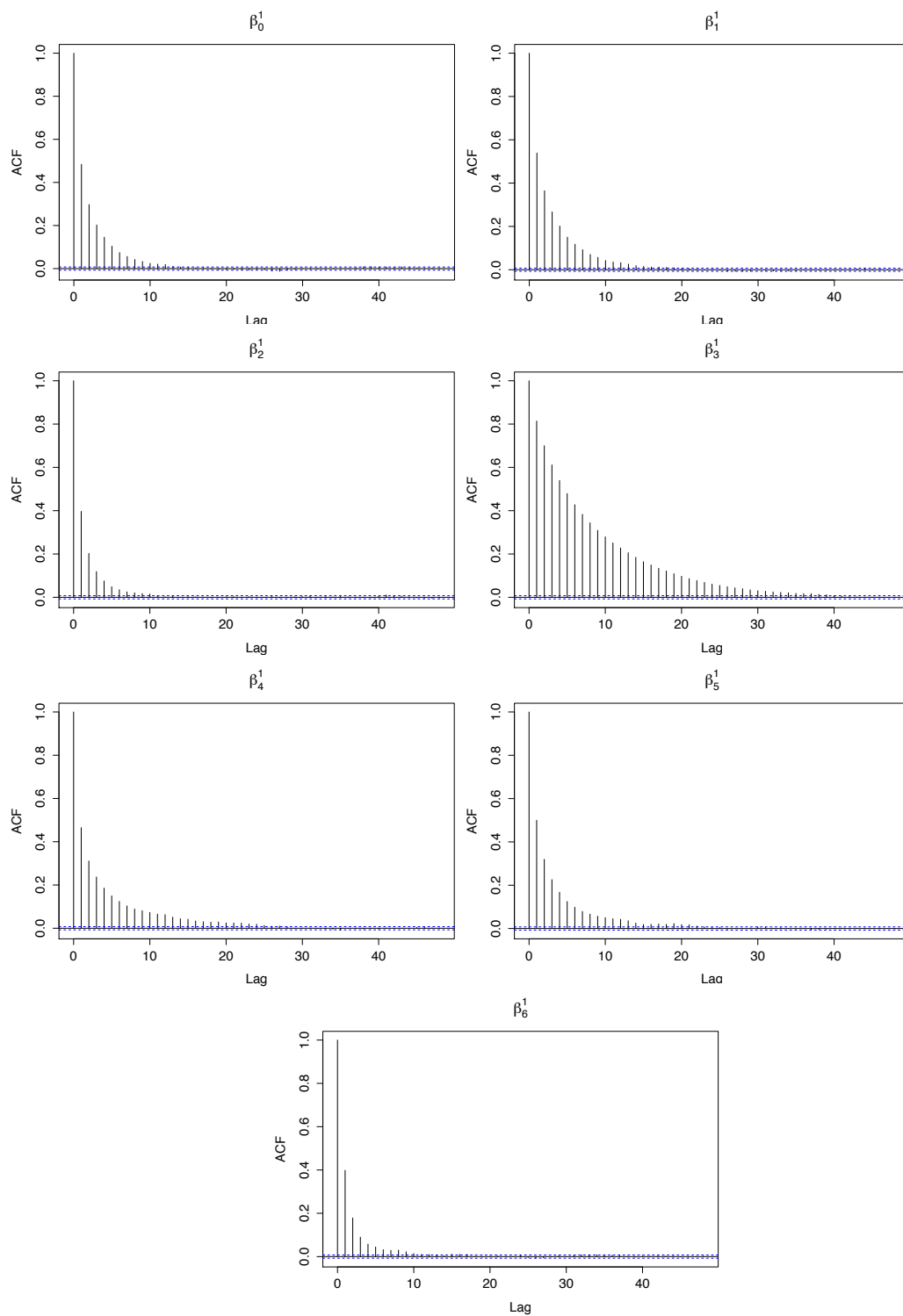
Por otro lado, el análisis sobre los resultados referentes a la variabilidad del error en la segunda componente muestra que, debido a que λ es distinto de 1, es pertinente la inclusión del éste parámetro en el modelo. Más aún, dado que los valores admitidos en dicho intervalo de credibilidad son estrictamente mayores a 1, es posible concluir que la variabilidad en las observaciones está conformada en una mayor medida por los aspectos de variabilidad propia de los individuos, así como por la variabilidad entre los sujetos. Por lo tanto el modelo propuesto en este trabajo permite mejorar la inferencia realizada sobre los datos de las pulgas de mar, en comparación al modelo propuesto por Núñez Antonio y Gutiérrez-Peña (2014).

