

*Modelos Para Datos Longitudinales: Una Propuesta
Bayesiana*

EDWIN JAVIER CASTILLO CARREÑO
MATEMÁTICO



UNIVERSIDAD NACIONAL DE COLOMBIA
FACULTAD DE CIENCIAS
DEPARTAMENTO DE ESTADÍSTICA
BOGOTÁ, D.C.
NOVIEMBRE DE 2017

*Modelos Para Datos Longitudinales: Una Propuesta
Bayesiana*

EDWIN JAVIER CASTILLO CARREÑO
MATEMÁTICO

DISERTACIÓN PRESENTADA PARA OPTAR AL TÍTULO DE
MAGISTER EN CIENCIAS, ESTADÍSTICA

DIRECTOR
EDILBERTO CEPEDA CUERVO, PH.D.
DOCTOR EN ESTADÍSTICA

LÍNEA DE INVESTIGACIÓN
ESTADÍSTICA BAYESIANA



UNIVERSIDAD NACIONAL DE COLOMBIA
FACULTAD DE CIENCIAS
DEPARTAMENTO DE ESTADÍSTICA
BOGOTÁ, D.C.
NOVIEMBRE DE 2017

Título en español

Modelos Para Datos Longitudinales: Una Propuesta Bayesiana.

Title in English

Models for Longitudinal Data: A Bayesian Proposal

Resumen: En este trabajo se presenta una propuesta Bayesiana para la estimación conjunta de estructuras de media y matriz de varianza y covarianza para modelos con datos longitudinales donde se suponen estructuras de covarianza SC , $AR(1)$, $ARMA(1,1)$ y algunos casos de estructuras de antedependencia no estructurada de orden 1 $ADS(1)$. Se generaliza el modelo para datos longitudinales con estructuras de antedependencia estructurada más general que la propuesta en Zimmerman, Núñez-antón & El-Barmi (1998). Se validan las distintas propuestas mediante simulaciones y aplicaciones a datos sugeridos en la literatura.

Abstract: In this work, is presented a Bayesian proposal for the joint estimation of structures of mean and variance and covariance matrix for models with longitudinal data where we assume structures of covariance SC , $AR(1)$, $ARMA(1.1)$ and some cases of structures of antedependence unstructured of order 1 $ADS(1)$. Is generalized a model for longitudinal data with structures of structured antedependence more general than the one proposed in citeasnoun Zimmerman2. The different proposals are validated through simulations and applications to data suggested in the literature.

Palabras clave: Datos Longitudinales, Estructuras de Covarianza, Bayesiana, Metropolis-Hasting, Antedependencia Estructurada

Keywords: Longitudinal Data, Covariance Structures, Bayesian, Metropolis-Hasting, Structured Antedependence

Dedicado a

A mis padres, gestores fundamentales de mi amor por la academia.

Agradecimientos

A todos los profesores que me guiaron por este camino. En especial a mi director, por sus consejos y comentarios permanentes.

A mi familia que ha sido mi apoyo constante en cada uno de los instantes de mi vida académica. A mis amigos, los cuales compartieron conmigo su conocimiento y su apoyo incondicional durante cada instante de esta maestría.

Por último a la universidad Nacional de Colombia la que me permitió realizar este posgrado.

Índice general

Índice general	I
Índice de tablas	III
Índice de figuras	IV
Introducción	V
1. Marco Teórico	1
1.1. Datos Longitudinales	1
1.1.1. Modelo con Estructura de Covarianza Específica	1
1.1.2. Estimación del Modelo	2
1.1.2.1. Máxima Verosimilitud Restringida (MVR)	3
1.2. Inferencia Bayesiana	5
1.2.1. Metropolis Hasting (MH)	5
2. Algunas Estructuras de Covarianza	8
2.1. Estructura de Simetría Compuesta (SC)	9
2.2. Estructura Autoregresiva de orden 1 AR(1)	10
2.3. Estructura Autoregresiva de Promedio Móvil de orden (1,1), ARMA(1,1)	11
2.4. Modelos de Antedependencia	12
2.4.1. Modelo de Antedependencia Estructurado ADS(1)	12
2.5. Sobre la Estimación de las Modelos de Covarianza SC, AR(1), ARMA(1,1), AD(1) y ADS(1)	13
3. Propuestas Bayesianas	15
3.1. Propuesta Bayesiana para la Estructura de Media	16

3.2. Propuesta Bayesiana para la Estimación de Modelos Longitudinales con Estructura de Covarianza SC y AR(1)	17
3.3. Propuesta Bayesiana para la Estimación de Modelos Longitudinales con Estructura de Covarianza ARMA(1,1)	19
3.4. Propuesta Bayesiana para la Estimación de Modelos Longitudinales con Estructura de Covarianza ADS(1)	20
3.4.1. Modelo Antedependiente Estructurado Tipo 1	20
3.4.2. Modelo Antedependiente Estructurado Tipo 2	22
3.4.3. Modelo Antedependiente Estructurado Tipo 3	23
4. Simulaciones y Aplicaciones	25
4.1. Simulaciones	25
4.1.1. Modelo de Simetría Compuesta	25
4.1.1.1. Una Distribución Apriori Informativa para la Correlación	27
4.1.2. Modelo AR(1)	28
4.1.2.1. Una Distribución Apriori Informativa para la Correlación	30
4.1.3. Modelo ARMA(1,1)	31
4.2. Aplicaciones	33
4.2.1. Aplicación sobre Small Mice Data	33
4.2.1.1. Small Mice Data con Estructura de Simetría Compuesta	35
4.2.1.2. Small Mice Data con Estructura AR(1)	36
4.2.1.3. Small Mice Data con Estructura ARMA(1,1)	37
4.2.2. Aplicaciones a los datos de Reconocimiento de voz	39
4.2.2.1. Modelo ADE(1) Tipo 1	40
4.2.2.2. Modelo ADE(1) Tipo 2	41
4.2.3. Datos Carrera de 100 Km	44
4.2.4. Estudio de Sensibilidad	50
Conclusiones	51
Trabajo Futuro	53
Bibliografía	54

Índice de tablas

4.1. Valores Estimados simulación SC	26
4.2. Valores Estimados simulación SC (Apriori Informativa)	27
4.3. Gráficos Diagnosticos Simulación AR(1)	29
4.4. Valores Estimados Simulación AR(1) (Apriori Informativa)	30
4.5. Valores Estimados ARMA(1,1)	32
4.6. Matriz de Correlación para el SMD	34
4.7. Valores Estimados por Propuesta Bayesiana a SMD con Estructura SC y Valores Estimados Mediante MVR	35
4.8. Valores Estimados por Propuesta Bayesiana a SMD con Estructura AR(1) y Valores Estimados Mediante MVR	37
4.9. Valores Estimados por Propuesta Bayesiana a SMD con Estructura AR- MA(1,1) y Valores Estimados Mediante MVR	38
4.10. Criterios de Comparación	39
4.11. Valores Estimados Reconocimiento de Voz (ADS(1)) Modelo 1	41
4.12. Valores Estimados Estructura Media Reconocimiento de Voz (ADS(1)) Mo- delo 2	42
4.13. Valores Estimados Varianzas y Covarianzas Reconocimiento de Voz (ADS(1)) Modelo 2	42
4.14. Correlaciones y Varianzas Muestrales <i>Carrera 100 Km</i>	44
4.15. Valores Estimados Estructura de la Media <i>Carrera 100 km ADS(1)</i>	46
4.16. Valores Estimados Estructura de Covarianzas <i>Carrera 100 km ADS(1)</i>	46
4.17. Valores Estimados Ψ en la Estructura de Covarianzas <i>Carrera 100 km ADS(1)</i>	46
4.18. Valores Estimados Varianzas por Sección Carrera 100 km <i>ADS(1)</i>	47
4.19. Varianzas Estimadas y sus Desviaciones Estándar Para Distintos Valores de los Hiperparámetros	50

Índice de figuras

4.1. Gráficos Diagnosticos simulación SC	27
4.2. Gráficos Diagnosticos simulación SC (apriori Informativa)	28
4.3. Gráficos Diagnosticos Simulación AR(1)	30
4.4. Gráficos Diagnosticos Simulación AR(1) (Apriori Informativa)	31
4.5. Gráficos Diagnosticos ARMA(1,1)	33
4.6. Perfiles <i>SMD</i>	34
4.7. Gráficos Diagnosticos SMD SC	36
4.8. Gráficos Diagnosticos SMD AR(1)	37
4.9. Gráficos Diagnosticos SMD ARMA(1,1)	38
4.10. Perfiles <i>Reconocimiento de Voz</i>	40
4.11. Gráficos Diagnósticos Reconocimiento de Voz (ADS(1)) Modelo 1	41
4.12. Gráficos Diagnósticos Estructura Media Reconocimiento de Voz (ADS(1)) Modelo 2	43
4.13. Gráficos Diagnósticos Varianzas y Covarianzas Reconocimiento de Voz (ADS(1)) Modelo 2	43
4.14. Perfiles <i>Carrera 100 Km.</i>	44
4.15. Gráficos Diagnosticos Carrera 100 Km <i>ADS(1)</i>	47
4.16. Residuales Carrera <i>100 Km ADS(1)</i>	48
4.17. Box-plot de los Residuales por Sección Carrera <i>100 Km ADS(1)</i> , Ajustados Mediante MVR.	49
4.18. Box-plot de los Residuales por Sección Carrera <i>100 Km ADS(1)</i> , Ajustados Mediante la 3.4.3.	49

Introducción

Los datos longitudinales consisten en observaciones repetidas a lo largo del tiempo sobre varios individuos. Estas medidas repetidas sobre cada individuo se consideran correlacionadas. Distintos métodos estadísticos son propuestos para el manejo de covarianzas estacionarias y no estacionarias, así como también para el trabajo sobre conjuntos de datos balanceados o desbalanceados (Fitzmaurice (2008), Verbeke & Molenberghs (2009), Weiss (2005), Diggle (2002)).

Un modelo de regresión para datos longitudinales se define al suponer que la variable respuesta Y_i puede ser explicada mediante el modelo.

$$\mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\beta} + \boldsymbol{\epsilon}_i, \quad i = 1, \dots, n,$$

donde \mathbf{Y}_i es el vector de la respuesta $m_i \times 1$ para el individuo i , \mathbf{X}_i es una matriz $m_i \times q$ de rango q que contiene los valores de las variables explicativas para el i -ésimo individuo; $\boldsymbol{\epsilon}_i$ es un vector que se distribuye multinormal con media 0 y matriz de varianzas y covarianzas $\boldsymbol{\Sigma}(\boldsymbol{\theta})$; mientras, $\boldsymbol{\theta} = (\theta_1, \dots, \theta_k)'$ y $\boldsymbol{\beta} = (\beta_1, \dots, \beta_m)$ son vectores de parámetros desconocidos para la varianza y la media del modelo respectivamente; m_i indica la cantidad de observaciones sobre el individuo i . Si la cantidad de observaciones sobre todos los individuos es la misma $m_i = m, \forall i$, entonces, se denomina un estudio balanceado, (Las observaciones no siempre son igualmente espaciadas) y n representa el número de individuos en el estudio los cuales no presentan correlación entre ellos (Núñez-Antón & Zimmerman (2001)).

Distintos métodos son utilizados para el análisis del modelo de regresión anteriormente descrito, tres tipos de modelos se usán con mayor frecuencia en el estudio de la estructura de covarianza interna de cada individuo: estructura multivariada, efectos aleatorios y modelos autoregresivos estacionarios (Fitzmaurice (2008)). De ellos el modelo de efectos aleatorios es el más estudiado ya que explica la correlación intrínseca de los datos longitudinales mediante un modelo mixto (Verbeke & Molenberghs (2009)), no obstante, se han desarrollado de manera paralela los modelos paramétricos para la estructura de la matriz de varianzas y covarianzas (Núñez-Antón & Zimmerman (2001)).

El modelo multivariado es útil cuando se tienen diseños desbalanceados y datos faltantes, sin embargo, presenta algunos inconvenientes cuando el número de medidas en el tiempo para cada individuo es mucho más grande que la cantidad de individuos, además de generar modelos poco parsimoniosos. Por otro lado el ajuste de la estructura de medias y de varianzas suele realizarse por medio de máxima verosimilitud usando métodos numéri-

cos tales como el algoritmo de Newton-Rhapson o el algoritmo de Esperanza-Maximización (Ware (1985)).

El modelo de efectos aleatorios, presentado por (Rao (1975)), para el análisis de curvas de crecimiento, brinda óptimos resultados para estudios balanceados y encuentra estimaciones mucho más parsimoniosas que el modelo multivariado. Las estimaciones son comúnmente halladas a través de la implementación de técnicas de máxima verosimilitud, máxima verosimilitud restringida o mínimos cuadrados ordinarios.

Los algoritmos computacionales usados para la estimación de este tipo de modelos a pesar de hacer uso del algoritmo Esperanza Maximización o del método de Newton-Rhapson tienen menos complejidad computacional que los algoritmos usados en la estimación de los modelos multivariados (Ware (1985)).

Los modelos autoregresivos se proponen como una herramienta cuando las observaciones en el tiempo son bastantes en comparación con la cantidad de individuos involucrados en el estudio. En primer lugar su ajuste es más práctico ya que se puede plantear como un modelo dinámico. Además, el uso de técnicas combinadas, como lo son el método de mínimos cuadrados y la máxima verosimilitud, permiten que las técnicas numéricas para el ajuste sean más eficaces (Ware (1985)).

Cuando se trata de ajustar un modelo de regresión sobre datos longitudinales mediante la especificación de la estructura de la matriz de varianzas covarianzas, se tienen muchas alternativas para la selección de dicha estructura; y se pueden considerar estructuras tanto estacionarias como no estacionarias y heterogéneas como no heterogéneas. Dentro de la consideración de estructuras con estacionariedad en la varianza o en la correlación, se destacan las estructuras de *simetría compuesta (SC)*, *autoregresivas de orden p AR(p)*, *autoregresivas de promedio móvil de orden (p,q) (ARMA(p,q))* (Weiss (2005)). Respecto a la no estacionariedad en la varianza y/o correlación, las versiones heterogéneas de los modelos anteriormente mencionados no siempre son la mejor alternativa para modelar la matriz de varianzas. Por lo tanto, es necesario implementar modelos más generales como el modelo *autoregresivo integrado de media móvil (ARIMA)* o generalizaciones de los modelos autoregresivos como los modelos *antedependientes de orden p AD(p)* o la versión estructurada de los mismos, *antedependientes estructurados de orden p ADS(p)* (Núñez-Antón & Zimmerman (2001), Zimmerman & Núñez-Antón (2009)). De este modo, las estimaciones se realizan mediante el método de máxima verosimilitud restringida aplicando algoritmos recursivos. Algunas herramientas computacionales como SAS o SPSS estiman esta clase de modelos.

Las Propuestas bayesianas aplicadas a la estimación de modelos de regresión lineal en distintos contextos, se han desarrollado en los últimos años. Cepeda & Gamerman (2001) implementan la metodología bayesiana presentada en Gamerman & Lopes (2006) para ajustar de manera conjunta los parámetros de la media y la matriz de varianza en una regresión heteroscedástica. Posteriormente en Cepeda & Gamerman (2004), Cuervo (2001) se extiende la metodología de estimación conjunta para modelos más generales. Las propuestas de estimación bayesiana para modelos de regresión con datos longitudinales en donde no se tienen en cuenta una estructura específica para la matriz de varianzas y covarianzas son presentadas en Brown, Kenward & Bassett (2001) asumiendo para esta la distribución a priori inversa de Wishart. También, Hui & Berger (1983) presentan la estimación conjunta de estructura de medias y algunas estructuras simples de covarianza mediante la implementación de aprioris normales para la media y gamma inversa para la varianza general del modelo, los resultados obtenidos se usan para ajustar un modelo de

curvas de crecimiento en los datos de la pérdida de calcio óseo en mujeres post menopausia. Daniels & Pourahmadi (2002) desarrolla una propuesta en la cual se consideran aprioris normales tanto para la estructura de medias como para la matriz de varianza, pero no se incluye ningún tipo de algoritmo específico.

En (Cepeda-Cuervo & Núñez-Antón (2007), Cuervo (2001), Cepeda Cuervo (2011)) se hace una extensión del algoritmo presentado en (Gamerman (1997)), el cual permite estimar de manera conjunta los parámetros de la media y los de la varianza sobre modelos longitudinales en los cuales se considera una estructura de antedependencia no estructurada. Para este fin, se hace uso de la descomposición matricial que se trabaja en Pourahmadi (1999) y la inclusión de variables de trabajo. Los algoritmos se presentan detalladamente y se incluyen simulaciones y casos de estudio para comparar los resultados de estas propuestas bayesianas con la estimación clásica.

En este trabajo, se extiende la metodología bayesiana propuesta en (Cepeda Cuervo (2011)), realizando la estimación conjunta de media y matriz de varianzas y covarianzas en una regresión de datos longitudinales bajo el supuesto de normalidad, suponiendo distintas estructuras para la matriz de varianzas. Las propuestas se realizan de manera constructiva empezando por propuestas sobre modelos con varianzas y/o correlaciones estacionarias como los son el modelo de SC , $AR(1)$ y $ARMA(1,1)$, para luego trabajar sobre la generalización no estacionaria de estos modelos suponiendo que la matriz de varianzas sigue un modelo de antedependencia estructurada de orden 1 $ADS(1)$.

Para cada una de las estructuras se presenta detalladamente el algoritmo mediante el cual se obtiene la estimación de los parámetros. Además, en todos los casos se hace uso del muestreo de Gibbs y del algoritmo de Metropolis-Hasting por bloques. Para mostrar el desempeño de los algoritmos propuestos para la estimación si se consideran estructuras de la matriz de varianzas SC , $AR(1)$ y $ARMA(1,1)$. Se presentan tanto simulaciones como aplicaciones sobre el conjunto de datos balanceados *Small Mice Data* presentados en (Izenman & Williams (1989)) y estudiados en (Weiss (2005)). Respecto a las propuestas para la estructura $ADS(1)$ dado que varían según la aplicación, se realiza el ajuste y comparación de dos modelos basados en los datos del *Reconocimiento de Voz*, los cuales son presentados en (Tyler (1988)) y ajustados en (Núñez-antón & Woodworth (1994), Zimmerman et al. (1998)). Además, se realiza el ajuste y comparación del modelo presentado en (Núñez-antón & Woodworth (1994)) para los datos de *Carrera de 100 km*.

En el capítulo 1 se realiza una aclaración del contexto y de las herramientas utilizadas en el proceso de estimación tanto desde la perspectiva clásica como las que se usan en el desarrollo de los capítulos posteriores. En el capítulo 2 se presentan las principales características y generalidades de las estructuras de covarianza para las cuales se proponen la metodología bayesiana. En el capítulo 3 se muestra el desarrollo de los algoritmos propuestos. En el capítulo 4 se muestran los resultados obtenidos mediante la implementación de la metodología bayesiana propuesta haciendo uso de simulaciones y ajustes de datos que permiten comparar la metodología clásica con la metodología acá propuesta. Finalmente en el capítulo 5 se presentan las respectivas conclusiones y perspectivas del trabajo próximo.

CAPÍTULO 1

Marco Teórico

1.1. Datos Longitudinales

Algunos autores como Weiss (2005) y Fitzmaurice (2008) definen los datos longitudinales como mediciones tomadas sobre varios individuos en más de dos instantes del tiempo. En este tipo de datos, se considera que las mediciones intraindividuo están correlacionadas mientras que las mediciones entre individuos se consideran no correlacionadas.

Los métodos estadísticos utilizados para el análisis de modelos de regresión con este tipo de datos se pueden realizar desde la perspectiva del modelo de efectos aleatorios dada la naturaleza multinivel intrínseca de los estudios longitudinales (Verbeke & Molenberghs (2009)) o dotando de estructuras específicas la matriz de varianzas covarianzas. En este documento, se presenta el ajuste de modelos de regresión lineal simple con estructura longitudinal sobre algunos modelos en los cuales se especifica la estructura de covarianza.

1.1.1. Modelo con Estructura de Covarianza Específica

El modelo de efectos aleatorios es la herramienta más utilizada en el estudio de datos longitudinales. El tratamiento de este modelo se detalla en textos como (Verbeke & Molenberghs (2009), Diggle (2002), Fitzmaurice (2008)). Sin embargo, en algunas ocasiones se hace necesario dotar a la matriz de varianzas y covarianzas de alguna estructura, definiendo el modelo de la siguiente manera:

Sea \mathbf{Y}_{it} la medición realizada en el sujeto i -ésimo en el tiempo t , sea $\mathbf{Y}_i = (\mathbf{Y}_{i1}, \mathbf{Y}_{i2}, \dots, \mathbf{Y}_{im})'$ y $\mathbf{t}_i = (t_{i1}, t_{i2}, \dots, t_{im})'$ el vector que contiene los tiempos correspondientes en que se realizaron las m mediciones sobre el individuo i . Además, se supone que se observa un vector p variante de covariables asociadas con \mathbf{Y}_{ij} , \mathbf{x}_{ij} .

Usando una notación matricial, sean $N = nm$, $\mathbf{Y} = (\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_m)'$, $\mathbf{t} = (\mathbf{t}_1, \mathbf{t}_2, \dots, \mathbf{t}_m)'$, $\mathbf{X}_i = (\mathbf{x}_{i1}, \mathbf{x}_{i2}, \dots, \mathbf{x}_{im})'$, $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n)$ y sea $\sigma^2 \mathbf{V} = \Sigma$ una matriz diagonal por bloques con bloques no nulos $\sigma^2 \mathbf{V}_0$ de tamaño $m \times m$, donde cada bloque no nulo representa la matriz de varianzas y covarianzas para cada individuo.

Siguiendo Núñez-antón & Woodworth (1994), si se supone :

- Las mediciones realizadas entre individuos distintos son independientes entre sí, aunque las mediciones en un solo individuo pueden ser dependientes .
- La respuesta media (para los individuos) es una combinación lineal de funciones de las covariables y/o el tiempo.
- Las respuestas están normalmente distribuidas.

Se tiene el modelo

$$\mathbf{Y} \sim NMV(\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma}(\boldsymbol{\alpha})) \quad (1.1)$$

Donde $\boldsymbol{\beta}$ es un vector p -dimensional de parámetros fijos y desconocidos.

En los modelos de efectos aleatorios, la estructura de varianzas y covarianzas proviene de regresiones que varían con cada individuo del estudio, mientras que el modelo con estructura de varianzas y covarianzas específica, considera la similitud intraindividuos. El modelo de efectos aleatorios, por su parte, ha sido mayormente usado de modo que brinda gran utilidad en los estudios longitudinales. Aunque los modelos que consideran una estructura paramétrica en su matriz de varianzas y covarianzas ofrecen algunas ventajas:

1. Permite estimar de manera más eficiente los parámetros relacionados con la media del modelo.
2. Ofrece estimaciones más adecuadas sobre los errores estándar de los estimadores que modelan la estructura de la media.
3. En algunas ocasiones son la solución a la estimación de modelos con datos faltantes o donde los tiempos no sean los mismos para todos los individuos del estudio.
4. Puede usarse incluso cuando el número de mediciones repetidas en el tiempo es bastante grande comparado con el número de individuos.

Núñez-antón & Woodworth (1994)

En (Weiss (2005) , Núñez-antón & Woodworth (1994)), se presentan algunas estructuras para la matriz de varianzas y covarianzas: Simetría Compuesta (*SC*), Intercepto Aleatorio (*IA*), Autoregresiva de orden p (*AR(P)*), Independiente (*ID*), Incrementos Independientes (*II*), Antedependientes (*ADP*), Toeplitz(*TL*), Factor Analítico (*FA*), Autoregresivas de Promedio Móvil de Orden (p,q) *ARMA(p,q)*.

Algunas de estas estructuras son tratadas en este documento y se definen de manera detallada en el capítulo 2.

1.1.2. Estimación del Modelo

La estimación del modelo para datos longitudinales puede hacerse utilizando el método de mínimos cuadrados o mínimos cuadrados ponderados, aunque autores como (Verbeke & Molenberghs (2009)) sugieren la estimación por medio de máxima verosimilitud y máxima verosimilitud restringida.

Dado el modelo 1.1 donde $\Sigma = \sigma^2 V$ bajo el supuesto de que los errores siguen una distribución normal, la log-verosimilitud de los datos viene dada:

$$L(\boldsymbol{\beta}, \sigma^2, \mathbf{V}_0) = -\frac{1}{2} \{nm \log(\sigma^2) + n \log(|\mathbf{V}_0|) + \sigma^{-2} (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})' \mathbf{V}^{-1} (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})\} \quad (1.2)$$

Así, el proceso de maximización de la función de log-verosimilitud para $\boldsymbol{\beta}$ y σ^2 para \mathbf{V}_0 conocido, se obtienen las estimaciones:

$$\boldsymbol{\beta}(\hat{\mathbf{V}}_0) = (\mathbf{X}' \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}' \mathbf{V}^{-1} \mathbf{Y} \quad (1.3)$$

$$\hat{\sigma}^2(\mathbf{V}_0) = \frac{SCE(\mathbf{V}_0)}{n * m} \quad (1.4)$$

$$\text{Donde } SCE(\mathbf{V}_0) = \{ \mathbf{Y} - \mathbf{X} \hat{\boldsymbol{\beta}}(\mathbf{V}_0) \}' \mathbf{V}^{-1} \{ \mathbf{Y} - \mathbf{X} \hat{\boldsymbol{\beta}}(\mathbf{V}_0) \}.$$

Al reemplazar los valores estimados de $\boldsymbol{\beta}(\hat{\mathbf{V}}_0)$ y $\hat{\sigma}^2(\mathbf{V}_0)$ en la ecuación 1.2 se tiene:

$$L(\hat{\boldsymbol{\beta}}(\mathbf{V}_0), \hat{\sigma}^2(\mathbf{V}_0), \mathbf{V}_0) = -\frac{1}{2} m \{ n \log(SCE(\mathbf{V}_0)) + n \log(|\mathbf{V}_0|) \} \quad (1.5)$$

Las estimaciones de máxima verosimilitud de $\boldsymbol{\beta}$ y σ^2 se obtendrán al optimizar la función 1.5 respecto a los diferentes parámetros que contenga \mathbf{V}_0 . Esta optimización, en algunos casos, debe realizarse mediante la implementación de métodos numéricos como por ejemplo el algoritmo de Newthton-Rhapsson, ya que para la estructura de matriz de varianzas y covarianzas que se trate, puede que no existan expresiones algebraicas para las estimaciones (Diggle (2002)).

El método de máxima verosimilitud anteriormente descrito no siempre funciona de la mejor manera, ya que, algunos problemas como la alta dimensionalidad de la matriz de covariables o si se tienen diferentes cantidades de observaciones para cada individuo (Estudios desbalanceados), son algunas de las causas de que el método de máxima verosimilitud sea poco efectivo. Por lo tanto, es necesario utilizar una metodología alterna como la máxima verosimilitud restringida.

1.1.2.1. Máxima Verosimilitud Restringida (MVR)

La estimación mediante *MVR* fue propuesta por (Patterson & Thompson (1971)) y perfeccionado por (Harville (1977)). El método maximiza la verosimilitud sobre un conjunto de parámetros restringidos, si bien es aplicable a una gran variedad de modelos, se usa principalmente para estimar los componentes de varianza en un modelo lineal general con errores distribuidos de manera multinormal. El procedimiento se ajusta al hecho de que los efectos fijos son desconocidos al estimar los componentes de varianza. El proceso de estimación mediante máxima verosimilitud restringida remueve el efecto de las variables fijas al aplicar el proceso de máxima verosimilitud sobre los residuales obtenidos mediante mínimos cuadrados.

Los estimadores de *MVR* se obtienen al operar sobre una transformación lineal del vector respuesta. $\mathbf{Y}^* = A\mathbf{Y}$, de modo que la distribución de \mathbf{Y}^* no dependa de $\boldsymbol{\beta}$. Para

esta transformación, se puede considerar:

$$\mathbf{A} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$$

Así, \mathbf{Y}^* tendrá distribución multinormal con media cero sin importar los distintos valores de $\boldsymbol{\beta}$, de modo que las estimaciones de σ^2 y \mathbf{V}_0 no dependerán de la dimensión de $\boldsymbol{\beta}$.

La matriz usada para la transformación \mathbf{A} puede ser cualquier matriz de rango completo que cumpla con la condición que $E(\mathbf{Y}^*) = 0$.

En el modelo 1.1, se asume que la matriz de varianzas y covarianzas $\boldsymbol{\Sigma}$ depende de un conjunto de parámetros α , $\boldsymbol{\Sigma}(\alpha)$. Sea $\mathbf{A} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ y \mathbf{B} una matriz de tamaño $(nm \times (nm - p))$ tales que, $\mathbf{B}\mathbf{B}' = \mathbf{A}$ y $\mathbf{B}'\mathbf{B} = \mathbf{I}_{(nm-p)}$, sea $\mathbf{Z} = \mathbf{B}'\mathbf{Y}$, si el conjunto de parámetros α , es fijado, se obtiene:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{Y}$$

En (Diggle (2002)) se muestra que bajo las anteriores condiciones $E(\mathbf{Z}) = 0$ y $cov(\mathbf{Z}, \hat{\boldsymbol{\beta}}) = 0$, por lo tanto el supuesto de normalidad implica la independencia de \mathbf{Z} y $\hat{\boldsymbol{\beta}}$.

Entonces se obtiene que la función de log-verosimilitud restringida para α es:

$$L^*(\alpha) = \frac{1}{2}\log(|\boldsymbol{\Sigma}|) - \frac{1}{2}\log|\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}| - \frac{1}{2}(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})'\boldsymbol{\Sigma}^{-1}(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}) \quad (1.6)$$

De tal manera que, si los bloques de la matriz \mathbf{V}_0 son dados, se tiene:

$$\hat{\boldsymbol{\beta}}(\mathbf{V}_0) = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})'\mathbf{X}'\mathbf{V}^{-1}\mathbf{Y} \quad (1.7)$$

$$\hat{\sigma}^2(\mathbf{V}_0) = \frac{SCE(\mathbf{V}_0)}{nm - p} \quad (1.8)$$

Y la estimación de \mathbf{V}_0 resulta al maximizar la función:

$$L^*(\mathbf{V}_0) = -\frac{1}{2}m \{n\log(SCE(\mathbf{V}_0)) + \log|\mathbf{V}_0|\} - \frac{1}{2}\log(|\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}|)$$

Finalmente, al reemplazar los valores de la estimación de los bloques de la matriz de varianzas y covarianzas $\hat{\mathbf{V}}_0$ en los valores 1.7 y 1.8 se obtiene la estimación $\hat{\boldsymbol{\beta}}$ y $\hat{\sigma}^2$.

1.2. Inferencia Bayesiana

En algunos casos, cuando se realiza un estudio estadístico, se puede tener información previa de los parámetros. Para incorporar esta información en el desarrollo del análisis estadístico, se puede hacer uso de la función de densidad para el parámetro θ , $p(\theta)$. Esta distribución dependerá de un conjunto de parámetros θ' denominados hiperparámetros, los cuales se suponen conocidos. A la distribución $p(\theta)$ se le denomina distribución a priori, (Gelman, Carlin, Stern & Rubin (2014)).

Después de obtener los valores observados de la variable de estudio Y , se tienen dos fuentes de información sobre los parámetros: la función de verosimilitud $L(\theta|Y) = \prod f(Y_i|\theta)$ y la distribución a priori $p(\theta)$. Por lo tanto en el estudio estadístico la inferencia se puede basar en la distribución de θ después de observados los datos. Esta distribución se denomina distribución a posteriori, que se denota $\pi(\theta)$ y se obtiene mediante el teorema de Bayes así:

$$\pi(\theta) \propto L(\theta|Y)p(\theta)$$

Desde la perspectiva bayesiana, la inferencia sobre θ se basa en la distribución a posteriori $\pi(\theta)$.

Cuando la distribución a priori y la distribución a posteriori pertenecen a la misma familia de distribuciones se conocen como distribuciones a priori conjugadas. La distribución normal es una distribución a priori conjugada así como también lo son: la distribución gamma con la distribución muestral exponencial, la distribución gamma con la distribución muestral poisson o la distribución beta con distribución muestral beta, (Hoff (2009)).

Dado que no siempre es posible obtener información de los parámetros de manera analítica, en este documento se hace uso de simulación estocástica basada en cadenas de Markov (*MCMC*) para obtener muestras de la distribución a posteriori $\pi(\theta)$, (Gamerman & Lopes (2006)). Estos métodos darán una aproximación a la distribución a posteriori y deben ser utilizados únicamente, cuando no se pueda obtener dicha distribución de manera analítica.

1.2.1. Metropolis Hasting (MH)

El algoritmo de Metrópolis Hasting es una herramienta que usa cadenas de Markov para generar muestras de una distribución π , este algoritmo es detallado en (Gamerman & Lopes (2006), Gelman et al. (2014), Albert (2009)), entre otros. Para su desarrollo es necesario construir un núcleo de transición $p(\theta, \phi)$, de tal manera que π sea la distribución de equilibrio de la cadena. El caso más sencillo resulta cuando se cumple la condición de reversibilidad en la cadena es decir:

$$\pi(\theta)p(\theta, \phi) = \pi(\phi)p(\phi, \theta)$$

Esto se da para todo θ y ϕ . La condición anterior es suficiente para que π sea la distribución de equilibrio de la cadena. Si se calcula la integral en ambos lados de la igualdad, se obtiene:

$$\int \pi(\theta)p(\theta, \phi)d\theta = \pi(\phi) \quad \text{para todo } \phi$$

La construcción del núcleo de transición $p(\theta, \phi)$ se hace a partir de un núcleo arbitrario $q(\theta, \phi)$ que satisfaga que $\int q(\theta, \phi)d\phi = 1$ y de una probabilidad de aceptación $\alpha(\theta, \phi)$ tal que

$$p(\theta, \phi) = \begin{cases} q(\theta, \phi)\alpha(\theta, \phi) & \theta \neq \phi \\ 1 - \int q(\theta, \phi)\alpha(\theta, \phi) & \theta = \phi \end{cases}$$

El núcleo de transición $q(\theta, \phi)$ propone el movimiento de la cadena de Markov del punto θ al nuevo valor propuesto ϕ . La probabilidad de aceptación se define como:

$$\alpha(\theta, \phi) = \min \left\{ 1, \frac{\pi(\phi)q(\phi, \theta)}{\pi(\theta)q(\theta, \phi)} \right\}$$

El algoritmo para la generación de una muestra de una distribución π es el siguiente.

1. Iniciar el contador de la cadena, $j = 1$ y dar un valor inicial $\theta^{(0)}$ para el parámetro θ
2. Obtener ϕ desde el kernel de transición arbitrario $q(\theta^{j-1}, \cdot)$
3. Calcular la probabilidad de aceptación $\alpha(\theta, \phi)$, si se acepta $\theta^j = \phi$, en caso contrario $\theta^j = \theta^{j-1}$
4. Cambiar el contador $j = j + 1$ y retornar al paso 2 hasta encontrar convergencia.

La aceptación del nuevo valor se hace mediante el uso de una distribución uniforme en el intervalo $(0, 1)$ independiente de θ , se toma una muestra u^* de la distribución $u(0, 1)$. Si $u^* \leq \alpha$ se acepta el movimiento de la cadena, en caso contrario, es decir si $u^* > \alpha$, no se acepta el movimiento, Gamerman & Lopes (2006).

Los valores obtenidos por este método se pueden considerar una muestra de la distribución π después de que la cadena pase la fase transitoria y el efecto de los valores iniciales de la cadena sean pequeños.

Si el conjunto de parámetros θ es demasiado grande, este se puede particionar en bloques de dimensiones más pequeñas y de esta manera obtener la estimación para cada partición de manera sucesiva. En este escenario el algoritmo de Metrópolis Hasting tomará la siguiente forma.

1. Iniciar el contador de la cadena, $j = 1$ y dar un valor inicial $\theta^{(0)}$ para el parámetro θ
2. Inicie un contador de componentes i (para el bloque deseado o de manera aleatoria)
3. Obtener ϕ_i desde el núcleo de transición arbitrario $q(\theta^{j-1}, \cdot)$ para obtener un movimiento de la cadena en la i -ésima componente de θ
4. Calcular la probabilidad de aceptación, si se acepta $\theta_i^j = \phi_i$, en caso contrario $\theta_i^j = \theta_i^{j-1}$
5. Cambiar el contador $j = j + 1$ y retornar al paso 2 hasta encontrar convergencia.

La probabilidad de aceptación para el algoritmo de Metropolis-Hasting por bloques queda definida como

$$\alpha(\theta, \phi) = \min \left\{ 1, \frac{\pi_i(\phi_i)q(\phi_i, \theta_i)}{\pi_i(\theta_i)q(\theta_i, \phi_i)} \right\}$$

con $\pi_i(\theta_i) = \pi(\theta_i|\theta_{-i})$. Donde π_i se conoce como la distribución condicional completa, (Gelman et al. (2014)).

Para el estudio de la convergencia de la cadena, en (Gamerman & Lopes (2006)) se presentan algunas metodologías teóricas, donde se estudia la variación total de la distancia entre la iteración j de la cadena y el límite de la distribución π , o las metodologías que el autor denomina estadísticas. Estas tienen como fin estudiar el comportamiento de las salidas de la cadena, usualmente mediante el uso de gráficos estadísticos.

En (Gelman et al. (2014), Gamerman & Lopes (2006), Hoff (2009) y (Albert 2009)), se detallan casos particulares del algoritmo de Metrópolis -Hasting los cuales resultan al escoger algún núcleo de transición específico, entre los que se encuentran las cadenas simétricas, las caminatas aleatorias o las cadenas independientes.

Algunas Estructuras de Covarianza

Algunas ventajas de la modelación paramétrica de la matriz de varianza y covarianza en un conjunto de datos longitudinales mencionadas por (Weiss (2005) , Núñez-Antón & Zimmerman (2001)) son:

- a) Facilita y optimiza la obtención de las estimaciones de los parámetros de la estructura de medias asociada a los datos.
- b) Permite obtener de manera más adecuada los estimadores de los errores estándar de los parámetros usados para modelar la estructura de medias, (β) .
- c) En la mayoría de los casos, brinda una solución efectiva cuando se desean estimar modelos con datos faltantes o cuando los tiempos de medición no son iguales para todos los individuos del estudio.
- d) Se puede hacer uso de este tipo de estimaciones incluso cuando el número de mediciones en el tiempo sobre cada individuo es relativamente grande, comparado con el número de individuos.

Las estructuras específicas de la matriz de varianzas y covarianzas, considera que todas las varianzas y covarianzas intraindividuos son funciones de un vector de parámetros con una cantidad pequeña o moderada de elementos, α . El modelo de covarianza $\Sigma\alpha$ define una familia de posibles matrices dependientes de α . La estimación del vector de parámetros de la estructura de covarianza por lo general se realiza mediante máxima verosimilitud o máxima verosimilitud restringida (Diggle (2002)).

Sin embargo, uno de los principales problemas al modelar las estructuras de covarianza, para un conjunto de datos longitudinales, es la elección de la misma. Los métodos gráficos, análisis descriptivo y análisis de perfiles de las observaciones brindan la información necesaria para hacer una selección óptima de la estructura (Fitzmaurice (2008), Verbeke & Molenberghs (2009)).

En este capítulo, se presentan las estructuras de la matriz de varianzas y covarianzas $\Sigma(\alpha)$, que son utilizadas en el desarrollo de este trabajo. Para cada una de las estructuras, se dará la forma teórica de un bloque de la matriz y ejemplos numéricos que ilustren el comportamiento de la correlación o covarianza intraindividuos. Con el fin de ilustrar la

forma algebraica y dar ejemplos numéricos para cada estructura, se asumen cinco observaciones en el tiempo igualmente espaciadas sobre un individuo.

Además de las estructuras de matrices de varianzas y covarianzas que se presentan en este capítulo, otros tipos de estructuras se pueden consultar en (Weiss (2005), Núñez-Antón & Zimmerman (2001)).

2.1. Estructura de Simetría Compuesta (SC)

La estructura de la matriz de varianzas y covarianzas más simple que se presenta en la literatura es la de simetría compuesta, esta se define por los parámetros, σ^2 y ρ , es decir, para esta estructura $\alpha = (\sigma^2, \rho)$. El modelo de simetría compuesta se considera un modelo homogéneo, ya que la varianza sobre la diagonal de la matriz permanece constante y es definida como: $\sigma^2 = Var(Y_{ij})$. Mientras el parámetro ρ , representa la correlación entre dos observaciones cualesquiera sobre el mismo individuo, $\rho = Corr(Y_{ij}, Y_{il})$ para $j \neq l$. La principal característica de la estructura (SC) es que considera la covarianza constante para todos los instantes de tiempo .

La forma algebraica del bloque que representa la matriz de varianzas y covarianzas con estructura de simetría compuesta, considerando cinco observaciones en el tiempo, es:

$$Var(Y_i) = \sigma^2 \begin{pmatrix} 1 & \rho & \rho & \rho & \rho \\ \rho & 1 & \rho & \rho & \rho \\ \rho & \rho & 1 & \rho & \rho \\ \rho & \rho & \rho & 1 & \rho \\ \rho & \rho & \rho & \rho & 1 \end{pmatrix} \quad (2.1)$$

Esta estructura de matriz de varianzas y covarianzas indica que para todo $j \neq l$ la correlación entre Y_{ij} y Y_{il} es constante.

Si se supone un modelo con la estructura de simetría compuesta donde $\sigma^2 = 4$ y $\rho = 0.6$, con cinco observaciones en el tiempo, el bloque de la matriz para un individuo es:

$$Var(Y_i) = \begin{pmatrix} 4 & 2.4 & 2.4 & 2.4 & 2.4 \\ 2.4 & 4 & 2.4 & 2.4 & 2.4 \\ 2.4 & 2.4 & 4 & 2.4 & 2.4 \\ 2.4 & 2.4 & 2.4 & 4 & 2.4 \\ 2.4 & 2.4 & 2.4 & 2.4 & 4 \end{pmatrix} \quad (2.2)$$

La estructura (SC) puede ser encontrada en la literatura con el nombre de equicovarianza o equicorrelación. Esta última en algunos casos se denomina de esta manera para referirse a la versión heterogénea, (varianzas no constantes) del modelo de simetría compuesta (Núñez-Antón & Zimmerman (2001)).

2.2. Estructura Autoregresiva de orden 1 AR(1)

La estructura AR(1), para la matriz de varianzas y covarianzas, al igual que la de simetría compuesta, está conformada por dos parámetros $\alpha = (\sigma^2, \rho)$, en donde ρ es el parámetro de la correlación entre las mediciones de un individuo en n tiempos t_1, t_2, \dots, t_n , y donde la correlación entre las mediciones tomadas en el individuo i -ésimo en los tiempos j y l viene dada por $Cor(Y_{ij}, Y_{il}) = \rho^{|t_{ij}-t_{il}|}$. Por tal razón, este tipo de correlación serial difiere de la correlación tratada en la sección 2.1, ya que en el caso del modelo autoregresivo de orden 1 la covarianza decrece de manera exponencial a medida que avanza el tiempo.

En teoría, el parámetro de correlación ρ puede ser negativo cuando las observaciones son balanceadas, pero autores como (Weiss (2005), Verbeke & Molenberghs (2009) y Fitzmaurice (2008)), mencionan que en la práctica sobre datos longitudinales es poco común encontrar correlaciones negativas, por lo tanto, sugieren que se considere el parámetro de correlación positivo $\rho \in [0, 1]$. Si se tiene que $t = 1, 2, 3, 4, 5$ considerando mediciones balanceadas, la forma algebraica del bloque que representa la matriz de varianzas y covarianzas del i -ésimo individuos

$$Var(Y_i) = \sigma^2 \begin{pmatrix} 1 & \rho & \rho^2 & \rho^3 & \rho^4 \\ \rho & 1 & \rho & \rho^2 & \rho^3 \\ \rho^2 & \rho & 1 & \rho & \rho^2 \\ \rho^3 & \rho^2 & \rho & 1 & \rho \\ \rho^4 & \rho^3 & \rho^2 & \rho & 1 \end{pmatrix} \quad (2.3)$$

La diferencia de la estructura AR(1) y SC es difícil de observar en el análisis exploratorio o desde la gráfica de perfiles de los datos, sobre todo cuando se tienen pocas observaciones repetidas en el tiempo (Fitzmaurice (2008)). A pesar de esto, la diferencia numérica que existe se observa claramente. Si se supone como en la sección 2.1 que se tiene que $\sigma^2 = 4$ y $\rho = 0.6$, el bloque de la matriz para un individuo con cinco observaciones tomadas en el tiempo con estructura AR(1) es:

$$Var(Y_i) = \begin{pmatrix} 4 & 2.4 & 1.44 & 0.864 & 0.5184 \\ 2.4 & 4 & 2.4 & 1.44 & 0.864 \\ 1.44 & 2.4 & 4 & 2.4 & 1.44 \\ 0.864 & 1.44 & 2.44 & 2.4 & \\ 0.5184 & 0.864 & 1.44 & 2.4 & 4 \end{pmatrix} \quad (2.4)$$

Con esto se muestra la diferencia considerable que existe entre la estructura SC mostrada en 2.2 y la estructura AR(1) representada en 2.4 .