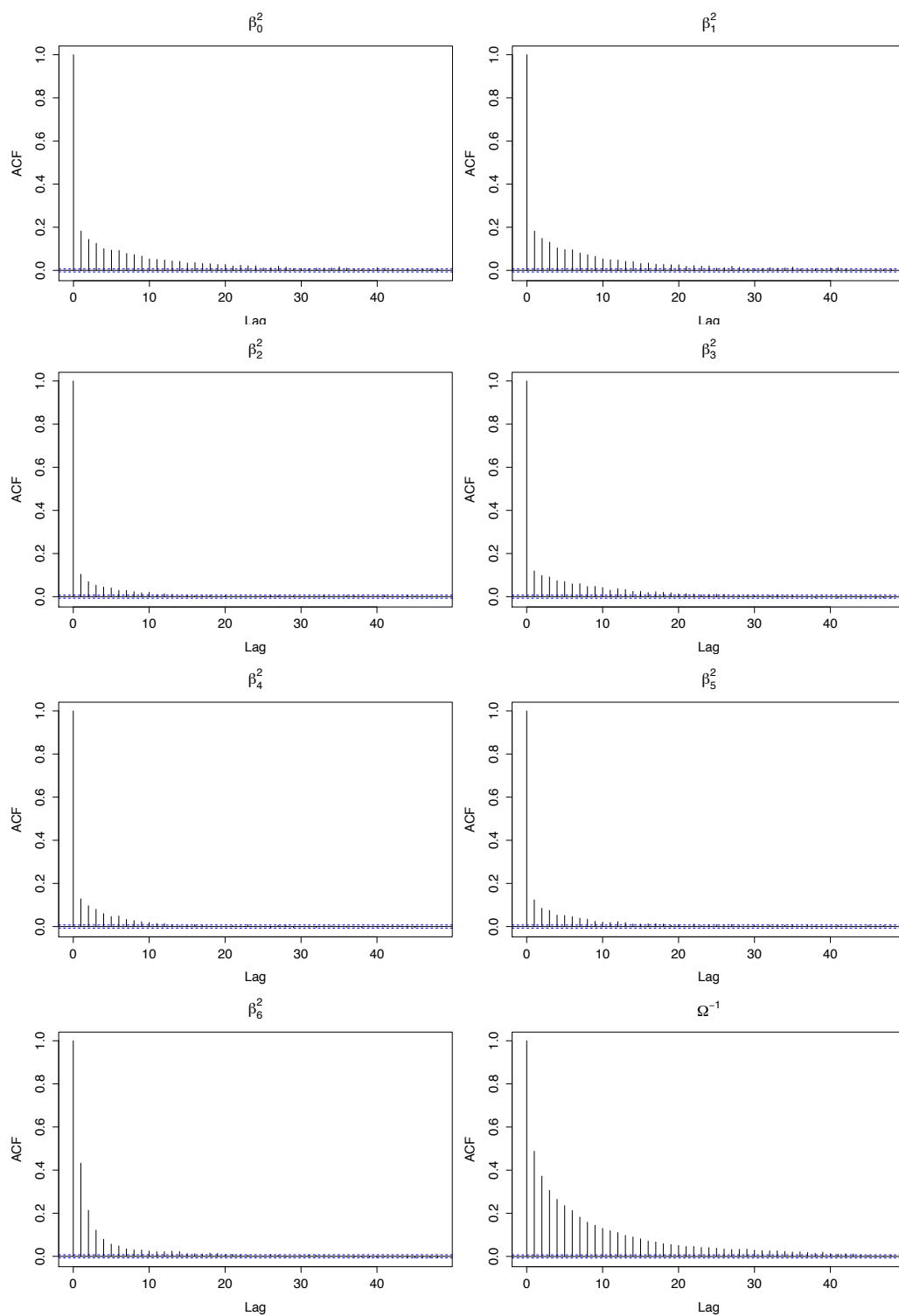
Cuadro 5.3: Intervalos de credibilidad al 95 % para los parámetros