En (Núñez-Antón & Zimmerman (2001)) se puede encontrar una generalización de este tipo de las estructuras tratadas en esta sección y la sección 2.1 que resulta al considerar varianzas diferentes para cada instante de tiempo, es decir, si $\sigma_i^2 \neq \sigma_j^2$ para $i \neq j$. Este tipo de estructuras se denominan estructuras heterogéneas SCH y ARH(1). Los modelos con estructuras de varianza heterogéneas tendrán más de dos parámetros para estimar, de modo que genera modelos menos parsimoniosos.

2.3. Estructura Autoregresiva de Promedio Móvil de orden (1,1), ARMA(1,1)

La estructura $ARMA(1, 1)$ para la matriz de varianzas y covarianzas es una generalización de las estructuras mostradas en las secciones 2.1 y 2.2. El modelo $ARMA(1, 1)$ depende de tres parámetros $\alpha = (\sigma^2, \rho, \phi)$, donde ρ y ϕ son parámetros de correlación; el parámetro ϕ determina la correlación de las observaciones espaciadas por una unidad de tiempo, es decir, $\phi = Corr(Y_{ij}, Y_{ij-1})$, mientras que ρ es un parámetro adicional que modela una correlación que decrece de manera exponencial a medida que aumenta la distancia entre las observaciones. por lo cual, bajo el supuesto de un modelo $ARMA(1, 1)$ se define la correlación entre las mediciones tomadas en el individuo i -ésimo en los tiempos j y l por la expresión

$$Corr(Y_{ij}, Y_{il}) = \begin{cases} \phi & |t_{ij} - t_{il}| = 1 \\ \phi\rho^{|t_{ij} - t_{il}| - 1} & |t_{ij} - t_{il}| > 1 \end{cases}$$

con $\rho \in [0, 1]$ y $\phi \in [0, 1]$.

Si se tiene que $t = 1, 2, 3, 4, 5$ y se suponen observaciones balanceadas, la forma algebraica del bloque que representa la matriz de varianzas y covarianzas con estructura $ARMA(1, 1)$ para un individuo es:

$$Var(Y_i) = \sigma^2 \begin{pmatrix} 1 & \phi & \phi\rho & \phi\rho^2 & \phi\rho^3 \\ \phi & 1 & \phi & \phi\rho & \phi\rho^2 \\ \phi\rho & \phi & 1 & \phi & \phi\rho \\ \phi\rho^2 & \phi\rho & \phi & 1 & \phi \\ \phi\rho^3 & \phi\rho^2 & \phi\rho & \phi & 1 \end{pmatrix} \quad (2.5)$$

La generalización de las estructuras SC y $AR(1)$ que genera la estructura $ARMA(1, 1)$ se produce para valores puntuales de los parámetros de la siguiente manera:

- Si $\rho = 1$, se obtiene la estructura de simetría compuesta, SC .
- Si $\phi = \rho$, se obtiene la estructura autoregresiva de orden 1, $AR(1)$.
- Si $\rho = 0$, se obtiene la estructura de promedio móvil de orden 1, $MA(1)$.¹

Para observar un ejemplo numérico del modelo de covarianza $ARMA(1, 1)$, se consideran cinco mediciones balanceadas en el tiempo sobre cada individuo del estudio y se supone que $\sigma^2 = 1$, $\rho = 0.95$ y $\phi = 0.65$; se tiene que un bloque de la matriz de varianzas covarianzas es la matriz:

$$Var(Y_i) = \begin{pmatrix} 1 & 0.65 & 0.6175 & 0.587 & 0.557 \\ 0.65 & 1 & 0.65 & 0.6175 & 0.587 \\ 0.6175 & 0.65 & 1 & 0.65 & 0.6175 \\ 0.587 & 0.6175 & 0.65 & 1 & 0.65 \\ 0.557 & 0.587 & 0.6175 & 0.65 & 1 \end{pmatrix} \quad (2.6)$$

¹Aspectos teóricos y prácticos de esta estructura se pueden consultar en Weiss (2005)

En Núñez-Antón & Zimmerman (2001) se encuentra un modelo más general donde se dota la estructura de estacionariedad, especificando diferencias entre mediciones consecutivas.

2.4. Modelos de Antedependencia

El concepto de antedependencia fue introducido en Gabriel (1962). Mientras la estructura de antedependencia para matrices de varianzas y covarianzas en datos longitudinales, fue definida en Zimmerman & Núñez-Antón (2009).

Si se tienen las observaciones Y_1, \dots, Y_n y estas se distribuyen normal multivariadas; se denominan antedependientes de orden s si, Y_j e Y_{j+k+1} condicionadas a las observaciones intermedias $y_j \dots y_{j+k}$ son independientes, para todo $j = 1, \dots, n - k - 1$ y para todo $k \geq s$.

Una forma paramétrica para escribir el modelo longitudinal con estructura de antedependencia es:

$$\begin{aligned} Y_1 &= X_1' \beta + \epsilon_1 \\ Y_j &= X_j' \beta + \sum_{k=1}^{s^*} \rho_{jk} (y_{j-k} - X_{j-k}' \beta) + \epsilon_j \quad (j = 2, \dots, n) \end{aligned}$$

Donde $s^* = \min(s, j - 1)$, ϵ_j variables aleatorias normales con media cero y varianza $\sigma^2 > 0$, y ϕ_{jk} parámetros no estructurados.

En el modelo antedependiente de orden 1, si se tiene que $t = 1, 2, 3, 4, 5$ y considerando mediciones balanceadas, la forma algebraica del bloque que representa la matriz de varianzas y covarianzas del i -ésimo individuo, se representa mediante la matriz:

$$\text{Var}(Y_i) = \begin{pmatrix} \sigma_1^2 & \sigma_1 \sigma_2 \rho_1 & \sigma_1 \sigma_3 \rho_1 \rho_2 & \sigma_1 \sigma_4 \rho_1 \rho_2 \rho_3 & \sigma_1 \sigma_5 \rho_1 \rho_2 \rho_3 \rho_4 \\ & \sigma_2^2 & \sigma_2 \sigma_3 \rho_2 & \sigma_2 \sigma_4 \rho_2 \rho_3 & \sigma_2 \sigma_5 \rho_2 \rho_3 \rho_4 \\ & & \sigma_3^2 & \sigma_3 \sigma_4 \rho_3 & \sigma_3 \sigma_5 \rho_3 \rho_4 \\ & & & \sigma_4^2 & \sigma_4 \sigma_5 \rho_4 \\ & & & & \sigma_5^2 \end{pmatrix} \quad (2.7)$$

Donde ρ_j es la correlación entre las observaciones j y $j + 1$.

Se dice que el modelo es antedependiente no estructurado, si ρ_j y σ_i^2 no pueden ser expresados como funciones de un número menor de parámetros.

Los modelos de antedependencia de orden s $AD(s)$ pueden ser poco parsimoniosos, ya que, la dimensión de α para un modelo de estas características tiene $(s + 1)(2n - s)/2$ parámetros (Zimmerman & Núñez-antón (1997)).

2.4.1. Modelo de Antedependencia Estructurado ADS(1)

En busca de un modelo con un menor número de parámetros a estimar en la estructura de covarianzas, en Zimmerman & Núñez-antón (1997), se proponen modelos de antedependencia estructurados. En estos modelos los parámetros de correlación son determinados por

una función de potencias de box-cox y las varianzas de cada observación son determinadas por una función polinomial o a trozos.

La especificación de covarianzas, propuesta por Núñez-antón & Woodworth (1994), se define a partir de la estructura de covarianzas mostrada en (2.7), si se supone que:

$$\rho_j = \rho^{f(t_j, \lambda_k) - f(t_{j+k}, \lambda_k)} \quad (j = k + 1, \dots, n; k = 1, \dots, s) \quad (2.8)$$

$$\sigma_i^2 = \sigma^2 G(t_j, \psi) \quad (j = 1, \dots, n) \quad (2.9)$$

donde

$$f(t, \lambda) = \begin{cases} (t^\lambda - 1)/\lambda & \lambda \neq 0 \\ \log(t) & \lambda = 0 \end{cases} \quad (2.10)$$

donde

$$0 \leq \rho_k \leq 1 \quad \sigma_i^2 > 0 \quad \{\psi | G(t, \psi) > 0\}$$

De tal manera que la matriz de varianzas y covarianzas Σ sea definida positiva.

2.5. Sobre la Estimación de las Modelos de Covarianza SC, AR(1), ARMA(1,1), AD(1) y ADS(1)

Para realizar las estimaciones sobre estructuras de covarianza específicas, autores como Diggle (2002) y Verbeke & Molenberghs (2009), recomiendan el uso del método de máxima verosimilitud restringida el cuál fue descrito en la sección 1.1.2.

El método de máxima verosimilitud restringida suele presentar problemas en la estimación, cuando el diseño del modelo tiene datos faltantes o cuando la cantidad de mediciones en el tiempo son bastantes ya que la función de log-verosimilitud puede ser plana maximizarla por medio de métodos numéricos puede ser difícil Fitzmaurice (2008).

Dado que el desarrollo de este trabajo está enmarcado sobre diseños balanceados la estimación de máxima verosimilitud restringida es una propuesta óptima para el ajuste de modelos longitudinales dotados por las estructuras (SC), AR(1) y ARMA(1,1).

La estimación de las estructuras de antedependencia de primer orden se realiza fácilmente, ya que en Byrne & Arnold (1983) se presentan expresiones para los determinantes y las inversas de las matrices tridiagonales de la función 1.6. Estas son:

$$|\Sigma| = \prod_{i=1}^p \sigma_i^2 \prod_{i=1}^{p-1} (1 - \rho_i^2) \quad (2.11)$$

$$\Sigma^{-1} = \begin{cases} [\sigma_1^2(1 - \rho_1^2)]^{-1} & i = j = 1 \\ [\sigma_p^2(1 - \rho_{p-1}^2)]^{-1} & i = j = p \\ (1 - \rho_{i-1}^2 \rho_i^2) [\sigma_i^2(1 - \rho_{i-1}^2)(1 - \rho_{i-1}^2)]^{-1} & i = j \neq 1, p \\ -\rho_i [\sigma_i \sigma_j (1 - \rho_{i-1}^2)]^{-1} & |i - j| = 1 \\ 0 & |i - j| > 1 \end{cases} \quad (2.12)$$

En la práctica, estos modelos se pueden ajustar mediante un software estadístico, en el caso de las estructuras (*SC*), *AR*(1), *ARMA*(1,1) y *AD*(1) SPSS permite realizar el ajuste de modelos longitudinales que tengan este tipo de estructuras y permite seleccionar el método de estimación entre máxima verosimilitud o máxima verosimilitud restringida, mientras que mediante la herramienta PROC MIXED de SAS se pueden ajustar todas las estructuras presentadas en este capítulo.

Propuestas Bayesianas

La característica esencial de los métodos bayesianos es su uso explícito de la probabilidad de cuantificar incertidumbre en inferencias basadas en el análisis de datos estadísticos.

Los métodos Bayesianos proporcionan:

- estimaciones de parámetros con buenas propiedades estadísticas;
- descripciones parsimoniosas de datos observados;
- predicciones de datos faltantes y pronósticos de datos futuros;
- un marco computacional para la estimación, selección y validación de modelos. (Hoff (2009))

Una propuesta bayesiana, para el ajuste sobre modelos generalizados mixtos, se presenta en (Gamerman (1997)). Igualmente, en (Cepeda & Gamerman (2001)), se ajusta de manera conjunta la media y la estructura de varianza para un modelo de regresión lineal heteroscedástica, usando aprioris normales tanto para los parámetros de la media como los de la varianza. Una generalización de ese tipo de propuestas para la estimación conjunta de estructura de medias y varianzas es presentada en (Cepeda & Gamerman (2004)) y (Cuervo (2001)). Independientemente, en (Daniels & Pourahmadi (2002)), se desarrollan propuestas bayesianas para el ajuste de modelos dinámicos con estructura longitudinal, usando la descomposición de cholesky presentada en (Pourahmadi (1999)) y suponiendo aprioris normales para los parámetros de la media y la varianza. En (Cepeda-Cuervo & Núñez-Antón (2007) y Cepeda Cuervo (2011)), se proponen metodologías bayesianas para hacer el ajuste conjunto de la media y varianza en un modelo longitudinal, suponiendo un modelo de antedependencia no estructurado para la varianza. Este tipo de estimación se aplica tanto para modelos lineales como no lineales.

Siguiendo la metodología de Cuervo (2001), en este capítulo se presentan las propuestas bayesianas para realizar la estimación conjunta de medias y varianza para modelos longitudinales lineales con respuesta normal, suponiendo que la matriz de covarianzas sigue alguna de las estructuras presentadas en el capítulo (2).

El proceso de estimación se hace mediante muestreo de Gibbs y el algoritmo de Metropolis-Hasting por bloques (Gamerman & Lopes (2006)). Para los modelos de an-

tedependencia estructurados de orden 1 se hacen tres distinciones según el tipo de función que genere la varianza.

3.1. Propuesta Bayesiana para la Estructura de Media

Si se tiene el modelo de regresión longitudinal para el i -ésimo individuo:

$$Y_{it} = \mathbf{X}_i \boldsymbol{\beta} + \epsilon_t \quad (3.1)$$

donde Y_{it} es la medición realizada en el sujeto i -ésimo en el tiempo $t = 1, \dots, m$, y se supone que $\epsilon_t \sim N(0, \sigma^2 \mathbf{V}_0)$; además, se supone que se observa un vector de covariables asociadas \mathbf{x}_{it} donde la matriz $\mathbf{X}_i = (\mathbf{1}, \mathbf{x}_{i1}, \mathbf{x}_{i2}, \dots, \mathbf{x}_{ip})'$ de tamaño $m \times (p+1)$ es la matriz de covariables asociada al i -ésimo individuo, sea $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_p)'$ un vector de parámetros desconocidos. Si se denota $\mathbf{Y}_i = (\mathbf{Y}_{i1}, \mathbf{Y}_{i2}, \dots, \mathbf{Y}_{im})'$, $\mathbf{Y} = (\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_n)'$ el cual es el vector que contiene las $N = n \times m$ observaciones de la variable respuesta. $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_p)'$ y sea $\sigma^2 \mathbf{V} = \boldsymbol{\Sigma}(\boldsymbol{\alpha})$ una matriz diagonal por bloques con bloques no nulos $\sigma^2 \mathbf{V}_0$ de tamaño $m \times m$, donde cada bloque no nulo representa la matriz de varianzas y covarianzas para cada individuo, con $\boldsymbol{\alpha}$ un conjunto de parámetros dependiente del tipo de estructura de varianza la cual sea seleccionada para modelar los datos.

Para el modelo 3.1, dado que $\mathbf{Y} \sim N(\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma}(\boldsymbol{\alpha}, \mathbf{t}))$, la función de verosimilitud está dada por:

$$\mathbb{L}(\boldsymbol{\beta}, \boldsymbol{\alpha}) = \prod |\boldsymbol{\Sigma}|^{-\frac{1}{2}} (2\pi)^{N/2} \exp \left\{ -\frac{1}{2} (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})' \boldsymbol{\Sigma}^{-1} (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) \right\}$$

Reescribiendo esta expresión se tiene:

$$\mathbb{L}(\boldsymbol{\beta}, \boldsymbol{\alpha} | \mathbf{Y}) \propto |\boldsymbol{\Sigma}|^{-\frac{N}{2}} \exp \left\{ -\frac{1}{2} [\mathbf{Y}' \boldsymbol{\Sigma}^{-1} \mathbf{Y} - 2\boldsymbol{\beta}' \mathbf{X}' \boldsymbol{\Sigma}^{-1} \mathbf{Y} + \boldsymbol{\beta}' \mathbf{X}' \boldsymbol{\Sigma}^{-1} \mathbf{X} \boldsymbol{\beta}] \right\} \quad (3.2)$$

Si se considera que $\boldsymbol{\Theta} = (\boldsymbol{\beta}, \sigma^2, \boldsymbol{\alpha})$, dado que se desea realizar una propuesta bayesiana, es necesario definir una distribución a priori $p(\boldsymbol{\Theta})$. Si se supone independencia entre $\boldsymbol{\beta}$, σ^2 y $\boldsymbol{\alpha}$, entonces la distribución a posteriori $\pi(\boldsymbol{\Theta}) = \mathbb{L}(\boldsymbol{\beta}, \boldsymbol{\alpha} | \mathbf{Y}) p(\boldsymbol{\beta}) p(\sigma^2) p(\boldsymbol{\alpha})$. Por lo tanto, si se suponen las ditribuciones a priori:

$$p(\boldsymbol{\beta}) \sim N(\mathbf{b}_0, \mathbf{B}_0) \quad (3.3)$$

$$p\left(\frac{1}{\sigma^2}\right) = p(\gamma) \sim \Gamma\left(\frac{g_0}{2}, \frac{g_0 \sigma_0^2}{2}\right) \quad (3.4)$$

Donde la distribución apriori $p(\boldsymbol{\alpha})$ se definirá según la estructura de la matriz de varianzas y covarianzas. De 3.4, se deduce que la distribución $p(\boldsymbol{\beta})$ es:

$$p(\boldsymbol{\beta}) \propto \exp \left\{ -\frac{1}{2} (\boldsymbol{\beta} - \mathbf{b}_0)' \mathbf{B}_0^{-1} (\boldsymbol{\beta} - \mathbf{b}_0) \right\}, \quad (3.5)$$

Entonces, por medio del teorema de Bayes, se concluye que la distribución condicional completa a posteriori $\pi(\boldsymbol{\beta} | \boldsymbol{\alpha}, \mathbf{Y})$ está dada:

$$\begin{aligned}\pi_{\beta} &\propto \exp \left\{ -\frac{1}{2} \left[-2\beta' \mathbf{X}' \boldsymbol{\Sigma}^{-1} \mathbf{Y} + \beta' \mathbf{X}' \boldsymbol{\Sigma}^{-1} \mathbf{X} \beta - 2\beta' \mathbf{B}_0^{-1} \mathbf{b}_0 + \beta' \mathbf{B}_0 \beta \right] \right\} \\ &= \exp \left\{ -\frac{1}{2} \left[\beta' (\mathbf{X}' \boldsymbol{\Sigma}^{-1} \mathbf{X} + \mathbf{B}_0^{-1}) \beta - 2\beta' (\mathbf{X}' \boldsymbol{\Sigma}^{-1} \mathbf{Y} + \mathbf{B}_0^{-1} \mathbf{b}_0) \right] \right\}.\end{aligned}\quad (3.6)$$

En consecuencia:

$$\pi_{\beta} \sim N(\mathbf{b}^*, \mathbf{B}^*) \quad (3.7)$$

Donde $\mathbf{b}^* = \mathbf{B}^* (\mathbf{X}' \boldsymbol{\Sigma}^{-1} \mathbf{Y} + \mathbf{B}_0^{-1} \mathbf{b}_0)$ y matriz de varianza $\mathbf{B}^* = (\mathbf{X}' \boldsymbol{\Sigma}^{-1} \mathbf{X} + \mathbf{B}_0^{-1})^{-1}$ (Cepeda & Gamerman (2004)).

Por lo tanto, para obtener los valores estimados de $\boldsymbol{\beta}$ mediante muestreo de Gibbs, se tomarán muestras de la distribución dada en 3.7 con probabilidad de aceptación $r = 1$ (Gamerman & Lopes (2006), Gelman et al. (2014)).

3.2. Propuesta Bayesiana para la Estimación de Modelos Longitudinales con Estructura de Covarianza SC y AR(1)

Las estructuras de la matriz de covarianza $\boldsymbol{\Sigma}(\alpha)$, presentadas en 2.1 y 2.3, dependen de dos parámetros $\boldsymbol{\alpha} = (\sigma^2, \rho)$. Así, para el ajuste de los modelos mediante métodos bayesianos es necesario establecer las distribuciones a posteriori condicionales completas para σ^2 y ρ .

Para calcular una distribución condicional completa para σ^2 se debe notar que $\boldsymbol{\Sigma} = \sigma^2 \mathbf{C}$ y $\boldsymbol{\Sigma}^{-1} = \frac{\mathbf{C}^{-1}}{\sigma^2}$, donde \mathbf{C} es una matriz diagonal por bloques que depende únicamente del parámetro de correlación ρ . Dado que se supone una distribución a priori para $\frac{1}{\sigma^2} = \gamma$ en 3.4 se tiene:

$$P(\gamma) = \frac{1}{\Gamma(\frac{g_0}{2})} \left(\frac{g_0 \sigma_0^2}{2} \right)^{\frac{g_0}{2}} \gamma^{\left(\frac{g_0}{2}-1\right)} \exp\left(-\frac{g_0 \sigma_0^2}{2} \gamma\right) I_{(0,\infty)} \quad (3.8)$$

Sea $R = (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})' \mathbf{C}^{-1} (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})$. De tal modo, la verosimilitud del modelo se puede escribir:

$$\mathbb{L}(\mathbf{Y} | \boldsymbol{\beta}, \sigma^2, \rho) \propto \gamma^{\frac{N}{2}} \exp(\gamma(-\frac{1}{2}R))$$

Y por lo tanto, la distribución a posteriori condicional completa de γ es:

$$\begin{aligned}\pi_{\gamma} &\propto \gamma^{\frac{N}{2}} \exp(\gamma(-\frac{1}{2}R)) \gamma^{\left(\frac{g_0}{2}-1\right)} \exp\left(-\frac{g_0 \sigma_0^2}{2} \gamma\right) \\ &\propto \gamma^{\left(\frac{N}{2} + \frac{g_0}{2} - 1\right)} \exp\left(-\gamma \frac{g_0 \sigma_0^2 + R}{2}\right) \\ &\propto \gamma^{\left(\frac{N+g_0}{2} - 1\right)} \exp\left(-\gamma \frac{g_0 \sigma_0^2 + R}{2}\right)\end{aligned}\quad (3.9)$$

De este modo, los valores de γ se pueden obtener de la distribución $\Gamma(\frac{N+g_0}{2}, \frac{g_0 \sigma_0^2 + R}{2})$. Debido a que $\gamma = 1/\sigma^2$ por propiedades de la distribución gamma, se tiene que los valores

de σ^2 pueden ser obtenidos directamente mediante muestreo de Gibbs de una distribución gamma inversa.

$$\pi(\sigma^2) \sim I\Gamma(c^*, d^*) \quad (3.10)$$

Donde $c^* = \frac{N+g_0}{2}$ y $d^* = \frac{g_0\sigma_0^2+R}{2}$ (Gelman et al. (2014), Hoff (2009), Albert (2009)).

Para construir una distribución a posteriori condicional completa, para el parámetro de correlación ρ , dado que se sugiere que la correlación sea positiva, en (Hoff (2009)) se propone una distribución apriori $p(\rho) \sim U(0, 1)$, en busca de una propuesta más flexible y con el fin de dar la oportunidad de utilizar una distribución apriori informativa en caso de que el investigador tenga algún conocimiento del comportamiento de la correlación, en este documento se supone una distribución apriori para el parámetro de correlación ρ .

$$p(\rho) \sim Beta(a, b) \quad (3.11)$$

Si no se tiene información del comportamiento de ρ , se puede hacer uso de la distribución $Beta(1, 1)$, la cuál coincide con la distribución uniforme continua en el intervalo $(0, 1)$.

Si $\rho \sim Beta(a, b)$, su función de densidad de probabilidad es:

$$p(\rho) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \rho^{a-1} (1-\rho)^{b-1} I_{(0,1)}$$

Por lo tanto, la distribución a posteriori condicional completa para ρ está dada por

$$\begin{aligned} \pi_\rho &\propto \exp \left\{ -\frac{1}{2} (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})' \boldsymbol{\Sigma}(\boldsymbol{\alpha})^{-1} (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) \right\} \rho^{a-1} (1-\rho)^{b-1} \\ &\propto \exp \left\{ -\frac{1}{2} (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})' \boldsymbol{\Sigma}(\boldsymbol{\alpha})^{-1} (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) + (a-1)\ln(\rho) + (b-1)\ln(1-\rho) \right\} \end{aligned} \quad (3.12)$$

Esta función no tiene forma de una distribución conocida. De modo que, para obtener muestras de esta, es necesario proponer un kernel de transición $q(\theta, \phi)$ para hacer uso de simulación de montecarlo mediante cadenas de markov (MCMC) (Gamerman & Lopes (2006)).

Algunos autores, como (Hoff (2009)) proponen una caminata aleatoria uniforme como Kernel de transición:

$$q(\rho^{(*)} | \rho^{(k)}) = \rho^{(*)} \sim U(\rho^{(k)} - \delta, \rho^{(k)} + \delta),$$

Este tipo de kernel muestrea valores fuera del intervalo $(0, 1)$. Este inconveniente se resuelve con transformaciones sobre los valores muestreados, por ejemplo $\rho^{(*)} = |\rho^{(*)}|$.

Para evitar muestrear valores fuera del intervalo de interés y no tener la necesidad de usar transformaciones sobre los valores muestreados, se propone como kernel de transición la caminata aleatoria:

$$q(\rho^{(*)} | \rho^{(k)}) = \begin{cases} \rho^{(*)} \sim U(0, 2\rho^{(k)}) & \rho^{(k)} \leq 0.5 \\ \rho^{(*)} \sim U(2\rho^{(k)} - 1, 1) & \rho^{(k)} > 0.5 \end{cases} \quad (3.13)$$

Con estas condiciones, para encontrar las estimaciones conjuntas de los parámetros de media y varianza para modelos longitudinales con estructura de covarianza SC o $AR(1)$, el algoritmo es el siguiente:

1. Dar valores iniciales $(\beta_0, \sigma_0^2, \rho_0)$
2. Mediante muestreo de Gibbs obtener $\beta^{(k+1)}$ a partir de la distribución normal multivariada 3.7
3. Mediante muestreo de Gibbs obtener $\sigma^{2(k+1)}$ a partir de la distribución Gamma Inversa 3.10
4. Muestrear $q(\rho^{(*)}|\rho^{(k)})$ desde el kernel de transición 3.13
5. Se calcula el radio de aceptación

$$r = \min \left(1, \frac{P(\mathbf{Y}|\beta^{(k+1)}, \sigma^{2(k+1)}, \rho^{(*)})P(\rho^{(*)})}{P(\mathbf{Y}|\beta^{(k+1)}, \sigma^{2(k+1)}, \rho^{(k)})P(\rho^{(k)})} \right)$$

6. Se muestrea $a \sim U(0, 1)$ si $a < r$ entonces $\rho^{(k+1)} = \rho^{(*)}$, si $r \geq a$ $\rho^{(k+1)} = \rho^{(k)}$
7. Retornar al paso 2 hasta convergencia.

3.3. Propuesta Bayesiana para la Estimación de Modelos Longitudinales con Estructura de Covarianza ARMA(1,1)

Si se tiene un modelo longitudinal como 3.1 con estructura de matriz de varianzas y covarianzas como la presentada en 2.5, $\Sigma(\alpha)$ depende de tres parámetros, $\alpha = (\sigma^2, \rho, \phi)'$. Al ser ϕ un parámetro de correlación con las mismas características de ρ , si se suponen distribuciones a priori $Beta(a_1, b_1)$ y $Beta(a_2, b_2)$ para ρ y ϕ respectivamente, se obtienen las siguientes distribuciones condicionales completas para las componentes de α

$$\begin{aligned} \pi_{\sigma^2} &\propto \gamma^{\left(\frac{N+g_0}{2}-1\right)} \exp\left(-\gamma \frac{g_0 \sigma_0^2 + R}{2}\right) \\ \pi_{\rho} &\propto \exp\left\{-\frac{1}{2}(Y - X\beta)' \Sigma^{-1}(Y - X\beta) + (a_1 - 1)\ln(\rho) + (b_1 - 1)\ln(1 - \rho)\right\} \\ \pi_{\phi} &\propto \exp\left\{-\frac{1}{2}(Y - X\beta)' \Sigma^{-1}(Y - X\beta) + (a_2 - 1)\ln(\phi) + (b_2 - 1)\ln(1 - \phi)\right\} \end{aligned} \quad (3.14)$$

Donde, para ϕ , se propone el kernel de transición:

$$q(\phi^{(*)}|\phi^{(k)}) = \begin{cases} \phi^{(*)} \sim U(0, 2\phi^{(k)}) & \phi^{(k)} \leq 0.5 \\ \phi^{(*)} \sim U(2\phi^{(k)} - 1, 1) & \phi^{(k)} > 0.5 \end{cases} \quad (3.15)$$

Siguiendo la metodología utilizada en ??, si se supone la estructura $ARMA(1, 1)$, el algoritmo para la estimación conjunta de media y matriz de varianzas y covarianzas es:

1. Dar valores iniciales $(\beta_0, \sigma_0^2, \rho_0, \phi_0)$
2. Mediante muestreo de Gibbs obtener $\beta^{(k+1)}$ a partir de la distribución normal multivariada 3.7
3. Mediante muestreo de Gibbs obtener $\sigma^{2(k+1)}$ a partir de la distribución Gamma Inversa 3.10
4. Muestrear $q(\rho^{(*)}|\rho^{(k)})$ desde el kernel de transición (3.13)
5. Se calcula el radio de aceptación ρ

$$r_1 = \min \left(1, \frac{P(\mathbf{Y}|P(\beta^{(k+1)}, \sigma^{2(k+1)}, \rho^{(*)}, \phi^{(k)}))P(\rho^{(*)})}{P(P(\mathbf{Y}|P(\beta^{(k+1)}, \sigma^{2(k+1)}, \rho^{(k)}, \phi^{(k)}))P(\rho^{(k)})} \right)$$

6. Se muestrea $a_1 \sim U(0, 1)$ si $a_1 < r_1$ entonces $\rho^{(k+1)} = \rho^{(*)}$, si $r_1 \geq a_1$ $\rho^{(k+1)} = \rho^{(k)}$
7. Muestrear $q(\phi^{(*)}|\phi^{(k)})$ desde el kernel de transición (3.15)
8. Se calcula el radio de aceptación para ϕ

$$r_2 = \min \left(1, \frac{P(\mathbf{Y}|\beta^{(k+1)}, \sigma^{2(k+1)}, \rho^{(k+1)}, \phi^{(*)})P(\phi^{(*)})}{P(\mathbf{Y}|\beta^{(k+1)}, \sigma^{2(k+1)}, \rho^{(k+1)}, \phi^{(k)})P(\phi^{(k)})} \right)$$

9. Se muestrea $a_2 \sim U(0, 1)$ si $a_2 < r_2$ entonces $\phi^{(k+1)} = \phi^{(*)}$, si $r_2 \geq a_2$ $\phi^{(k+1)} = \phi^{(k)}$
10. Retornar al paso 2 hasta convergencia.

Al igual que en la sección anterior, al usar como kernel de transición una caminata aleatoria, los dos cocientes de aceptación calculados en este algoritmo se reducen al cociente de los productos de la verosimilitud y las distribuciones a priori (Gamerman & Lopes (2006), Gelman et al. (2014)).

3.4. Propuesta Bayesiana para la Estimación de Modelos Longitudinales con Estructura de Covarianza ADS(1)

Dado que la cantidad de parámetros de la matriz de varianza en los modelos de antedependencia estructurados de orden 1 dependen del tipo de funciones $G(\Psi, t)$ que se consideren en 2.9, es necesario realizar una propuesta bayesiana diferente para cada tipo de función que se proponga. En el desarrollo de este trabajo, se tienen en cuenta tres tipos de funciones posibles para las cuales se describe por separado las distribuciones completas que se emplean y el algoritmo de estimación .

3.4.1. Modelo Antedependiente Estructurado Tipo 1

En primer lugar, si se supone que la matriz de varianzas covarianzas está dada por 2.7, con función de potencias definida como en 2.10 y $G(t, \psi) = 1$, para un modelo de regresión longitudinal como el presentado en 3.1. Si, además, se suponen cuatro observaciones en el

tiempo sobre cada individuo, el bloque de la matriz de varianzas covarianzas para este modelo, tiene la forma:

$$\text{Var}(Y_i) = \sigma^2 \begin{pmatrix} 1 & \rho_1 & \rho_1\rho_2 & \rho_1\rho_2\rho_3 \\ & 1 & \rho_2 & \rho_2\rho_3 \\ & & 1 & \rho_3 \\ & & & 1 \end{pmatrix}$$

$$\rho_j = \rho^{f(t_j, \lambda) - f(t_{j+1}, \lambda)} \quad (j = 1, \dots, n)$$

Los parámetros a estimar en la matriz de covarianzas son $(\sigma^2, \rho, \lambda)$. Para estimar σ^2 , se tomán muestras directamente de una distribución a posteriori condicional completa como la presentada en 3.10; y para tomar muestras de ρ se hace uso de MCMC tomando un kernel de transición como el presentado en 3.13.

Con el fin de encontrar una distribución a posteriori condicional completa para λ , se supone como distribución a priori $p(\lambda) \sim U(-a, a)$; por lo tanto, la distribución condicional completa para λ viene dada por la expresión:

$$\pi_\lambda \propto \exp \left\{ -\frac{1}{2}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})' \boldsymbol{\Sigma}_\lambda^{-1} (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) \right\} \frac{1}{2a} \quad (3.16)$$

$$= \exp \left\{ -\frac{1}{4a}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})' \boldsymbol{\Sigma}_\lambda^{-1} (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) \right\} \quad (3.17)$$

Esta distribución no tiene una forma de distribución conocida para λ , por lo cual no se puede hacer uso del muestreo de Gibbs y es necesario proponer un kernel de transición. Se propone como kernel de transición la caminata aleatoria:

$$q(\lambda^{(*)} | \lambda^{(k)}) = \lambda^{(*)} \sim U(\lambda^{(k)} - \delta, \lambda^{(k)} + \delta) \quad (3.18)$$

Por lo tanto, el algoritmo para realizar la estimación conjunta de media y estructura de matriz de varianzas y covarianzas en este modelo es:

1. Dar valores iniciales $(\boldsymbol{\beta}_0, \sigma_0^2, \rho_0, \phi_0)$
2. Mediante muestreo de Gibbs obtener $\boldsymbol{\beta}^{(k+1)}$ a partir de la distribución normal multivariada 3.7
3. Mediante muestreo de Gibbs obtener $\sigma^{2(k+1)}$ a partir de la distribución Gamma Inversa 3.10
4. Muestrear $\rho^{(*)}$ desde el kernel de transición 3.13
5. Se calcula el radio de aceptación ρ

$$r_1 = \min \left(1, \frac{P(\mathbf{Y} | P(\boldsymbol{\beta}^{(k+1)}, \sigma^{2(k+1)}, \rho^{(*)}, \phi(k))) P(\rho^{(*)})}{P(P(\mathbf{Y} | P(\boldsymbol{\beta}^{(k+1)}, \sigma^{2(k+1)}, \rho^{(k)}, \phi(k))) P(\rho^{(k)})} \right)$$

6. Se muestrea $a_1 \sim U(0, 1)$ si $a_1 < r_1$ entonces $\rho^{(k+1)} = \rho^{(*)}$, si $r_1 \geq a_1$ $\rho^{(k+1)} = \rho^{(k)}$
7. Muestrear $\lambda^{(*)}$ desde el kernel de transición 3.18

8. Se calcula el radio de aceptación

$$r_2 = \min \left(1, \frac{P(\mathbf{Y}|\boldsymbol{\beta}^{(k+1)}, \sigma^{2(k+1)}, \rho^{(k+1)}, \lambda^{(*)})P(\lambda^{(*)})}{P(\mathbf{Y}|\boldsymbol{\beta}^{(k+1)}, \sigma^{2(k+1)}, \rho^{(k+1)}, \lambda^{(k)})P(\lambda^{(k)})} \right)$$

9. Se muestrea $a_2 \sim U(0, 1)$ si $a_2 < r_2$ entonces $\lambda^{(k+1)} = \lambda^{(*)}$, si $r_2 \geq a_2$ $\lambda^{(k+1)} = \lambda^{(k)}$

10. Retornar al paso 2 hasta convergencia.

3.4.2. Modelo Antependiente Estructurado Tipo 2

Si en el modelo 3.1, se supone una matriz de varianzas covarianzas dada por 2.7, con función de potencias como la presentada en 2.10 y se define:

$$G(t, \psi) = \begin{cases} 1 & t = 1 \\ \psi & \text{En otro caso.} \end{cases}$$

Al suponer 4 observaciones en el tiempo, el bloque de la matriz de varianzas covarianzas para este modelo tiene como estructura:

$$\text{Var}(Y_i) = \sigma^2 \begin{pmatrix} 1 & \sqrt{\psi}\rho_1 & \sqrt{\psi}\rho_1\rho_2 & \sqrt{\psi}\rho_1\rho_2\rho_3 \\ & 1 & \psi\rho_2 & \sqrt{\psi}\rho_2\rho_3 \\ & & 1 & \psi\rho_3 \\ & & & 1 \end{pmatrix}$$

Los parámetros a estimar para la matriz de varianzas y covarianzas en este modelo son $\boldsymbol{\alpha} = (\sigma^2, \rho, \lambda, \psi)$. Para la estimación bayesiana tanto de σ^2 , ρ y λ se hace uso de la misma estrategia presentada en la sección 3.4.1; mientras, para ψ se supone que $\psi = \exp(\Delta)$; y se propone como distribución a priori para Δ , una distribución normal $p(\Delta) \sim N(0, \sigma^2)$. De este modo la distribución posteriori condicional completa $\pi(\Delta)$ no tiene forma de distribución conocida, por lo tanto se propone el kernel de transición para ψ

$$\begin{aligned} q(\psi^{(*)}, \psi^{(k)}) &= \psi^{(*)} + N(0, \sigma^2) = \Delta^{(k)} \\ \psi^{(k)} &= \exp(\Delta^{(k)}) \end{aligned} \quad (3.19)$$

Después del paso (9) del algoritmo mostrado en la sección 3.4.1, se agregan los siguientes pasos:

10 Muestrear $\psi^{(k)}$ tal como en 3.19

11 Se calcula el radio de aceptación

$$r_3 = \min \left(1, \frac{P(\mathbf{Y}|\boldsymbol{\beta}^{(k+1)}, \sigma^{2(k+1)}, \rho^{(k+1)}, \lambda^{(k+1)}, \psi^{(*)})P(\psi^{(*)})}{P(\mathbf{Y}|\boldsymbol{\beta}^{(k+1)}, \sigma^{2(k+1)}, \rho^{(k+1)}, \lambda^{(k+1)}, \psi^{(k)})P(\psi^{(k)})} \right)$$

12 Se muestrea $a_3 \sim U(0, 1)$ si $a_3 < r_3$ entonces $\psi^{(k+1)} = \psi^{(*)}$, si $r_3 \geq a_3$ $\psi^{(k+1)} = \psi^{(k)}$

3.4.3. Modelo Antedependiente Estructurado Tipo 3

En Zimmerman et al. (1998) se propone el ajuste de un modelo como 3.1, considerando una estructura de covarianzas $ADS(1)$ determinada por la función de potencias 2.10 y se define:

$$G(t, \psi) = 1 + \psi_1 t + \psi_2 t^2 \quad (3.20)$$

Donde los parámetros ψ_1 y ψ_2 son seleccionados de modo que $G(t, \psi)$ sea una función positiva.

Con el fin de obtener una propuesta no restrictiva sobre los parámetros, se supone:

$$\sigma_i^2 = \exp(\psi_0 + \psi_1 t_i + \psi_2 t_i^2) \quad (3.21)$$

De tal manera, los parámetros a estimar en la estructura de la matriz de varianzas y covarianzas son $(\rho, \lambda, \psi_0, \psi_1, \psi_2)$. Así, la propuesta bayesiana sobre ρ y λ se realizan como en 3.4.2; mientras que para la estimación de (ψ_0, ψ_1, ψ_2) se hace uso de las varianzas muestrales de cada una de las secciones de la carrera s_i^2 con $i = 1, \dots, 10$ para construir la variable de trabajo:

$$\widetilde{w}_i = \log(s_i^2) = \psi_0 + \psi_1 t_i + \psi_2 t_i^2 + \widetilde{\epsilon}_i \quad (3.22)$$

Si se supone que en 3.22 $\widetilde{\epsilon}_i \sim N(0, \widetilde{\sigma}^2)$ con $\widetilde{\sigma}^2$ constante y conocida, se tiene un modelo de regresión cuadrática en el cual se deben estimar los parámetros (ψ_0, ψ_1, ψ_2) .

Sea $\mathbf{W}=(w_1, w_2, \dots, w_t)$, $\mathbf{\Psi}=(\psi_0, \psi_1, \psi_2)$ y la matriz $\mathbf{X}_t=(\mathbf{1}, x_{1t}, x_{2t})$ con $i = 1, 2, \dots, n$, de donde se obtiene la función de verosimilitud:

$$\mathbb{L}(\mathbf{\Psi}|\mathbf{W}) \propto \exp \left\{ -\frac{1}{2} (\mathbf{W}'\mathbf{W} - 2\mathbf{\Psi}'\mathbf{X}_t'\mathbf{W} + \mathbf{\Psi}'\mathbf{X}_t'\mathbf{X}_t\mathbf{\Psi}) \right\} \quad (3.23)$$

Si se supone una distribución a priori para $\mathbf{\Psi} \sim N(\mathbf{\Psi}_0, \mathbf{K}_0)$, entonces:

$$p(\mathbf{\Psi}) \propto \exp \left\{ -\frac{1}{2} (\mathbf{\Psi}'\mathbf{K}^{-1}\mathbf{\Psi} - 2\mathbf{\Psi}'\mathbf{K}^{-1}\mathbf{\Psi}_0) \right\} \quad (3.24)$$

De modo que, al multiplicar las ecuaciones 3.23 y 3.24 se obtiene la distribución a posteriori:

$$\pi_{\mathbf{\Psi}} \propto \exp \left\{ -\frac{1}{2} \mathbf{\Psi}'(\mathbf{X}_t'\mathbf{\Sigma}_{\mathbf{\Psi}}^{-1}\mathbf{X}_t + \mathbf{K}_0^{-1})\mathbf{\Psi} - 2\mathbf{\Psi}'(\mathbf{X}_t'\mathbf{\Sigma}_{\mathbf{\Psi}}^{-1}\mathbf{W} + \mathbf{K}_0^{-1}\mathbf{\Psi}_0) \right\}$$

con $\mathbf{\Sigma}_{\mathbf{\Psi}}^{-1} = \text{diag}(\widetilde{\sigma}^2) = \text{diag}(\text{var}(w_i))$

$$\pi_{\mathbf{\Psi}} \sim N(\mu_{\mathbf{\Psi}}, \mathbf{K}_{\mathbf{\Psi}}) \quad (3.25)$$

Donde $\mu_{\mathbf{\Psi}} = \mathbf{K}_{\mathbf{\Psi}}(\mathbf{X}_t'\mathbf{\Sigma}_{\mathbf{\Psi}}^{-1}\mathbf{W} + \mathbf{K}_0^{-1}\mathbf{\Psi}_0)$ y $\mathbf{K}_{\mathbf{\Psi}} = (\mathbf{X}_t'\mathbf{\Sigma}_{\mathbf{\Psi}}^{-1}\mathbf{X}_t + \mathbf{K}_0^{-1})^{-1}$ Cuervo (2001).