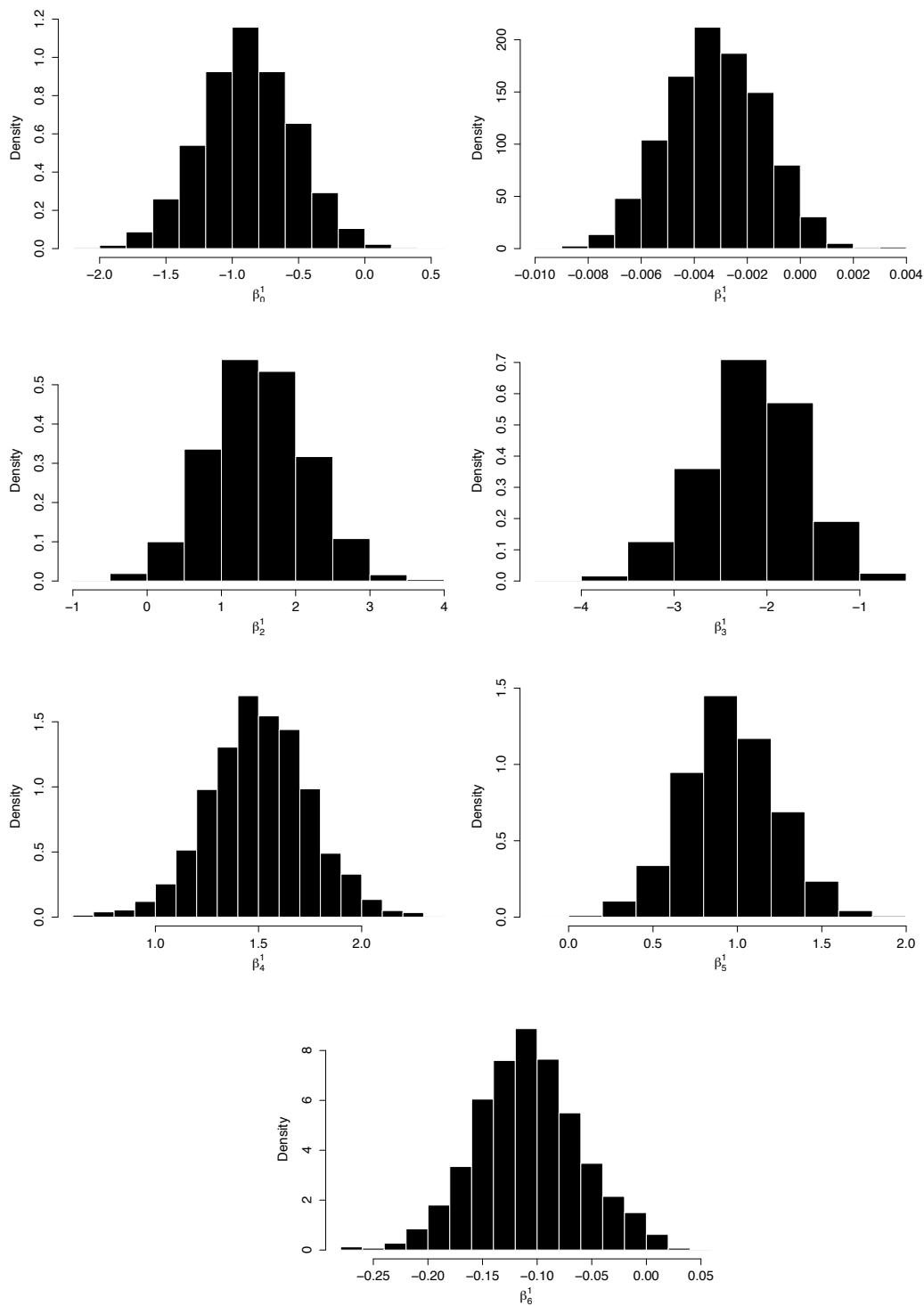
	Componente I	Componente II
β_0	(-1.618352 , -0.252842)	(3.037431 , 5.703815)
β_1	(-0.006491 , 0.000486)	(-0.028834 , -0.014730)
β_2	(0.250883 , 2.764886)	(-0.081541 , 5.636991)
β_3	(-3.274942 , -1.149711)	(-0.591864 , 2.143386)
β_4	(1.012730 , 1.986393)	(-0.664370 , 1.310612)
β_5	(0.395236 , 1.498992)	(0.854931 , 3.005431)
β_6	(-0.200139 , -0.008073)	(-0.298181 , -0.142436)
Ω^{-1}		(0.745315 , 1.950174)
λ		(1.072828 , 2.436998)