Con lo anterior para la estimación conjunta de media y estructura de covarianzas en este modelo se da el algoritmo:

1. Se dan valores iniciales $(\beta_0, \Psi_0, \rho_0, \lambda_0)$
2. Mediante muestreo de Gibbs obtener $\beta^{(k+1)}$ a partir de la distribución normal multivariada 3.7
3. Se obtiene los valores de $\Psi^{(*)}$ desde la ecuación (3.25)
4. Se calcula el radio de aceptación

$$r_1 = \min \left(1, \frac{P(\mathbf{Y}|\beta^{(k+1)}, \Psi^{(*)}, \rho^{(k)}, \lambda^{(k)})P(\Psi^{(*)})}{P(\mathbf{Y}|\beta^{(k+1)}, \Psi^{(k)}, \rho^{(k)}, \lambda^{(k)})P(\Psi^{(k)})} \right)$$

5. Se muestrea $a_1 \sim U(0, 1)$ si $a_1 < r_1$ entonces $\Psi^{(k+1)} = \Psi^{(*)}$, si $r_2 \geq a_2$ $\Psi^{(k+1)} = \Psi^{(k)}$
6. Muestrear $\rho^{(*)}$ desde el kernel de transición (3.13)
7. Se calcula el radio de aceptación ρ

$$r_2 = \min \left(1, \frac{P(\mathbf{Y}|P(\beta^{(k+1)}, \sigma^{2(k+1)}, \rho^{(*)}, \phi^{(k)})P(\rho^{(*)})}{P(P(\mathbf{Y}|P(\beta^{(k+1)}, \sigma^{2(k+1)}, \rho^{(k)}, \phi^{(k)})P(\rho^{(k)})} \right)$$

8. Se muestrea $a_2 \sim U(0, 1)$ si $a_2 < r_2$ entonces $\rho^{(k+1)} = \rho^{(*)}$, si $r_2 \geq a_2$ $\rho^{(k+1)} = \rho^{(k)}$
9. Muestrear $\lambda^{(*)}$ desde el kernel de transición (3.18)
10. Se calcula el radio de aceptación

$$r_3 = \min \left(1, \frac{P(\mathbf{Y}|\beta^{(k+1)}, \sigma^{2(k+1)}, \rho^{(k+1)}, \lambda^{(*)})P(\lambda^{(*)})}{P(\mathbf{Y}|\beta^{(k+1)}, \sigma^{2(k+1)}, \rho^{(k+1)}, \lambda^{(k)})P(\lambda^{(k)})} \right)$$

11. Se muestrea $a_3 \sim U(0, 1)$ si $a_3 < r_3$ entonces $\lambda^{(k+1)} = \lambda^{(*)}$, si $r_3 \geq a_3$ $\lambda^{(k+1)} = \lambda^{(k)}$
12. Retornar al paso 2 hasta convergencia.

Simulaciones y Aplicaciones

En este capítulo se presentan simulaciones y aplicaciones para algunos conjuntos de datos, usando las propuestas bayesianas presentadas en el capítulo anterior. Además se presentan simulaciones para los modelos $AR(1)$, SC y $ARMA(1, 1)$, implementando distribuciones apriori informativas y no informativas. Las aplicaciones se realizan para el *Small Mice Data* considerando estructuras SC , $AR(1)$ y $ARMA(1, 1)$. Los resultados de la estimación bayesiana para estas estructuras son comparadas por las presentadas en Weiss (2005) donde se utilizó la metodología clásica (estimación via MVR).

Para los modelos $ADS(1)$, se usan tres aplicaciones dos se realizan con el conjunto de datos *Reconocimiento de Voz* y la tercera, con los datos *Carrera de 100 km*. Los resultados son comparados con los presentados por (Zimmerman & Núñez-Antón (2009), Zimmerman et al. (1998), Núñez-antón & Woodworth (1994)).

4.1. Simulaciones

4.1.1. Modelo de Simetría Compuesta

Para observar el comportamiento de la propuesta bayesiana se usa una simulación del modelo de regresión longitudinal:

$$y_{ij} = \beta_0 + \beta_1 t_{ij} + \beta_3 t_{ij}^2 + \epsilon_{ij}$$

Donde $i = 1, \dots, 20$ indica los individuos, $t = 1, 2, 3, 4$ representa los tiempos del modelo longitudinal y se supone que $\epsilon_{ij} \sim N(0, \Sigma(\rho, \theta))$, con $\Sigma(\rho, \theta)$ sigue una estructura como la presentada en 2.1, por lo tanto:

$$\mathbf{Y} \sim NMV(\mathbf{X}\boldsymbol{\beta}, \Sigma)$$

Para obtener el i -ésimo componente de $\mathbf{Y} = (Y_1, \dots, Y_{20})'$, donde $\mathbf{Y}_i = (Y_{i1}, Y_{i2}, Y_{i3}, Y_{i4})'$, se consideran como covariables los valores del tiempo en que se tomó la medición $t = (1, 2, 3, 4)$ y el cuadrado del mismo $t^2 = (1, 4, 9, 16)$. Si se supone un vector de parámetros:

$$\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)' = (30, 20, 20)',$$

Entonces, el vector de medias para el i -ésimo individuo viene dado:

$$\begin{pmatrix} \mu_{i1} \\ \mu_{i2} \\ \mu_{i3} \\ \mu_{i4} \end{pmatrix} = \begin{pmatrix} 1 & t_{i1} & t_{i1}^2 \\ 1 & t_{i2} & t_{i2}^2 \\ 1 & t_{i3} & t_{i3}^2 \\ 1 & t_{i4} & t_{i4}^2 \end{pmatrix} \begin{pmatrix} 30 \\ 20 \\ 20 \end{pmatrix} \quad (4.1)$$

Luego, se construye la matriz de varianzas y covarianzas que depende de los valores de σ^2 y ρ . Si se hace $\sigma^2 = 25$ y $\rho = 0.7$ siguiendo la estructura de la ecuación 2.1, la matriz de varianzas y covarianzas para las 3 observaciones realizadas sobre un individuo en particular está dada por:

$$\sigma^2 V_0 = \begin{pmatrix} 25 & 17.5 & 17.5 & 17.5 \\ 17.5 & 25 & 17.5 & 17.5 \\ 17.5 & 17.5 & 25 & 17.5 \\ 17.5 & 17.5 & 17.5 & 25 \end{pmatrix} \quad (4.2)$$

Bajo el supuesto de normalidad del modelo dado que $Y \sim (\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma}(\sigma^2, \rho))$, el valor de $\mathbf{Y}_i = (Y_{i1}, Y_{i2}, Y_{i3}, Y_{i4})'$, para un individuo en particular se obtienen de una distribución normal multivariada con media $(\mu_{i1}, \mu_{i2}, \mu_{i3})$, calculada en 4.1, donde la matriz de varianzas y covarianzas es dada en 4.2. Este mismo proceso se realiza para cada uno de los 20 individuos.

Considerando las distribuciones a priori,

$$\beta \sim N(b_n, B_n) \quad \sigma^2 \sim \Gamma\left(\frac{60 + 0.000000001}{2}, \frac{0.000000001 * 100 + R}{2}\right) \quad \rho \sim \text{Beta}(1, 1)$$

y dados los valores iniciales:

$$\beta_0 = (0, 0, 0) \quad \sigma_0^2 = 100 \quad \rho_0 = 0.5 \quad B_0 = \text{diag}(100^4).$$

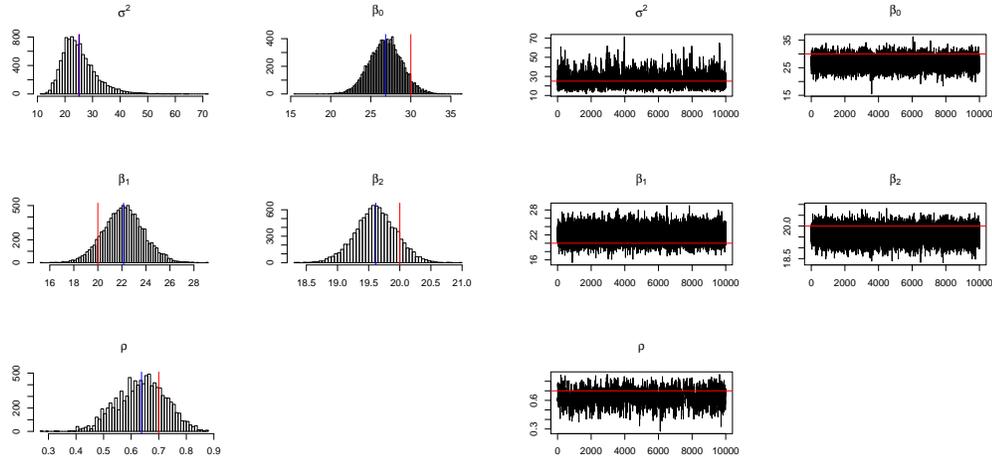
Realizando 20.000 iteraciones y con un Burn In de 10.000, los resultados obtenidos se resumen en la siguiente tabla.

Parámetro	Valor Simulado	Media	Mediana	D.Estandar
β_0	20	26.85134	26.86975	2.046394
β_1	2	22.1421	22.14507	1.681443
β_2	1	19.61449	19.615	0.3305445
σ^2	25	25.21477	24.08948	6.266709
ρ	0.7	0.6376172	0.6416781	0.08823302

TABLA 4.1. Valores Estimados simulación SC

Los histogramas obtenidos y las gráficas de las cadenas se presentan en la figura 4.1.

Además de las estimaciones y las gráficas obtenidas desde la aplicación del algoritmo, es necesario resaltar que, en esta simulación, la tasa de aceptación para la estimación de ρ fue de 32%.



(a) Histogramas (b) Cadenas

FIGURA 4.1. Gráficos Diagnosticos simulación SC

4.1.1.1. Una Distribución Apriori Informativa para la Correlación

Para observar el efecto que puede tener el uso de una distribución a priori informativa, se propone una distribución apriori para la correlación $Beta(8, 4)$.

Considerando las distribuciones a priori:

$$\beta \sim N(b_n, B_n) \quad \sigma^2 \sim \Pi\left(\frac{60 + 0.000000001}{2}, \frac{0.000000001 * 100 + R}{2}\right) \quad \rho \sim Beta(8, 4)$$

y dados los valores iniciales:

$$\beta_0 = (0, 0, 0) \quad \sigma_0^2 = 100 \quad \rho_0 = 0.5 \quad B_0 = diag(100^4).$$

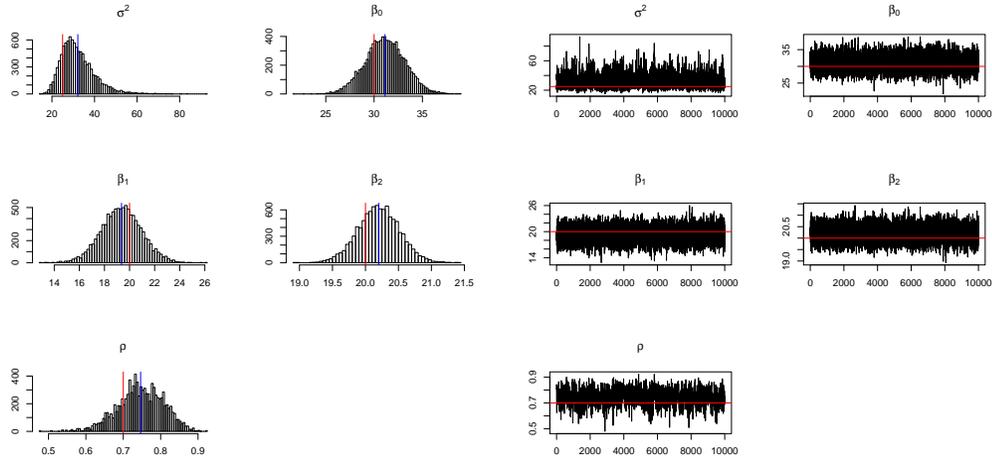
Realizando 20.000 iteraciones y con un periodo de quema o Burn In de 10.000, los resultados obtenidos se resumen en la siguiente tabla.

Parámetro	Valor Simulado	Media	Mediana	Desviación Estandar
β_0	30	31.13315	31.1178	2.063713
β_1	20	19.35872	19.36138	1.596044
β_2	20	20.20141	20.20014	0.3139896
σ^2	25	32.15983	30.76553	7.947891
ρ	0.7	0.7463998	0.7471375	0.06110183

TABLA 4.2. Valores Estimados simulación SC (Apriori Informativa)

La tasa de aceptación, usando una apriori informativa fue de 36%. Los histogramas y gráficas del comportamiento de las cadenas se muestran en la figura 4.2

Con el uso de la apriori informativa, se observa que disminuye la desviación estandar de las muestras para el parámetro de correlación y aumenta la tasa de aceptación del algoritmo, también se puede observar un mejor ajuste para los demás parámetros.



(a) Histogramas

(b) Cadenas

FIGURA 4.2. Gráficos Diagnosticos simulación SC (apriori Informativa)

4.1.2. Modelo AR(1)

Para observar el comportamiento de la propuesta bayesiana de la sección 3.2, se usa una simulación del modelo sin intercepto:

$$Y_i = \beta_1 x_{1ij} + \beta_2 x_{2ij} + \beta_3 x_{3ij} + \epsilon_{ij}$$

Donde $i = 1, \dots, 20$ indica los individuos, y $j = 1, 2, 3$ representa los tiempos del modelo longitudinal. Por lo cual al suponer que $\epsilon_{ij} \sim N(0, \Sigma(\rho, \theta))$, si Σ tiene una estructura AR(1) como la presentada en 2.3, entonces:

$$\mathbf{Y} \sim NMV(\mathbf{X}\boldsymbol{\beta}, \Sigma)$$

Para obtener el i -ésimo componente de $\mathbf{Y} = (Y_1, \dots, Y_{20})$ donde $\mathbf{Y}_i = (Y_{i1}, Y_{i2}, Y_{i3})'$, se consideran 3 covariables: X_1, X_2, X_3 Donde cada una de estas, al igual que la respuesta es medida en tres periodos de tiempo, es decir: $X_1 = (x_{i11}, x_{i21}, x_{i31})'$, $X_2 = (x_{i12}, x_{i22}, x_{i32})'$ y $X_3 = (x_{i13}, x_{i23}, x_{i33})'$. A diferencia de la simulación usada en el modelo de simetría compuesta, se supone que estas covariables no dependen del tiempo. Por lo tanto, los valores de estas covariables para el individuo i son obtenidas mediante tres muestras de las siguientes distribuciones uniformes: $X_1 \sim u(0, 20)$, $X_2 \sim u(5, 15)$ y $X_3 \sim u(4, 10)$. Suponiendo como vector de parámetros desconocidos:

$$\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3)' = (30, 20, 20)'$$

El vector de medias para el i -ésimo individuo viene dado por

$$\begin{pmatrix} \mu_{i1} \\ \mu_{i2} \\ \mu_{i3} \end{pmatrix} = \begin{pmatrix} x_{i11} & x_{i12} & x_{i13} \\ x_{i21} & x_{i22} & x_{i23} \\ x_{i31} & x_{i32} & x_{i33} \end{pmatrix} \begin{pmatrix} 30 \\ 20 \\ 20 \end{pmatrix} \quad (4.3)$$

Para construir la matriz de varianzas y covarianzas que depende de los valores de σ^2 y ρ , se hace $\sigma^2 = 81$ y $\rho = 0.6$ siguiendo la estructura de la ecuación (2.3) con esto el bloque $\sigma^2 V_0$ de la matriz de varianzas para las observaciones realizadas sobre el i -ésimo individuo es:

$$\sigma^2 V_0 = \begin{pmatrix} 81.00 & 48.6 & 29.16 \\ 48.60 & 81.0 & 48.60 \\ 29.16 & 48.6 & 81.00 \end{pmatrix} \quad (4.4)$$

Bajo el supuesto de normalidad del modelo, dado que $\mathbf{Y} \sim (\mathbf{X}\boldsymbol{\beta}, \Sigma(\sigma^2, \rho))$, los valores de \mathbf{Y}_{ij} para el individuo i se obtienen de una distribución normal multivariada con media $(\mu_{i1}, \mu_{i2}, \mu_{i3})$ calculada en 4.3 y con matriz de varianzas y covarianzas dada en 4.4. Este mismo proceso se realiza para cada uno de los 20 individuos, obteniendo los 60 valores que componen el vector de observaciones \mathbf{Y} .

Considerando las distribuciones a priori,

$$\beta \sim N(b_n, B_n) \quad \sigma^2 \sim \Gamma\left(\frac{60 + 0.000000001}{2}, \frac{0.000000001 * 100 + R}{2}\right) \quad \rho \sim Beta(1, 1)$$

y dados los valores iniciales:

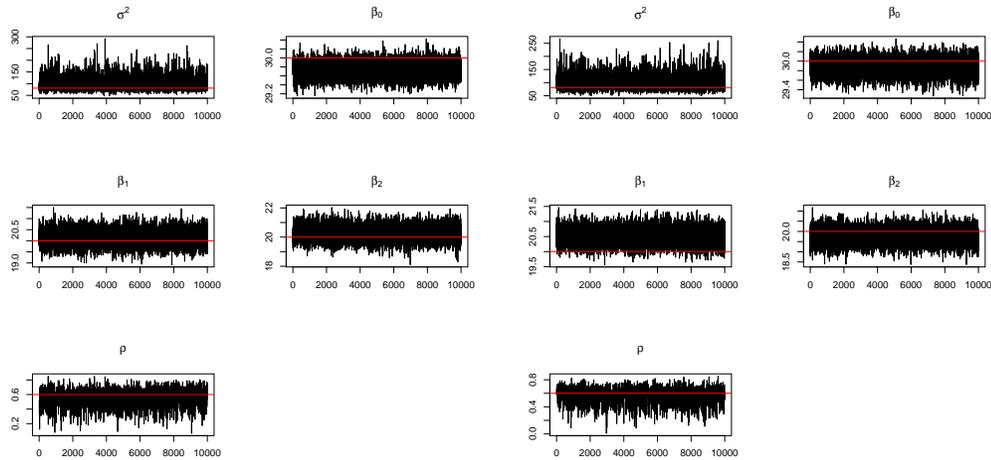
$$\beta_0 = (0, 0, 0) \quad \sigma_0^2 = 100 \quad \rho_0 = 0.5 \quad B_0 = \text{diag}(100^4).$$

Realizando 20.000 iteraciones, con un Burn In de 10.000, los resultados obtenidos se resumen en la tabla 4.3.

Parámetro	Valor Simulado	Media	Mediana	Desviación Estandar
σ^2	81	103.2703	99.16885	25.35267
β_1	30	29.73158	29.73177	0.1604566
β_2	20	20.14133	20.14142	0.3298006
β_3	20	20.28212	20.27972	0.4670534
ρ	0.6	0.5604326	0.5716748	0.118784

TABLA 4.3. Gráficos Diagnosticos Simulación AR(1)

La tasa de aceptación, para esta simulación fue de 38%. Los histogramas obtenidos y las gráficas de las cadenas semuestran en la figura 4.3.



(a) Histogramas

(b) Cadenas

FIGURA 4.3. Gráficos Diagnosticos Simulación AR(1)

4.1.2.1. Una Distribución Apriori Informativa para la Correlación

Para observar el efecto que puede tener utilizar una distribución a priori informativa, dado que para la simulación se uso el valor de $\rho = 0.6$ se propone una distribución apriori para la correlación $Beta(7, 5)$.

Considerando las distribuciones a priori,

$$\beta \sim N(b_n, B_n) \quad \sigma^2 \sim \Pi\left(\frac{60 + 0.000000001}{2}, \frac{0.000000001 * 100 + R}{2}\right) \quad \rho \sim Beta(7, 5)$$

y dando como valores iniciales:

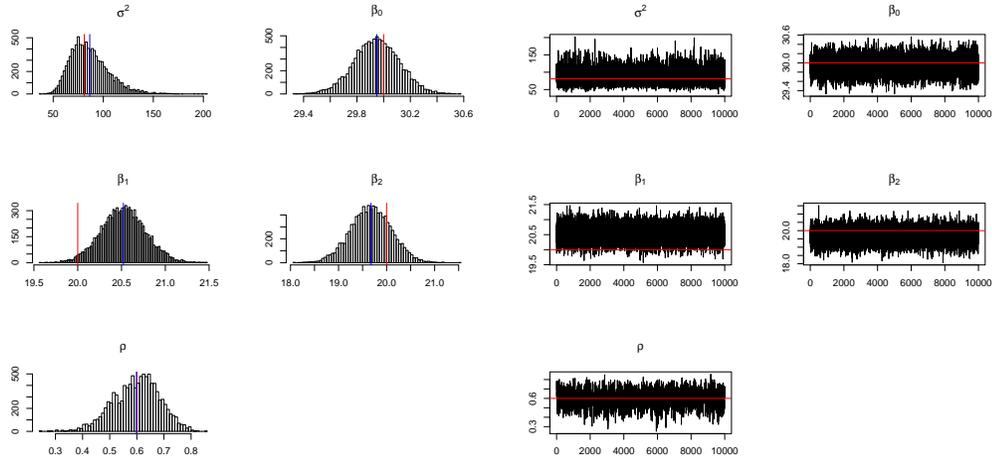
$$\beta_0 = (0, 0, 0) \quad \sigma_0^2 = 100 \quad \rho_0 = 0.5 \quad B_0 = \text{diag}(100^4).$$

Al realizar 20.000 iteraciones, tomando un BurnIn de 10.000, los resultados obtenidos se muestran en la tabla 4.4.

Parámetro	Valor Simulado	Media	Mediana	Desviación Estandar
σ^2	81	86.51554	83.78435	19.26938
β_1	30	29.9484	29.94821	0.1704781
β_2	20	20.52161	20.52103	0.2553712
β_3	20	19.67519	19.67305	0.4270217
ρ	0.6	0.5987282	0.6053743	0.08543353

TABLA 4.4. Valores Estimados Simulación AR(1) (Apriori Informativa)

La tasa de aceptación usando una apriori informativa fue de 42%. Los histogramas y gráficas del comportamiento de las cadenas son los presentados en la figura 4.4.



(a) Histogramas

(b) Cadenas

FIGURA 4.4. Gráficos Diagnosticos Simulación AR(1) (Apriori Informativa)

Se observa que el uso de la distribución apriori informativa disminuye la desviación estandar de las muestras para el parámetro de correlación y aumenta la tasa de aceptación del algoritmo de Metropolis-Hasting. Este mismo efecto, sobre la estimación del parámetro de correlación y sobre la tasa de aceptación del algoritmo, fue observado en la sección anterior en la cual se trabajó una simulación suponiendo una estructura de matriz de varianzas y covarianzas de simetría compuesta.