Figura 5.15: Promedios ergódicos para la componente I .

Figura 5.16: Promedios ergódicos para la componente II .

Figura 5.17: Autocorrelaciones para la componente I .

Figura 5.18: Autocorrelaciones para la componente *II*.

Figura 5.19: Distribuciones finales para los parámetros de la componente I .

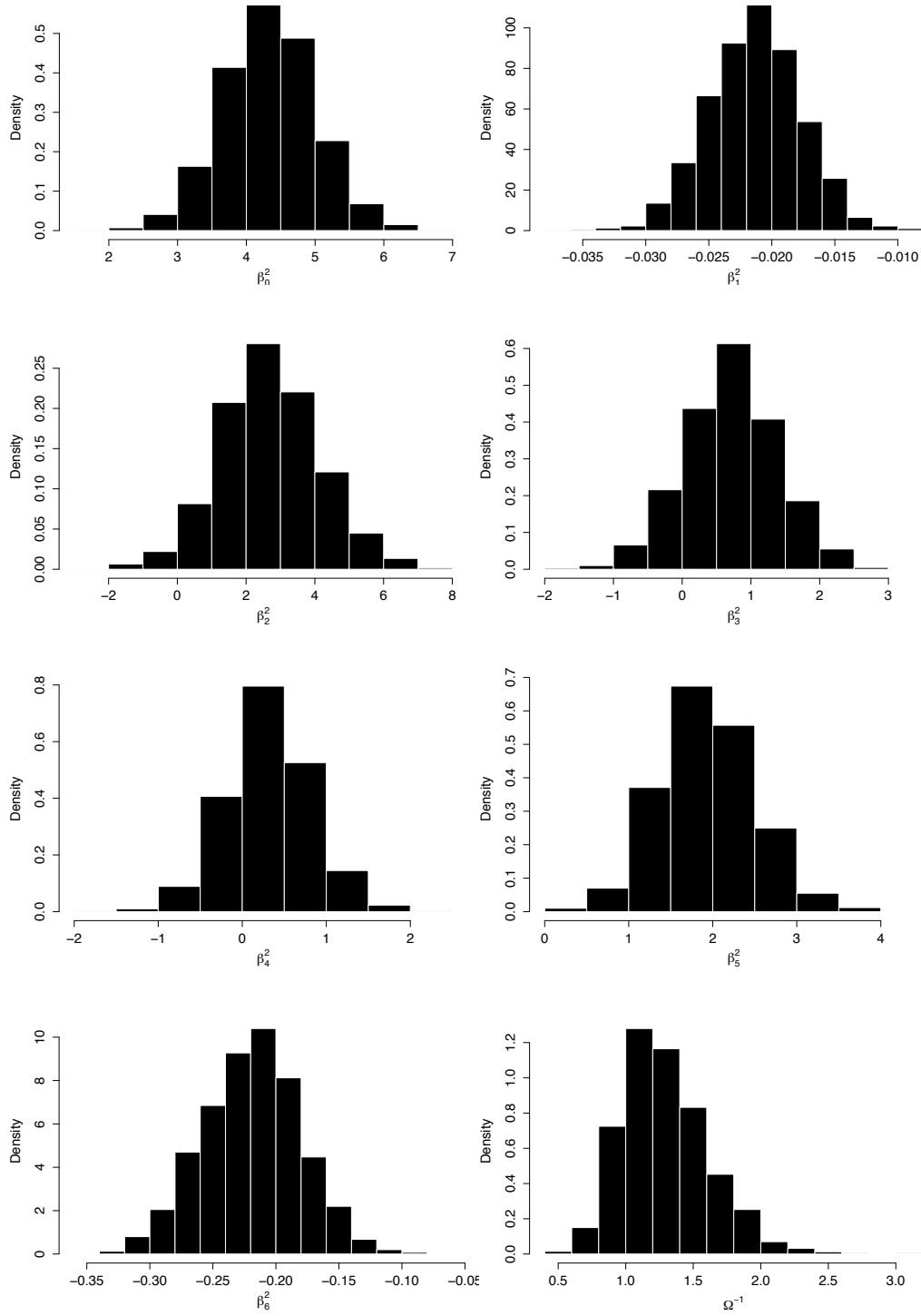


Figura 5.20: Distribuciones finales para los parámetros de la componente *II*.

Conclusiones y perspectivas

Este trabajo de investigación desarrolla una extensión al modelo longitudinal circular proyectado (LCP) propuesto por Núñez Antonio y Gutiérrez-Peña (2014), con el objetivo de obtener una mayor flexibilidad estadística. Dicha extensión consiste en la inclusión de un parámetro extra de variabilidad, que permita estudiar con mayor detalle el comportamiento de los datos longitudinales circulares. En particular, el análisis del modelo se realiza a partir de un enfoque bayesiano.

El modelo propuesto se basa en una versión de la distribución Normal proyectada, la cual se obtiene a partir de la proyección radial de una Normal bivariada. Esta distribución permite representar un vector direccional unitario a partir de un ángulo. Por lo tanto, para el desarrollo de este trabajo se supone que los datos circulares siguen dicha distribución y con lo cual, a través de la inclusión de una variable latente, es posible realizar el análisis a partir del vector Normal bivariado. Este enfoque facilita el estudio sobre los datos circulares.