4.1.3. Modelo ARMA(1,1)

Siguiendo la misma metodología aplicada en la sección anterior, se hace una simulación del modelo sin intercepto:

$$y_{ij} = \beta_1 x_{1ij} + \beta_2 x_{2ij} + \beta_3 x_{3ij} + \epsilon_{ij}$$

Donde $i = 1, \dots, 20$ indica los individuos y $j = 1, 2, 3, 4, 5$ representa los tiempos del modelo longitudinal. Además, al suponer que $\epsilon_{ij} \sim N(0, \Sigma(\sigma^2, \rho, \phi))$, si Σ sigue una estructura ARMA(1,1) como se describió en 2.5.

$$\mathbf{Y} \sim NMV(\mathbf{X}\beta, \Sigma)$$

Para obtener el i -ésimo componente de $\mathbf{Y} = (Y_1, \dots, Y_{20})$ se consideran 3 covariables: X_1, X_2, X_3 . Donde cada una de estas, al igual que la respuesta, es medida en cinco periodos de tiempo, es decir: $X_1 = (x_{i11}, x_{i21}, x_{i31}, x_{i41}, x_{i51})'$, $X_2 = (x_{i12}, x_{i22}, x_{i32}, x_{i42}, x_{i52})'$ y $X_3 = (x_{i13}, x_{i23}, x_{i33}, x_{i43}, x_{i53})'$. Los valores de estas covariables para el individuo i son obtenidas mediante cinco muestras de las siguientes distribuciones uniformes: $X_1 \sim u(0, 20)$, $X_2 \sim u(5, 15)$, $X_3 \sim u(4, 10)$, $X_4 \sim u(5, 20)$ y $X_5 \sim u(7, 30)$. Al suponer un vector de parámetros

$$\beta = (\beta_1, \beta_2, \beta_3)' = (30, 20, 20)'$$

el vector de medias para el i -ésimo individuo se da por:

$$\begin{pmatrix} \mu_{i1} \\ \mu_{i2} \\ \mu_{i3} \\ \mu_{i4} \\ \mu_{i5} \end{pmatrix} = \begin{pmatrix} x_{i11} & x_{i12} & x_{i13} \\ x_{i21} & x_{i22} & x_{i23} \\ x_{i31} & x_{i32} & x_{i33} \\ x_{i41} & x_{i42} & x_{i43} \\ x_{i51} & x_{i52} & x_{i53} \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} \quad (4.5)$$

La matriz de varianzas y covarianzas que depende de los valores de σ^2 , ρ y ϕ . Se obtiene, al suponer $\sigma^2 = 64$, $\rho = 0.45$ y $\phi = 0.6$, siguiendo la estructura de la ecuación 2.5. Por consiguiente la matriz de varianzas y covarianzas para las cinco observaciones realizadas sobre el i -ésimo individuo, es:

$$\sigma^2 V_0 = \begin{pmatrix} 64.0000 & 38.400 & 17.28 & & 7.776 & 3.4992 \\ 38.4000 & 64.000 & 38.40 & 17.280 & 7.7760 & \\ 17.2800 & 38.400 & 64.00 & 38.400 & 17.2800 & \\ 7.7760 & 17.280 & 38.40 & 64.000 & 38.4000 & \\ 3.4992 & 7.776 & 17.28 & 38.400 & 64.0000 & \end{pmatrix} \quad (4.6)$$

Suponiendo normalidad en el modelo, ya que $\mathbf{Y} \sim (\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma}(\sigma^2, \rho))$, los valores de $Y_i = (Y_{i1}, Y_{i2}, Y_{i3}, Y_{i4}, Y_{i5})'$ para el individuo i se obtienen de una distribución normal multivariada con media calculada en 4.5 y con matriz de varianzas y covarianzas dada en 4.6. Este mismo proceso se realiza para cada uno de los 20 individuos, obteniendo de esta manera los 100 valores que componen el vector de observaciones \mathbf{Y} .

Suponiendo las distribuciones a priori para cada parámetro:

$$\beta \sim N(b_n, B_n) \quad \sigma^2 \sim \text{IG}\left(\frac{60 + 1 \times 10^{-8}}{2}, \frac{1 \times 10^{-8} * 100 + R}{2}\right) \quad \rho \sim \text{Beta}(1, 1) \quad \phi \sim \text{Beta}(1, 1)$$

y dados los valores iniciales:

$$\beta_0 = (0, 0, 0) \quad \sigma_0^2 = 100 \quad \rho_0 = 0.5 \quad \phi_0 = 0.9 \quad B_0 = \text{diag}(100^4)$$

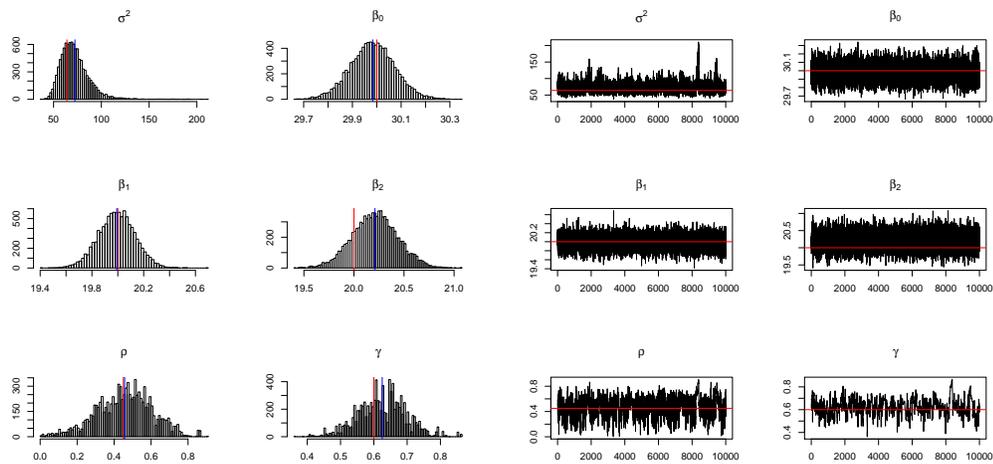
Al realizar 20.000 iteraciones con un o Burn In de 10.000, los resultados obtenidos se muestran en la tabla 4.5.

Parámetro	Valor Simulado	Media	Mediana	Desviación Estandar
σ^2	64	72.34198	70.03126	15.60649
β_1	30	29.98335	29.98314	0.09132147
β_2	20	19.9917	19.99318	0.1406774
β_3	20	20.21	20.21189	0.2270429
ρ	0.45	0.4569203	0.4644457	0.1512316
ϕ	0.6	0.6251613	0.6262012	0.07280917

TABLA 4.5. Valores Estimados ARMA(1,1)

La tasa de aceptación para el parámetro ρ fue de 35 %. Mientras que para el parámetro ϕ fue de 29 %.

Los histogramas obtenidos y las gráficas de las cadenas se pueden observar en la figura 4.5.



(a) Histogramas

(b) Cadenas

FIGURA 4.5. Gráficos Diagnosticos ARMA(1,1)

4.2. Aplicaciones

4.2.1. Aplicación sobre Small Mice Data

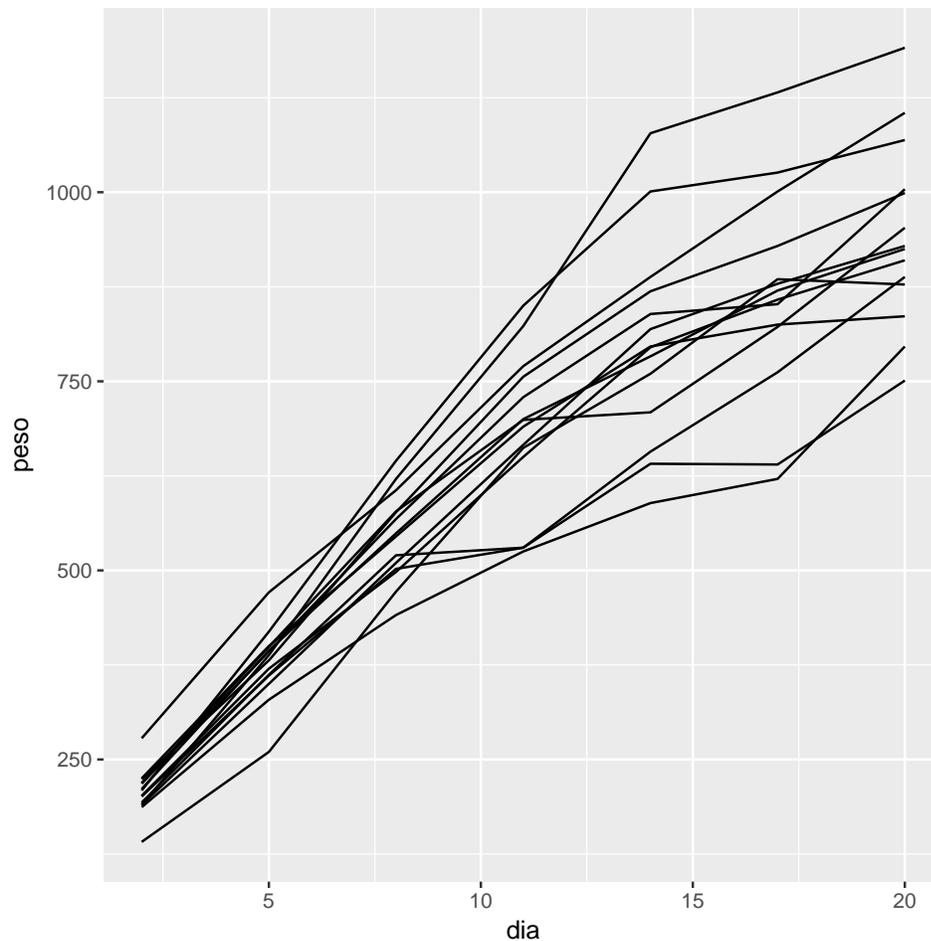
Los datos denominados *Small Mice Data* son tomados de Izenman & Williams (1989), el cual utiliza estos datos para los modelos espectrales para el análisis de datos longitudinales. En el estudio, se analizan más de 600 ratones en diferentes instantes: al momento de nacer, 7 días después del nacimiento (Inicio de crecimiento), 14 días (Cuando los ojos se abren e inicia el consumo de alimentos sólidos), 21 días (fin de la influencia materna para la alimentación) y 42 días (cuando la mayoría de los ratones alcanzan su madurez sexual). De estas 600 observaciones, un grupo particular de 35 ratones machos son divididos en 4 grupos. Los datos *Small Mice Data* forman un conjunto balanceado de datos longitudinales con los pesos en miligramos de 14 ratones los cuales conforman los grupos 3 y 4 del estudio realizado por Izenman & Williams (1989). Estos pesos fueron tomados en los días $t = 2, 5, 8, 11, 14, 17, 20$ después del nacimiento por la misma persona utilizando la misma escala de medición.

Weiss (2005), en el análisis exploratorio de los datos, presenta la siguiente matriz de correlación para los datos:

	2	5	8	11	14	17	20
2	1	0.92	0.57	0.36	0.23	0.23	0.38
5	0.92	1	0.77	0.54	0.45	0.41	0.55
8	0.57	0.77	1	0.86	0.8	0.76	0.81
11	0.36	0.54	0.86	1	0.93	0.92	0.87
14	0.23	0.45	0.8	0.93	1	0.96	0.89
17	0.23	0.41	0.76	0.92	0.96	1	0.92
20	0.38	0.55	0.81	0.87	0.89	0.92	1

TABLA 4.6. Matriz de Correlación para el SMD

A partir de la matriz de correlación 4.6, en Weiss (2005), se resalta una alta correlación entre las observaciones consecutivas (Supra Diagonal), de donde se muestra que los datos que presentan menos correlación son los pesos tomados entre los días 5 y 8 las restantes, aunque no son iguales, muestran bastante similitud. Se observa que las correlaciones de los tiempos 2,5,8 y 11 tiene alta correlación de orden 1, al igual que todas las correlaciones de tercer orden. Para observar el comportamiento de los datos en la gráfica 4.6 se muestran los perfiles para el conjunto de datos en cuestión.

FIGURA 4.6. Perfiles *SMD*

Dada la naturaleza creciente de los perfiles. Weiss (2005) ajusta tres modelos: un modelo lineal en el tiempo, un modelo cuadrático y un modelo cúbico. Debido a que el modelo que menos desviación estándar tiene sobre sus estimaciones es el modelo cuadrático, el autor propone estudiar el modelo:

$$Y_{ij} = \beta_0 + \beta_1 t + \beta_2 t^2 + \epsilon_{ij}, \quad (4.7)$$

Donde la variable respuesta es el peso de los ratones y $t = 2, 5, 8, 11, 14, 17, 20$. Para ajustar estos datos, el autor utiliza el método de máxima verosimilitud restringida para distintas estructuras de la matriz de varianzas y covarianzas, entre las que se encuentran las estructuras de simetría compuesta, no estructurada, $AR(1)$, $ARMA(1, 1)$, independientes, entre otras.

A través del mismo conjunto de datos en oprimer lugar se ajusta el modelo 4.7 utilizando la metodología bayesiana propuesta para la estructura de matriz de varianzas y covarianzas de simetría compuesta SC , autoregresivos de primer orden $AR(1)$ y autoregresivos de media móvil $ARMA(1, 1)$ y en segundo lugar, se compararán las estimaciones obtenidas mediante la propuesta bayesiana y las obtenidas mediante el método de máxima verosimilitud restringida ¹.

4.2.1.1. Small Mice Data con Estructura de Simetría Compuesta

Para hacer uso de la propuesta bayesiana para la estimación del parámetro de correlación, se utilizó una distribución apriori no informativa $Beta(1, 1)$. Se realizaron 20.000 iteraciones del algoritmo y un Burn In de 10.000.

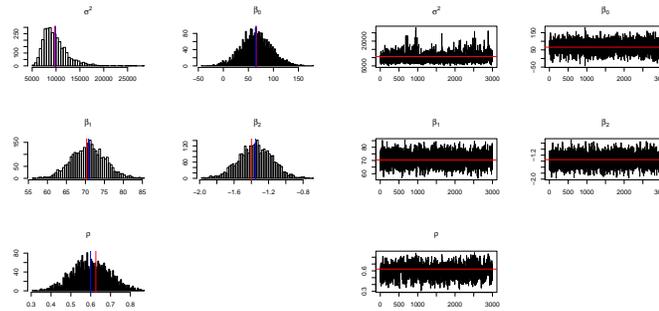
De este modo en la siguiente tabla se presentan los valores de la media y mediana a posteriori, obtenidos mediante la propuesta Bayesiana para cada parámetro, junto con su desviación estándar (valor entre paréntesis) y los valores estimados, obtenidos mediante la aplicación del método de máxima verosimilitud restringida.

Parámetro	Estimación Bayesiana		MVR
	Media	Mediana	Valor Estimado
$\hat{\beta}_0$	65.20008 (30.15849)	64.71494	65.745 (29.459)
$\hat{\beta}_1$	70.77607 (4.469975)	70.76598	70.3282 (4.4059)
$\hat{\beta}_2$	-1.351422 (0.1978872)	-1.350196	-1.349 (0.1949)
$\hat{\sigma}^2$	9889.113 (2663.29)	9348.963	9671.2533 (2621.89)
$\hat{\rho}$	0.6031316 (0.09559118)	0.6057473	0.6258 (0.1073)

TABLA 4.7. Valores Estimados por Propuesta Bayesiana a SMD con Estructura SC y Valores Estimados Mediante MVR

¹las estimaciones obtenidas por Weiss (2005) fueron verificadas mediante el software SPSS 20

Se puede observar que existe una aproximación óptima de las estimaciones proporcionadas por la metodología Bayesiana propuesta y las obtenidas mediante el ajuste del modelo marginal utilizando herramientas de estadística clásica. Los criterios de comparación se presentan al final de este capítulo en la tabla 4.10. Para observar el comportamiento de la estimación bayesiana alrededor de las estimaciones clásicas, se presentan los histogramas y gráficas de las cadenas para cada parámetro.



(a) Histogramas

(b) Cadenas

FIGURA 4.7. Gráficos Diagnosticos SMD SC

4.2.1.2. Small Mice Data con Estructura AR(1)

En esta sección, se muestran las estimaciones obtenidas para el modelo 4.7, si se supone que la matriz de varianzas y covarianzas tiene estructura autoregresiva de primer orden como la que se muestra en la ecuación 2.3.

Para la estimación del parámetro de correlación ρ se hace uso de una distribución apriori no informativa $Beta(1, 1)$. Además, se realizan 20.000 iteraciones con un periodo de quema de 10.000.

En la tabla 4.8 se presentan los valores de la media y mediana obtenidos mediante la propuesta bayesiana para cada parámetro, junto con su desviación estándar (valor entre paréntesis) y los valores estimados obtenidos mediante la aplicación del método de máxima verosimilitud restringida.

Parámetro	Estimación Bayesiana		MVR
	Media	Mediana	Valor Estimado
$\hat{\beta}_0$	73.8437 (28.12993)	73.5186	74.083 (28.1864)
$\hat{\beta}_1$	68.6131 (4.089173)	68.62998	68.588 (3.9943)
$\hat{\beta}_2$	-1.252414 (0.1729606)	-1.254393	-1.2515 (0.1688)
$\hat{\sigma}^2$	8622.617 (2663.29)	8130.642	8796.697 (2488)
$\hat{\rho}$	0.8558992 (0.03680179)	0.8574861	0.8737 (0.03784)

TABLA 4.8. Valores Estimados por Propuesta Bayesiana a SMD con Estructura AR(1) y Valores Estimados Mediante MVR

Considerando una estructura de matriz de varianzas y covarianzas para ajustar el modelo 4.7 la metodología bayesiana propuesta en la sección 3.2 hace una buena aproximación a las estimaciones obtenidas mediante la máxima verosimilitud restringida. A continuación se presentan los histogramas y las gráficas de las cadenas para el ajuste bayesiano de *Small Mice Data*.

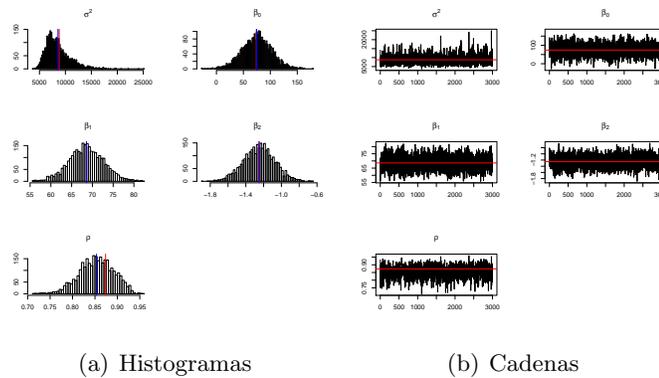


FIGURA 4.8. Gráficos Diagnosticos SMD AR(1)

4.2.1.3. Small Mice Data con Estructura ARMA(1,1)

Si se considera que el modelo 4.7 tiene como estructura para la matriz de varianzas y covarianzas autoregresiva de primer orden con promedio móvil $ARMA(1,1)$ como la de la ecuación 2.5 y considerando distribuciones apriori no informativas $Beta(1,1)$ para la estimación de los dos parámetros de correlación ρ, ϕ , al realizar el algoritmo, al igual que en los anteriores casos con 20.000 iteraciones y un BurnIn de 10.000; los resultados de la media, la mediana y la desviación estandar (valor entre parentesis) para las muestras a posteriori de todos los parámetros del modelo se presentan en la tabla 4.9.

Parámetro	Estimación Bayesiana		MVR
	Media	Mediana	Valor Estimado
$\hat{\beta}_0$	77.55587 (28.09933)	77.51176	77.4183 (28.494)
$\hat{\beta}_1$	67.68745 (4.53446)	67.734758	67.6197 (4.3165)
$\hat{\beta}_2$	-1.20763 (0.1943583)	-1.210046	-1.2043 (0.1825)
$\hat{\sigma}^2$	8169.547 (2293.082)	7684.294	8796.697 (2488)
$\hat{\rho}$	0.7924982 (0.05394172)	0.796965	0.8322 (0.05479)
$\hat{\phi}$	0.8421932 (0.03405925)	0.8448967	0.8732 (0.035312)

TABLA 4.9. Valores Estimados por Propuesta Bayesiana a SMD con Estructura ARMA(1,1) y Valores Estimados Mediante MVR

En el caso de considerar la estructura de matriz de varianzas y covarianzas ARMA(1,1), la propuesta bayesiana presentada en la sección 3.3 ofrece una óptima aproximación a los resultados obtenidos por el método de máxima verosimilitud restringida. Las gráficas de los histogramas y las cadenas obtenidas desde el algoritmo bayesiano para el ajuste del *Small Mice Data* son las siguientes:

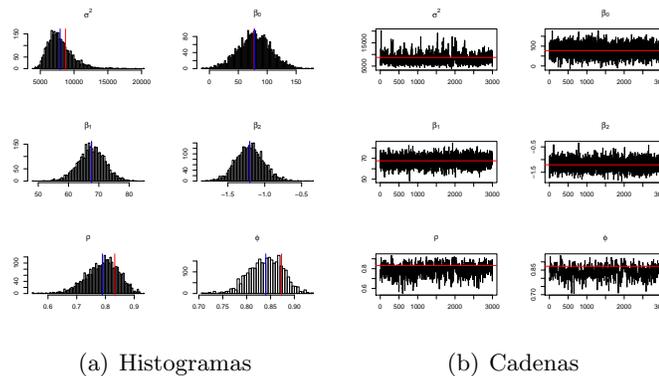


FIGURA 4.9. Gráficos Diagnosticos SMD ARMA(1,1)

Con el propósito de realizar una comparación para los modelos estimados, via la metodología bayesiana propuesta en este capítulo y los obtenidos por medio de la estimación clásica, se presenta los valores de AIC, BIC y DIC.

Modelo	AIC	BIC	DIC
SC (Clásico)	1109.3	1110.6	-
SC (Bayesiano)	1124.3	1137.2	1122.9
AR(1) (Clásico)	1039.3	1040.5	-
AR(1) (Bayesiano)	1054.7	1067.6	1053.3
ARMA(1,1) (Clásico)	1038.9	1040.8	-
ARMA(1,1) (Bayesiano)	1054.6	1070.1	1052.7

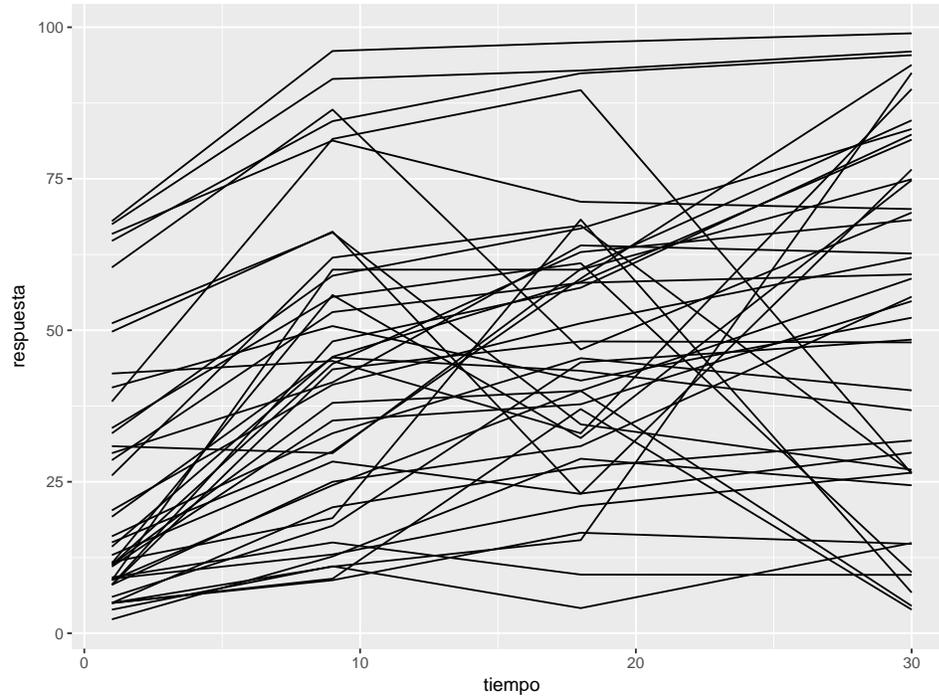
TABLA 4.10. Criterios de Comparación

4.2.2. Aplicaciones a los datos de Reconocimiento de voz

Este conjunto de datos proviene del estudio audiológico presentado en Tyler (1988). El estudio general se realizó con cinco tipos distintos de implantes: 3 implantes de canales simples y 2 implantes multicanal. Los implantes fueron insertados quirúrgicamente cinco a seis semanas antes de conectarse eléctricamente a un procesador de voz externo. El conjunto de datos consta de los puntajes obtenidos al realizar una prueba de reconocimiento de voz, efectuada sobre pacientes, los cuales tuvieron un implante coclear multicanal. Los pacientes se clasifican en dos grupos, dependiendo del tipo de implante que recibieron (A y B): 21 sujetos con el implante A y 21 sujetos con el implante B. En el estudio, los individuos eran bilateralmente sordos, por lo tanto los valores bases de la prueba son todos 0. Las pruebas se efectuaron en 4 intervalos de tiempo 1, 9, 18 y 30 meses luego de haber recibido el implante. En el estudio se presentó cierta variación en los tiempos reales de seguimiento, por lo cual estos tiempos no eran exactos. Además, algunos sujetos no se presentaron en uno o más de sus seguimientos programados, por lo que faltan algunos datos. Se tienen 8 datos faltantes en las observaciones del mes 19 y 21 datos faltantes en las observaciones del mes 30.

El interés del estudio de estos datos se centra en describir cómo el rendimiento audiológico de los individuos que reciben cada tipo de implante depende del tiempo transcurrido desde la implantación. Específicamente, en primer lugar, se desea saber cómo los perfiles medios de los dos tipos de implantes se comparan entre sí, y en segundo lugar, si el rendimiento auditivo de un sujeto tiende a ser más consistente en el tiempo (Zimmerman & Núñez-Antón (2009)).

La gráfica 4.10 muestra los perfiles de los datos. A partir del análisis exploratorio presentado en (Zimmerman & Núñez-Antón (2009)) se destaca el incremento de la media a medida que se avanza en el tiempo. También, las varianzas parecen incrementar ligeramente de la primera a la segunda medición, sin embargo permanecen constantes a partir de entonces. El autor ajusta modelos de simetría compuesta y autoregresivos de primer orden, tanto en su versión homogénea como heterogénea, pero finalmente decide que el modelo que mejor se ajusta es un modelo de antedependencia estructurada de orden 1 $ADS(1)$.

FIGURA 4.10. Perfiles *Reconocimiento de Voz*

4.2.2.1. Modelo ADE(1) Tipo 1

En Núñez-antón & Woodworth (1994) se propone ajustar el modelo:

$$Y_{it} = \beta_0 + \beta_1 t + \beta_2 t^2 + \epsilon_{it}$$

Donde ϵ_{it} sigue una distribución normal de media cero para el i -ésimo individuo; y se supone una matriz de varianzas covarianzas dada por 2.7, cuya función de potencias se define como en 2.10 y $G(t, \psi) = 1$. Ya que este conjunto de datos es desbalanceado los registros faltantes son obtenidos de manera aleatoria.

Para hacer uso de la propuesta Bayesiana en la estimación del modelo presentado en la sección 3.4.1, para la estimación del parámetro de correlación ρ se hace uso de una distribución apriori no informativa $Beta(1, 1)$. Por otro lado, para el parámetro λ se usa una distribución $U(-1, 1)$ y la estimación de β y σ^2 se obtienen mediante muestreo de Gibbs.

Después de 20.000 iteraciones con un periodo de quema de 10.000 y tomando como valores iniciales:

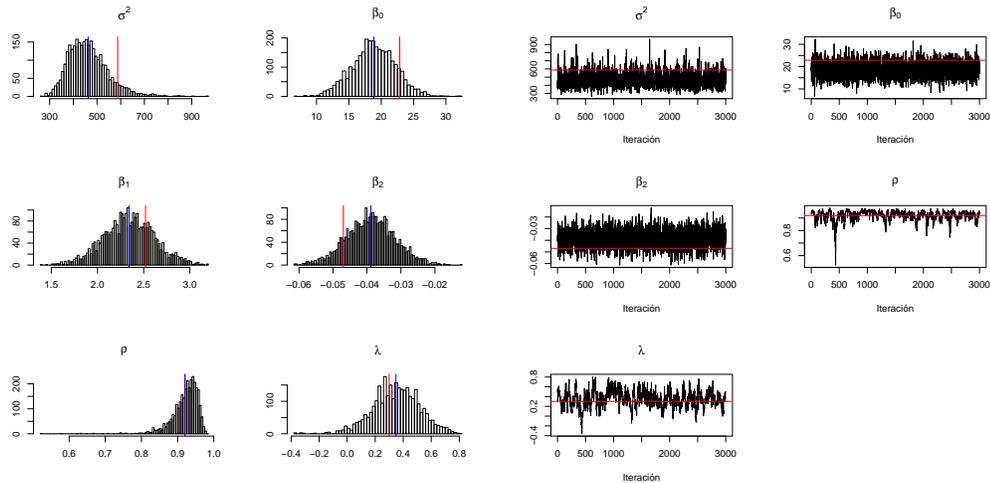
$$\beta_0 = (20, 1, 0) \quad \sigma_0^2 = 100 \quad \rho_0 = 0.5 \quad \lambda = 0.5 \quad B_0 = \text{diag}(100^4)$$

De este modo, los resultados obtenidos mediante la propuesta bayesiana así como los resultados presentados en Núñez-antón & Woodworth (1994) se pueden observar en la tabla 4.11.

	β_0	β_1	β_2	σ^2	ρ	λ
Media	22.33057	2.536971	-0.047993	602.0287	0.9339738	0.297134
D.E	4.294531	0.3132022	0.0080857	112.6369	0.02521215	0.1444972
Mediana	22.37508	2.540172	-0.048007	585.5624	0.9391685	0.3000189
E. Clásica	22.85	2.52	-0.04096	587.15	0.94	0.30

TABLA 4.11. Valores Estimados Reconocimiento de Voz (ADS(1)) Modelo 1

Para observar el comportamiento óptimo de la propuesta bayesiana se presentan a continuación los histogramas y cadenas obtenidas desde el algoritmo propuesto.



(a) Histogramas

(b) Cadenas

FIGURA 4.11. Gráficos Diagnósticos Reconocimiento de Voz (ADS(1)) Modelo 1

Las tasas de aceptación para los parámetros ρ y λ son respectivamente de 39% y 32%. los valores obtenidos para los criterios de comparación son $AIC = 1322.752$, $BIC = 1341.351$, $DIC = 1320.714$.

Con la información contenida en 4.11 y los gráficos en 4.11, se puede observar que la propuesta bayesiana brinda una aproximación cercana a los valores obtenidos en Núñez-antón & Woodworth (1994), la falta de ajuste se puede atribuir a que el cálculo de los datos faltantes se realizó de manera diferente en Núñez-antón & Woodworth (1994) y en este trabajo.

4.2.2.2. Modelo ADE(1) Tipo 2

En Zimmerman et al. (1998), se propone el modelo con estructura de antedependencia estructurada de orden 1 para los datos de reconocimiento de voz:

$$Y_{ij} = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 z_k + \beta_4 z_k t + \beta_5 z_k t^2 + \epsilon_{ij}$$

Con $k = 1, \dots, 42$ $t = 1, 9, 18, 30$. $z_k = 1$ si el individuo k recibió el implante tipo A o 0 en otro caso, además se supone que ϵ_{ij} sigue una distribución normal de media cero. Sí

para el i -ésimo individuo se supone una matriz de varianzas covarianzas dada por 2.7, con función de potencias como la presentada en 2.10 y

$$G(t, \psi) = \begin{cases} 1 & t = 1 \\ \psi & \text{En otro caso.} \end{cases}$$

Para hacer uso de la propuesta bayesiana sobre el modelo presentado en la sección 3.4.2 para la estimación del parámetro de correlación ρ , se usa una distribución a priori no informativa $Beta(1, 1)$. Para el parámetro λ , se usa una distribución $U(-1, 1)$. Mientras, para Δ , se consideró una distribución $N(0, 1)$. Para la estimación de β y σ^2 se suponen aprioris normales y gamma respectivamente como en 3.7, 3.10.

Después de 10.000 iteraciones, con un periodo de quema de 5.000 y tomando como valores iniciales:

$$\beta_0 = (20, 1, 0, 8, 1, 0) \quad \sigma_0^2 = 100 \quad \rho_0 = 0.5 \quad \lambda = 0.5 \quad \Psi_0 = 1 \quad B_0 = \text{diag}(100^4)$$

De este modo, los resultados obtenidos para la estimación de la estructura media se pueden observar en la siguiente tabla, sin embargo, no es posible realizar una comparación directa con las estimaciones por el método de máxima verosimilitud restringida ya que los autores no proporcionan estos valores estimados.

	β_0	β_1	β_2	β_3	β_4	β_5
Media	13.82724	2.249744	-0.04412	14.71942	0.3948032	-0.0097352
D.E	3.959064	0.3865282	0.010280	5.662908	0.5506937	0.01463349
Mediana	13.83731	2.246179	-0.04396	14.63447	0.3982116	-0.0098186

TABLA 4.12. Valores Estimados Estructura Media Reconocimiento de Voz (ADS(1)) Modelo 2

Los resultados obtenidos para la estimación de los parámetros de la matriz de varianzas y covarianzas se pueden observar en la siguiente tabla, así como los valores para estos mismos parámetros presentados en Zimmerman et al. (1998):

	σ^2	ρ	λ	ψ
Media	334.0463	0.9277529	0.3232773	1.773276
D.E	71.85849	0.0380726	0.1922905	0.3156079
Mediana	325.0523	0.9364507	0.3303191	1.745823
E. Clásica	388.7	0.935	0.240	1.615

TABLA 4.13. Valores Estimados Varianzas y Covarianzas Reconocimiento de Voz (ADS(1)) Modelo 2

Las gráficas siguientes muestran los histogramas y cadenas obtenidas mediante el proceso de estimación de los 10 parámetros considerados en el modelo.

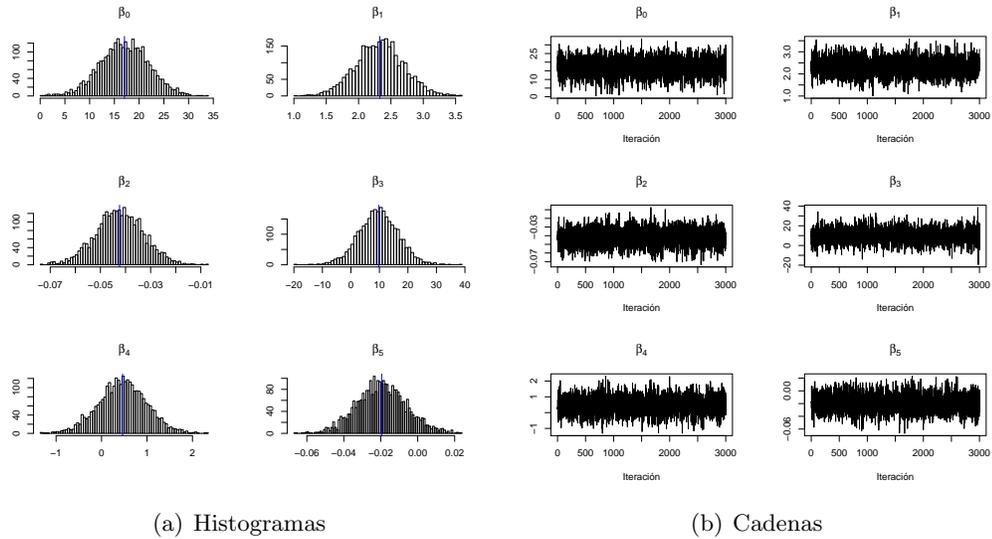


FIGURA 4.12. Gráficos Diagnósticos Estructura Media Reconocimiento de Voz (ADS(1)) Modelo 2

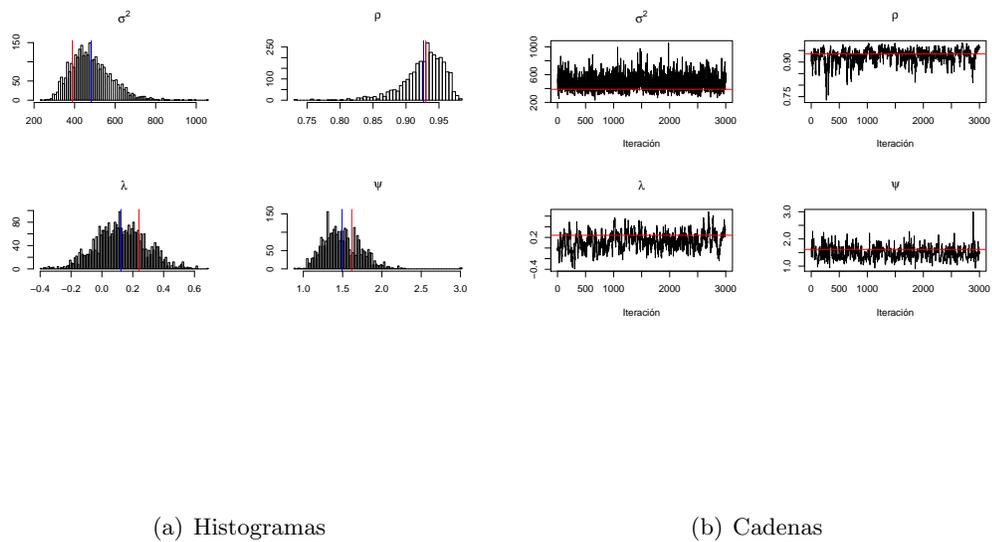


FIGURA 4.13. Gráficos Diagnósticos Varianzas y Covarianzas Reconocimiento de Voz (ADS(1)) Modelo 2

Las tasas de aceptación para los parámetros ρ , λ y ψ , fueron respectivamente de 35 %, 32 % y 34 %, mientras que los criterios de información son: $AIC = 1287.364$, $BIC = 1318.362$, $DIC = 1283.554$.

Las estimaciones para los parámetros que componen la matriz de varianzas y covarianzas del modelo se pueden considerar una buena aproximación a las obtenidas en otros trabajos, en los cuales se ajustan los modelos mediante métodos de estimación clásicos. Nuevamente, la falta de ajuste se le puede atribuir a que el cálculo de los datos faltantes se realizó de manera diferente en Zimmerman et al. (1998) y en este trabajo.

4.2.3. Datos Carrera de 100 Km

Estos datos presentados en Zimmerman & Núñez-Antón (2009) corresponden a los tiempos parciales en minutos para cada uno de los 80 competidores en las secciones de 10 kilómetros de una carrera de 100 kilómetros, la cual, se llevó a cabo en el Reino Unido en 1984. Además de los tiempos parciales, los datos contienen la edad de 76 de los 80 competidores. Algunos gráficos descriptivos de estos datos son mostrados y analizados en Everitt (1994). La matriz de correlación muestral para los datos de *Carrera de 100 km* es:

Sección	1	2	3	4	5	6	7	8	9	10
1	1	.95	.84	.78	.60	.60	.52	.45	.51	.38
2	.95	1	.89	.82	.63	.62	.54	.48	.51	.40
3	.84	.89	1	.92	.75	.72	.60	.61	.56	.44
4	.78	.82	.92	1	.88	.84	.69	.69	.65	.49
5	.60	.63	.75	.88	1	.94	.95	.78	.73	.52
6	.60	.62	.72	.84	.94	1	.84	.84	.77	.64
7	.52	.54	.60	.69	.95	.84	1	.78	.69	.72
8	.45	.48	.61	.69	.78	.84	.78	1	.75	.65
9	.51	.51	.56	.65	.73	.77	.69	.75	1	.77
10	.38	.40	.44	.49	.52	.64	.72	.65	.77	1
Varianza	26.9	34.8	49.0	58.9	91.4	149.9	107.9	152.2	145	167.2

TABLA 4.14. Correlaciones y Varianzas Muestrales *Carrera 100 Km*

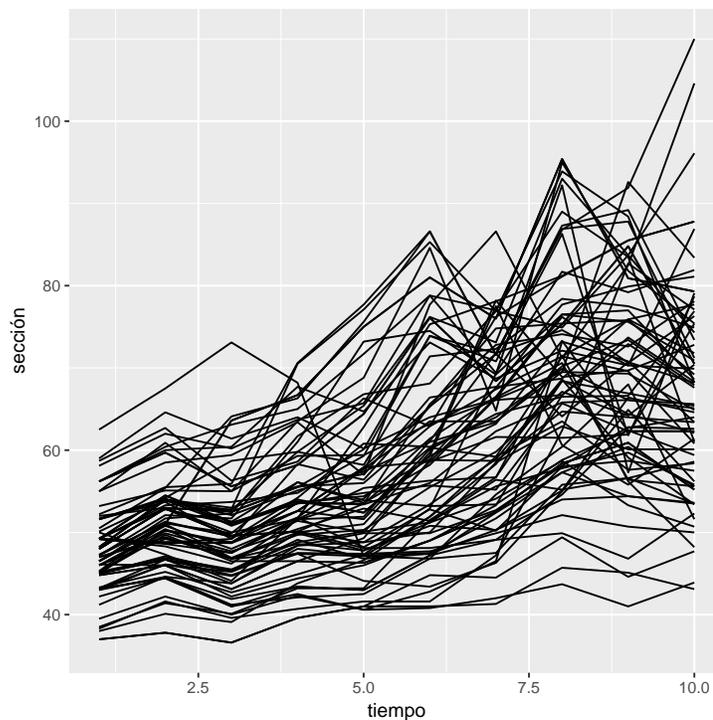


FIGURA 4.14. Perfiles *Carrera 100 Km*.

En Núñez-Antón & Zimmerman (2001) se resalta la heterogeneidad de las varianzas, el hecho de que todas las correlaciones sean positivas, la existencia de correlación serial, y que

las correlaciones entre observaciones separadas por la misma distancia no son constantes. Esto se puede atribuir a que los tiempos utilizados en una sección de la carrera serán mejores predictores de los tiempos en las secciones siguientes al principio de la carrera que al final de ella.

Con los datos de la *carrera de 100 km* se ajusta el modelo $ADE(1)$ con estructura para la media dada de la siguiente manera:

$$E(\mathbf{Y}_{ij}) = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3$$

$$i = 1, \dots, 74 \qquad y \qquad j = 1, \dots, 10$$

además con estructura de covarianzas dada en 2.7, donde:

$$\rho_j = \rho^{f(t_j, \lambda) - f(t_{j+1}, \lambda)} \qquad (j = 2, \dots, 10) \qquad (4.8)$$

Y

$$f(t, \lambda) = \begin{cases} (t^\lambda - 1)/\lambda & \lambda \neq 0 \\ \log(t) & \lambda = 0 \end{cases} \qquad (4.9)$$

$$\sigma_j^2 = \exp(\psi_0 + \psi_1 t_j + \psi_2 t_j^2) \qquad y \qquad (j = 1, \dots, 10) \qquad (4.10)$$

En Zimmerman et al. (1998) se ajustan estos datos, con la diferencia que se propone ajustar el modelo con estructura de covarianzas determinada por la función de potencias 2.10 y función para la estructura de covarianzas dada por:

$$G(t, \psi) = 1 + \psi_1 t + \psi_2 t^2 \qquad (4.11)$$

Para hacer uso del algoritmo presentado en la sección 3.4.3, para el parámetro de correlación ρ , se hace uso de una distribución a priori no informativa, $Beta(1, 1)$. Para el parámetro λ , se toma una distribución $U(-1, 1)$. Mientras que, para ψ como lo indica el algoritmo, se selecciona una distribución a priori normal $N(\Psi_0, K_0)$.

Después de 15.000 iteraciones con un periodo de quema de 5.000 y tomando como valores iniciales.

$$\beta_0 = (20, 1, 0, 0) \quad \rho_0 = 0.5 \quad \lambda = 0.5 \quad \Psi_0 = (1, 1, 1) \quad B_0 = \text{diag}(100^4) \quad K_0 = \text{diag}(0.1384),$$

De este modo, los resultados obtenidos para la estimación de la estructura media se muestran en la tabla 4.15, así como también los valores de las estimaciones de los mismos parámetros ajustados mediante la metodología de máxima verosimilitud restringida, los cuales son presentados en Zimmerman et al. (1998).