Por otro lado, es necesario establecer la relación de dependencia que existe en los datos de tipo longitudinal. Existen diversas propuestas para trabajar con datos longitudinales, sin embargo los modelos lineales de efectos mixtos presentan diversas ventajas. Por tal motivo se propone la especificación de un modelo lineal de efectos mixtos para cada componente del vector bivariado.

Si bien el modelo Normal proyectado posee diversas ventajas, es necesario resaltar que es posible encontrar problemas de identificabilidad de los parámetros involucrados. Dado que la distribución Normal proyectada no se modifica para $a > 0$ si se utilizan $\mu^* = a\mu$ y $\Sigma^* = a^2\Sigma$, entonces es necesario establecer restricciones adicionales necesarias para asegurar la identificabili-

dad de los parámetros.

Para evitar problemas de identificabilidad, Núñez Antonio y Gutiérrez-Peña (2014) la matriz identidad como matriz de covarianza. El modelo propuesto en este trabajo introduce un parámetro extra de variabilidad para la segunda componente del vector bivariado, lo cual permite una mayor flexibilidad para el análisis de diversos patrones longitudinales sin sacrificar la identificabilidad del modelo.

Como se ha mostrado, la introducción de dicho parámetro de variabilidad resulta apropiado para mejorar el estudio de los datos longitudinales. Este modelo permite realizar una mejor distinción entre la variabilidad intrínseca de cada sujeto, de la variabilidad poblacional. Además, es posible generar una mayor variedad de patrones longitudinales, lo que muestra que esta extensión del modelo original posee una mayor flexibilidad estadística.

Para la implementación de este modelo se desarrollaron algoritmos que permitieran obtener muestras de las distribuciones finales de los parámetros involucrados. De esta manera, se utilizaron métodos MCMC que permiten realizar inferencia. En particular se utilizaron los algoritmos de muestreo de Gibbs y Metrópolis de independencia. Sin embargo con la introducción del parámetro extra se observó un incremento importante en el número de iteraciones requeridas para conseguir la convergencias de las cadenas y un aumento en las autocorrelaciones de los estados de las mismas. Debido a lo anterior, se consideraron alternativas que permitan acelerar la convergencia de la cadena y mejorar la mezcla de las mismas. Dentro de estas alternativas se implementó el slice sampling, con el cual se redujo el tiempo de procesamiento pero no se observaron mejoras sustanciales en la velocidad de convergencia de las cadenas.

Actualmente se está en proceso de investigación de distintas alternativas de parametrización y métodos de muestreo que permitan realizar una implementación más eficiente del modelo. Con esto se busca lograr que el modelo propuesto sea una propuesta aún más importante para el análisis de datos longitudinales con variable de respuesta circular y para su generalización a otras dimensiones.

Capítulo 6

Anexo A: Programas

Programa de Gibbs para $NP(\mu, \lambda)$.

```
1 # Programa de Gibbs con lambda
2 # version:1
3 # Fecha de modificación: 07.11.2014
4 # Comentarios:
5 rm(list=ls())
6 time.ini<-Sys.time()
7 library(MASS)
8 library(CircStats)
9 #####
10 ##### Declaración de Funciones #####
11 #####
12 erg.mean<-function(x)
13 {
14     n<-length(x)
15     result<-cumsum(x)/(1:n)
16 }
17
18 dcuad<- function(tetha ,lambda2)
19 {
20     drop( (cos(tetha)^2)+lambda2*(sin(tetha)^2) )
21 }
22
23 dval<- function(tetha ,lambda2)
24 {
```

```

25     drop(sqrt( (cos(tetha)^2)+lambda2*(sin(tetha)^2) ) )
26 }
27
28 bval<- function(tetha ,mu1,mu2,lambda2)
29 {
30     drop( ( mu1*cos(tetha)+ ( lambda2*mu2*sin(tetha) ) )/( (cos(
31         tetha)^2)+lambda2*(sin(tetha)^2) ) )
32 }
33 kn<- function(bd){
34     drop( 1.0+( ( bd*pnorm(bd) )/dnorm(bd) ) )
35 }
36
37 moda0<- function(b,bd,d2)
38 {
39     drop( ( b+sqrt(((bd^2)+4)/d2) )/2 )
40 }
41
42 mun<-function(X,n,suma,lambda0,mu0)
43 {
44     drop( ( ( lambda0*mu0 ) + ( n*mean(X) ) )/suma )
45 }
46
47 lambdan<-function(suma,lambda)
48 {
49     drop( lambda*suma )
50 }
51
52 alfan<-function(n,alfa0)
53 {
54     drop( alfa0 + (n/2) )
55 }
56
57 betan<- function(X,n,beta0 ,suma,lambda0 ,mu0)
58 {
59     drop( beta0 + (0.5)*(( (n-1)*var(X) ) + ( ( lambda0*n*(
60         mean(X)-mu0 )^2 )/suma ) ) )
61 }
62 simula.tetha<-function(n,m1,m2,lambda1,lambda2)

```

```

63 {
64     tetha<-rep(0,n)
65     x<-rnorm(n,m1,sqrt(1/lambda1) )
66     y<-rnorm(n,m2,sqrt(1/lambda2))
67     tetha<-(atan2(y,x))%%(2*pi)
68     drop(tetha)
69 }
70
71 #####
72 ##### Declaración de densidad utilizadas #####
73 #####
74 frdt <- function(r,b,bd,d2)
75 {
76     drop( ( d2*r*exp( (-0.5)*d2*( (r^2)-(2*b*r) ) ) )/kn(bd) )
77 }
78
79 Frdt<-function(r,b,bd,d2,d)
80 {
81     ( ( 1.0-exp(d2*(b*r-((r^2)/2)) ) ) + bd*(1/dnorm(bd))*(pnorm(d*(r-b)
82         )-pnorm(-bd)) ) /kn(bd)
83 }
84
85 #####
86 ##### Algoritmos empleados #####
87 #####
88
89 #— Algoritmo de Newton–Rapson —
90 nr<- function(F,f,b,bd,d2,d,n,ite)
91 {
92     r<-moda0(b,bd,d2)
93     u<-runif(n)
94     for (i in 1:ite)
95     {
96         r<-r-(( F(r,b,bd,d2,d) - u )/( f(r,b,bd,d2) ))
97     }
98     drop( r )
99 }
100
101 #— Algoritmo de Gibbs —

```



```

102 gibbs<-function(F,f,tetha,tam,g.burn,g.salto,mu10,mu20,lambda0,alfa0,
    beta0)
103 {
104     total.ite<- g.burn+((g.salto+1)*tam)
105     t.tam<- length(tetha)
106     mu1.ma<<-rep(NA,tam)
107     mu2.ma<<-rep(NA,tam)
108     lambda.ma<<-rep(NA,tam)
109     r.ma<-rep(1,t.tam)
110     lsum<-lambda0+t.tam
111
112     mu1.e<-1
113     mu2.e<-1
114     lambda2.e<-0.0001
115     l<-1
116     for(i in 1:total.ite)
117     {
118         for(j in 1:t.tam)
119         {
120             b.temp <-bval(tetha[j],mu1.e,mu2.e,lambda2.e)
121             d2.temp<-dcuad(tetha[j],lambda2.e)
122             d.temp<-dval(tetha[j],lambda2.e)
123             bd.temp<-(b.temp)*(d.temp)
124             r.ma[j]<- nr(Frdt,frdt,b.temp,bd.temp,d2.temp
                ,d.temp,1,5)
125         }
126         X<-(r.ma*cos(tetha))
127         Y<-(r.ma*sin(tetha))
128
129         mu1.e<- rnorm(1,mun(X,t.tam,lsum,lambda0,mu10),sqrt
            (1/(lsum)))
130         lambda2.e<-rgamma(1,shape=alfan(t.tam,alfa0),scale
            =(1/betan(Y,t.tam,beta0,lsum,lambda0,mu20)))
131         cat("lambda es ",lambda2.e," \n")
132     mu2.e<-rnorm(1,mun(Y,t.tam,lsum,lambda0,mu20),sqrt(1/lambda(lsum
        ,lambda2.e)))
133     if((g.burn<i) & (l<=tam) &(((i-g.burn)%g.salto+1)
        ==0))
134     {
135         lambda.ma[1]<<- lambda2.e

```

```

136             mu1.ma[l]<<-mu1.e
137             mu2.ma[l]<<-mu2.e
138             l<-l+1
139         }
140         if(i%%1000==0)
141         {
142             cat(" iteracion=", i," valor1=",mu1.ma[l-1]," valor2
143                 =",mu2.ma[l-1],"\\n")
144         }
145     }
146 }
147
148 #####
149 ##### Programa Principal #####
150 #####
151 #— Simulación de datos tetha originales
152 if(1)
153 {
154     tetha<-simula.tetha(100,-1.5,2.5,1,2.5)
155     write(tetha,file="Dat7.R")
156 }
157 #— Lee datos del archivo
158 datos<-scan(file="Dat6.R")
159 #— Valores apriori —
160 mu10<-0.0
161 mu20<-0.0
162 lambda0<-0.001
163 alfa0<-0.1
164 beta0<-0.1
165
166 #— Valores iniciales para Gibbs —
167 calentamiento<-9000
168 salto<-4
169 tam.muestra<- 1000
170
171 #— Uso de gibbs para ontener muestra de parámetros —
172 gibbs(Frdt,frdt,datos,tam.muestra,calentamiento,salto,mu10,mu20,
173     lambda0,alfa0,beta0)
173 time.fin<-Sys.time()

```

```
174 time.tot <- time.fin - time.ini
175 time.tot
176 par(mfrow=c(3,1))
177 hist(mu1.ma)
178 hist(mu2.ma)
179 hist(lambda.ma)
```

Bibliografía

- [1] BATSCHELET, E. (1981). *Circular statistics in biology*. London: Academic Press.
- [2] BERNARDO, J. M. (1981). *Bioestadística: Una perspectiva bayesiana*. Barcelona: Vicens-Vives.
- [3] BERNARDO, J. M., & SMITH, A. F. M. (2000). *Bayesian theory*. Chichester: Wiley.
- [4] CONGDON, P. (2014) *Applied Bayesian modelling*. Chichester: Wiley
- [5] CONGDON, P. (2006) *Bayesian statistical modelling*. Chichester: John Wiley & Sons.
- [6] BOLSTAD, W. M. (2007). *Introduction to Bayesian statistics*. Hoboken, N.J: Wiley-Interscience.
- [7] BOX, G. E. P., & TIAO, G. C. (1973). *Bayesian inference in statistical analysis*. Reading, Mass: Addison-Wesley Pub. Co.
- [8] DIGGLE P. J., LIANG K. & ZEGER S.L. (2002). *Analysis of Longitudinal Data*. New York: Oxford University Press, Inc.
- [9] FISHER, N. I., LEWIS, T., & EMBLETON, B. J. J. (1987). *Statistical analysis of spherical data*. Cambridge: University Press.
- [10] FITZMAURICE G.M., LAIRD, N.M. Y WARE, J.H(2004). *Applied longitudinal Analysis*. Hoboken, N.J: Wiley-Interscience. Cambridge: University Press.

- [11] FITZMAURICE G.M., DAVIDIAN, M., VERBEKE, G. & MOLENBERGHS, G.(2009). *Longitudinal Data Analysis*. Boca Raton: Chapman & Hall/CRC. Cambridge: University Press.
- [12] GELMAN, A. B., CARLIN, J. B., STERN, H. S., & RUBIN, D. B. (2004). *Bayesian Data Analysis*. Boca Raton: Chapman & Hall/CRC.
- [13] GHOSH, J. K., DELAMPADY, M., & SAMANTA, T. (2006). *An introduction to Bayesian analysis: Theory and methods*. New York: Springer.
- [14] HEDEKER D. Y GIBBONS R.D. (2006). *Longitudinal Data Analysis*. New Jersey: John Wiley & Sons, Inc.
- [15] HOFF, P. D. (2009). *A first course in Bayesian statistical methods*. New York: Springer.
- [16] JAMMALAMADAKA, S. R., & SENGUPTA, A. (2001). *Topics in circular statistics*. Singapore: World Scientific.
- [17] KOCH, K.-R. (2007). *Introduction to Bayesian statistics*. Berlin: Springer.
- [18] LAIRD, N.M. & WARE, J.H. (1982). Random-effects models for longitudinal data. *Biometrics*. 38, 963-974.
- [19] MARDIA, K. V., & JUPP, P. E. (2000). *Directional Statistics*. Chichester: Wiley.
- [20] NÚÑEZ-ANTONIO, G., & GUTIÉRREZ-PEÑA, E. (2005). A Bayesian analysis of directional data using the projected normal distribution. *Journal of Applied Statistics*. 32,10, 995-1001.
- [21] NÚÑEZ-ANTONIO, G. (2010). *Análisis Bayesiano de Modelos Lineales para Datos Direccionales considerando la Distribución Normal bajo Proyección*. (Tesis Doctoral). Universidad Autónoma Metropolitana, Distrito Federal, México.
- [22] NÚÑEZ-ANTONIO, G., & GUTIÉRREZ-PEÑA, E. (2014). A Bayesian Model for Longitudinal Circular Data based on the Projected Normal Distribution. *Computational Statistics and Data Analysis*. 71, 506–519.

- [23] PRESS, S. J. (2003) *Subjective and objective Bayesian statistics: Principles, models, and applications*. Hoboken, N.J: Wiley-Interscience.
- [24] VERBEKE, G. & MOLENBERGS, G.(2000) *Linear mixed models for longitudinal data*. New York: Springer- Verlag.