	β_0	β_1	β_2	β_3
Media	44.55874	-2.398607	1.321879	-0.09647709
D.E	1.634959	2.113936	0.7487526	0.07142071
Mediana	44.56703	-2.381146	1.318993	-0.09602199
E Clásica	43.428	1.354	0.253	-0.0169

TABLA 4.15. Valores Estimados Estructura de la Media *Carrera 100 km ADS(1)*

Aunque se observan diferencias entre las estimaciones clásicas y las obtenidas mediante la propuesta bayesiana. Los valores obtenidos en Zimmerman et al. (1998) para $(\beta_1, \beta_2, \beta_3)$ están dentro de los intervalos de credibilidad del 95 % que se generan con los valores ajustados mediante la estimación bayesiana, es decir:

$$IC(0.95)_{\beta_1} = (-11.77, 6.95) \quad (4.12)$$

$$IC(0.95)_{\beta_2} = (-4.90, 7.547)$$

$$IC(0.95)_{\beta_3} = (-0.5433, 0.3513)$$

Las estimaciones mediante la propuesta bayesiana y los valores estimados en Zimmerman et al. (1998), para los parámetros ρ y λ , se muestran a continuación:

	ρ	λ
Media	0.9225286	1.722776
D.E	0.02771019	0.2643785
Mediana	0.92517817	1.694609
E. Clásica	0.929	1.6

TABLA 4.16. Valores Estimados Estructura de Covarianzas *Carrera 100 km ADS(1)*

Los valores estimados para la estructura de las varianzas σ_i no son comparables ya que el modelo aquí propuesto es diferente al utilizado en Zimmerman et al. (1998). En el modelo bayesiano, se obtiene que los valores estimados para $\Psi = (\psi_0, \psi_1, \psi_2)$ son:

	ψ_0	ψ_1	ψ_2
Media	2.785591	0.4491583	-0.02198412
D.E	0.01238902	0.02222586	0.01609199
Mediana	2.785489	0.4490198	-0.02207433

TABLA 4.17. Valores Estimados Ψ en la Estructura de Covarianzas *Carrera 100 km ADS(1)*

sin embargo, en Zimmerman et al. (1998) los parámetros estimados para la función que modela las varianzas es $\hat{\sigma}^2 = 16.952$, $\hat{\psi}_1 = 0.590$, $\hat{\psi}_2 = 0.450$.

Para observar el comportamiento de las cadenas propuestas en la estimación bayesiana, se presentan los histogramas y el recorrido de las cadenas en la siguiente figura:

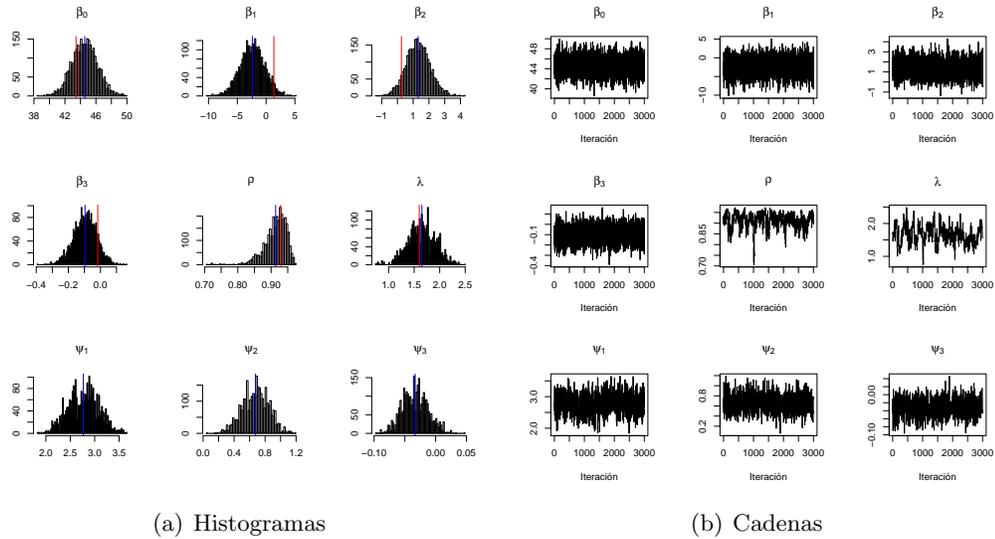


FIGURA 4.15. Gráficos Diagnosticos Carrera 100 Km $ADS(1)$

Para el modelo bayesiano, se tiene los criterios de información $AIC = 1401.88$, $BIC = 1425.31$, $DIC = 1403.078$. La tasa de aceptación para ρ es del 34 %, para λ es de 37 % y para Ψ , por provenir de una variable de trabajo, tiene una tasa de aceptación del 84 %.

Para observar el efecto de la estimación bayesiana sobre el modelado de la función de varianzas, se presenta en la siguiente tabla un comparativo entre los resultados obtenidos por la propuesta acá presentada y los valores muestrales de las varianzas de cada sección de la carrera. Los valores se obtienen al reemplazar las estimaciones sobre las funciones 4.10.

	Muestral	Bayesiana
$\hat{\sigma}_1^2$	26.89	24.84136
$\hat{\sigma}_2^2$	34.78	36.43719
$\hat{\sigma}_3^2$	49.01	51.14732
$\hat{\sigma}_4^2$	58.89	68.70830
$\hat{\sigma}_5^2$	91.41	88.32913
$\hat{\sigma}_6^2$	149.90	108.66937
$\hat{\sigma}_7^2$	107.85	127.94366
$\hat{\sigma}_8^2$	152.22	144.15801
$\hat{\sigma}_9^2$	144.99	155.44158
$\hat{\sigma}_{10}^2$	167.21	160.39989

TABLA 4.18. Valores Estimados Varianzas por Sección Carrera 100 km $ADS(1)$

Con el propósito de dar una mejor comparación de los valores estimados mediante la propuesta bayesiana y el ajuste presentado en Zimmerman et al. (1998), se presenta a continuación las gráficas de los residuales. Por un lado la gráfica superior muestra los residuales para el modelo ajustado mediante metodología clásica y por otro lado, la gráfi-

ca inferior muestra los residuales para el modelo ajustado por la metodología bayesiana propuesta en la sección 3.4.3.

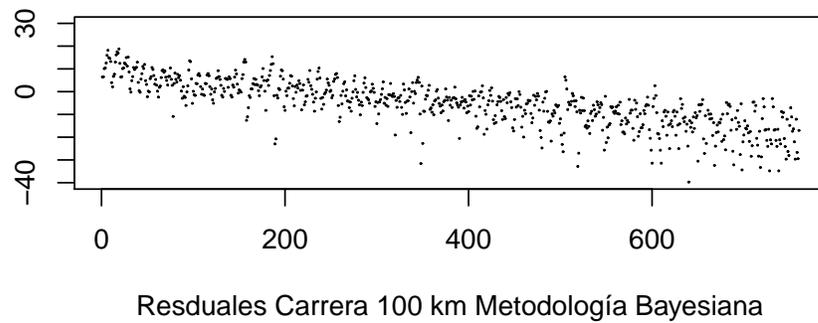
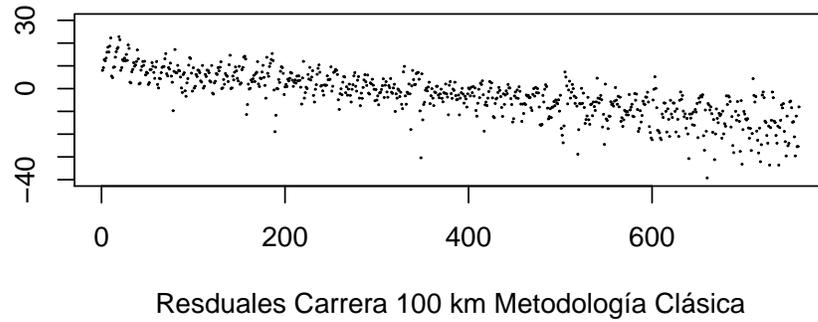


FIGURA 4.16. Residuales Carrera 100 Km $ADS(1)$

Las diferencias entre los residuales de la estimación del modelo mediante estimación clásica y estimación bayesiana no son significativamente grandes. Además, la suma de cuadrados del error calculada sobre el modelo estimado mediante máxima verosimilitud restringida es $SCE = 83455.21$, mientras que la suma de cuadrados del error calculada sobre el modelo bayesiano es de $SCE = 94598.3$. Con el propósito de comparar el comportamiento de los residuales para cada sección de la carrera, se muestran los boxplots de estos residuales tanto para el modelo clásico como el modelo bayesiano.

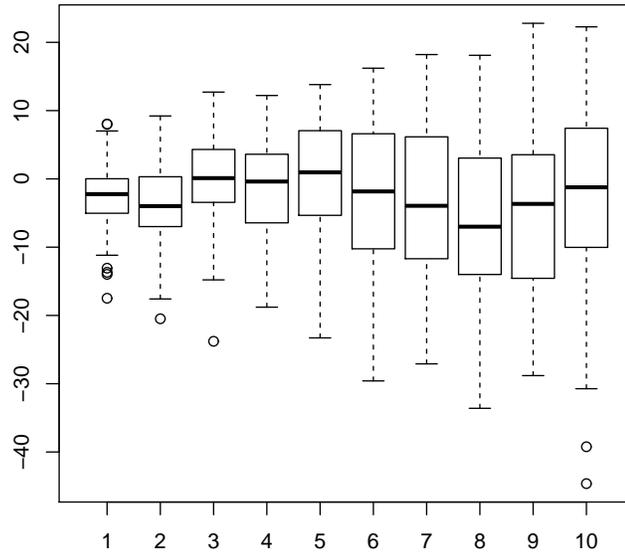


FIGURA 4.17. Box-plot de los Residuales por Sección Carrera 100 Km ADS(1), Ajustados Mediante MVR.

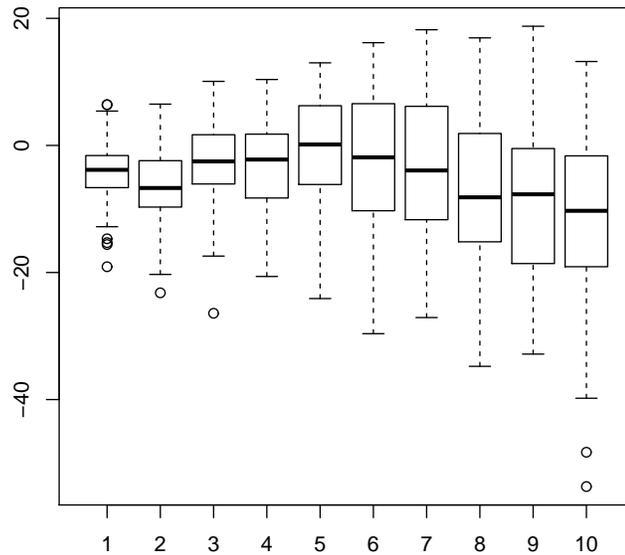


FIGURA 4.18. Box-plot de los Residuales por Sección Carrera 100 Km ADS(1), Ajustados Mediante la 3.4.3.

Estos gráficos evidencian que los residuales tomados para cada sección de la carrera, obtenidos mediante el ajuste clásico del modelo o el ajuste bayesiano, tampoco difieren en gran parte uno del otro. En conclusión la propuesta bayesiana brinda un ajuste aproximado al obtenido mediante la estimación vía máxima verosimilitud restringida presentado en Zimmerman et al. (1998).

4.2.4. Estudio de Sensibilidad

En esta sección, se presentan los resultados de la estimación bayesiana de la varianza σ^2 al cambiar los hiperparámetros usados en la distribución a priori 3.8, para los tres modelos ajustados del *SMD* y los dos modelos *ADS(1)* que se usaron para modelar los datos de *reconocimiento de voz*. Se muestran los cambios sobre las estimaciones de la varianza al utilizar como hiperparámetros los valores 0.001, 0.000001, 0.000000001 y 0.0000000001. La tabla 4.19 muestra los valores medios de la cadena obtenida mediante muestre de Gibbs de la distribución a posteriori 3.10 y el valor de la desviación de los mismos el cuál se encuentra entre paréntesis para cada uno de los modelos mencionados.

	0.001	0.000001	0.000000001	0.0000000001
SMD SC	12302.35 (4103.29)	10672.29 (3504.29)	9889.113 (2663.29)	9890.017 (2662.79)
SMD AR(1)	10025.256 (3234.567)	9989.302 (3012.23)	8622.617 (2663.29)	8621.850 (2661.95)
SMD ARMA(1.1)	10054.534 (2997.72)	9867.561 (2900.32)	8169.547 (2488.562)	8168.995 (2488.322)
R.V (Tipo 1)	768.341 (172.3452)	727.021 (156.143)	602.0287 (112.6369)	602.2392 (112.4522)
R.V (Tipo 2)	380.654 (83.579)	372.986 (78.934)	334.0463 (71.85849)	334.0463 (71.80849)

TABLA 4.19. Varianzas Estimadas y sus Desviaciones Estándar Para Distintos Valores de los Hiperparámetros

La información recopilada en la tabla 4.19 muestra que a medida que se aumenta la cantidad de 0's en los hiperparámetros de la distribución propuesta en 3.8, las desviaciones estándar de las estimaciones disminuyen y los valores estimados se hacen más aproximados a los valores ajustados para la varianza mediante (MVR). Dado que las estimaciones y desviaciones estándar para los valores 0.000000001 y 0.0000000001 son bastante similares, se puede deducir que, a partir del valor 1×10^{-8} los cambios sobre las medias de las cadenas son demasiado pequeños y por esta razón, se usó en las 5 aplicaciones en cuestión el valor 1×10^{-8} como hiperparámetro.

Conclusiones

- La propuesta Bayesiana para el núcleo de transición 3.13 permite la estimación de los parámetros de correlación dentro del intervalo $[0, 1]$ sin transformar o ignorar valores que caigan fuera de este intervalo. Este tipo de kernel mostró tener mejores tasas de aceptación que la propuesta presentada por Hoff (2009).
- La implementación de las propuestas bayesianas, para la estimación conjunta de estructura de medias y matriz de varianzas y covarianzas para modelos longitudinales en los cuales se consideran estructuras de SC , $AR(1)$, $ARMA(1, 1)$, retorna los valores de las simulaciones realizadas. Hacer uso de aprioris informativas para los parámetros de correlación de estos modelos proporciona un ajuste más eficaz y mejora las tasas de aceptación del algoritmo MCMC.
- Al ajustar los datos del *SMD* mediante la metodología bayesiana propuesta para estructuras de covarianza SC , $AR(1)$, $ARMA(1, 1)$, se obtuvieron resultados aproximados a los obtenidos por Weiss (2005) y a las estimaciones obtenidas desde SPSS. Los criterios de comparación AIC y BIC, obtenidos desde el ajuste bayesiano son también próximos a los valores de los mismos criterios, al ajustar este conjunto de datos mediante la técnica de máxima verosimilitud restringida.
- El ajuste de los modelos propuestos en Zimmerman et al. (1998) y Núñez-antón & Woodworth (1994) sobre los datos de *reconocimiento de voz* en los cuales se considera una estructura $ADS(1)$ mediante la metodología bayesiana aquí propuesta, retornó resultados aproximados a los presentados por los autores originales, con tasas de aceptación razonables desde el punto de vista teórico. Se resalta que al observar la comparación de los valores obtenidos mediante la estimación clásica y bayesiana se encontró una mayor diferencia que las obtenidas sobre los datos de *SMD*. Estas diferencias se pueden atribuir al hecho de que el conjunto de datos de reconocimiento de voz es desbalanceado y para obtener la estimación mediante la propuesta bayesiana descrita en el capítulo 3 fue necesario completar las observaciones faltantes de manera aleatoria.
- Para realizar una propuesta bayesiana que ajuste los datos de la carrera de 100 km, se propone un modelo alternativo el cual proporciona menos restricción sobre los posibles valores de los parámetros que hacen parte de la matriz de varianzas y covarianzas. La metodología bayesiana aplicada sobre este modelo ajustó valores distintos a los obtenidos mediante la metodología clásica para los parámetros de la media β , aunque los intervalos de credibilidad del 95 % contengan los valores estimados mediante la metodología clásica; los valores estimados de los parámetros ρ

y λ son muy similares a los presentados en Núñez-Antón & Zimmerman (2001). Las estimaciones de σ_0^2 y $\boldsymbol{\psi} = (\boldsymbol{\psi}_1, \boldsymbol{\psi}_2)$ no son comparables de manera directa.

Para obtener algún punto de comparación es necesario calcular los valores de las varianzas de cada sección de la carrera σ_i^2 , $i = 1, \dots, 10$, al hacer esto se observa una gran diferencia entre los valores estimados mediante la metodología clásica y los obtenidos mediante el uso de nuestra propuesta, obteniendo una mayor proximidad a las varianzas muestrales al hacer uso del algoritmo que se postuló en el capítulo 3 para este modelo. No obstante, el análisis de residuales de los modelos no muestra diferencias grandes entre la metodología clásica y la bayesiana.

- La implementación de los algoritmos propuestos para la estimación bayesiana en estudios longitudinales con estructura $ADS(1)$ en una librería de R, se hace muy complicado ya que las funciones que estructuran la matriz de varianzas y covarianzas son muy distintas de un problema a otro.

Trabajo Futuro

- Generalizar los algoritmos presentados para la estimación bayesiana de modelos $AR(1)$ y $ARMA(1,1)$ para modelos de mayor orden es decir modelos $AR(p)$ y $ARMA(p, q)$.
- Hacer propuestas bayesianas para el ajuste de datos longitudinales desbalanceados que presenten estructuras como las trabajadas en este documento.
- Extender las propuesta bayesiana para el ajuste de datos longitudinales con estructuras de antedependencia estructuradas de mayor orden $ADS(s)$
- Implementar, en el paquete R que se desarrolló, las funciones y algoritmos para la estimación de modelos longitudinales con estructuras de antedependencia no estructuradas de orden 1.

Bibliografía

- Albert, J. (2009), *Bayesian computation with R*, Springer Science & Business Media.
- Brown, P. J., Kenward, M. G. & Bassett, E. E. (2001), ‘Bayesian discrimination with longitudinal data’, *Biostatistics* **2**(4), 417–432.
- Byrne, P. J. & Arnold, S. F. (1983), ‘Inference about multivariate means for a nonstationary autoregressive model’, *Journal of the American Statistical Association* **78**(384), 850–855.
- Cepeda Cuervo, E. (2011), ‘Generalized spatio-temporal models’, *SORT* **35**(2), 165–178.
- Cepeda Cuervo, E. & Achcar, J. A. (2009), ‘Regression models with heteroscedasticity using bayesian approach’, *Revista Colombiana de Estadística* **32**(2), 267–287.
- Cepeda-Cuervo, E. & Núñez-Antón, V. (2007), ‘Bayesian joint modelling of the mean and covariance structures for normal longitudinal data’, *SORT* **31**(2), 181–200.
- Cepeda-Cuervo, E. & Nunez-Anton, V. (2009), ‘Bayesian modelling of the mean and covariance matrix in normal nonlinear models’, *Journal of Statistical Computation and Simulation* **79**(6), 837–853.
- Cepeda, E. C. & Gamerman, D. (2004), ‘Bayesian modeling of joint regressions for the mean and covariance matrix’, *Biometrical journal* **46**(4), 430–440.
- Cepeda, E. & Gamerman, D. (2001), ‘Bayesian modeling of variance heterogeneity in normal regression models’, *Brazilian journal of probability and statistics* **14**(1), 207–221.
- Cuervo, E. C. (2001), ‘Modelagem da variabilidade em modelos lineares generalizados’, *Ph.D. thesis*.
- Daniels, M. J. & Pourahmadi, M. (2002), ‘Bayesian analysis of covariance matrices and dynamic models for longitudinal data’, *Biometrika* **89**(3), 553–566.
- Diggle, P. (2002), *Analysis of longitudinal data*, Oxford University Press.
- Everitt, B. (1994), ‘Exploring multivariate data graphically: a brief review with examples’, *Journal of applied Statistics* **21**(3), 63–94.
- Fitzmaurice, e. a. (2008), *Longitudinal data analysis*, CRC Press.

- Gabriel, K. (1962), ‘Ante-dependence analysis of an ordered set of variables’, *The Annals of Mathematical Statistics* pp. 201–212.
- Gamerman, D. (1997), ‘Sampling from the posterior distribution in generalized linear mixed models’, *Statistics and Computing* **7**(1), 57–68.
- Gamerman, D. & Lopes, H. F. (2006), *Markov chain Monte Carlo: stochastic simulation for Bayesian inference*, CRC Press.
- Gelman, A., Carlin, J. B., Stern, H. S. & Rubin, D. B. (2014), *Bayesian data analysis*, Vol. 2, Chapman & Hall/CRC Boca Raton, FL, USA.
- Harville, D. A. (1977), ‘Maximum likelihood approaches to variance component estimation and to related problems’, *Journal of the American Statistical Association* **72**(358), 320–338.
- Hoff, P. D. (2009), *A first course in Bayesian statistical methods*, Springer Science & Business Media.
- Hui, S. L. & Berger, J. O. (1983), ‘Empirical bayes estimation of rates in longitudinal studies’, *Journal of the American Statistical Association* **78**(384), 753–760.
- Izenman, A. J. & Williams, J. s. (1989), ‘A class of linear spectral models and analyses for the study of longitudinal data’, *Biometrics* pp. 831–849.
- Núñez-antón, V. & Woodworth, G. G. (1994), ‘Analysis of longitudinal data with unequally spaced observations and time-dependent correlated errors’, *Biometrics* pp. 445–456.
- Núñez-Antón, V. & Zimmerman, D. L. (2001), ‘Modelización de datos longitudinales con estructuras de covarianza no estacionarias: Modelos de coeficientes aleatorios frente a modelos alternativos’, *Qüestiió. 2001, vol. 25, núm. 2*.
- Patterson, H. D. & Thompson, R. (1971), ‘Recovery of inter-block information when block sizes are unequal’, *Biometrika* pp. 545–554.
- Potthoff, R. F. & Roy, S. (1964), ‘A generalized multivariate analysis of variance model useful especially for growth curve problems’, *Biometrika* pp. 313–326.
- Pourahmadi, M. (1999), ‘Joint mean-covariance models with applications to longitudinal data: Unconstrained parameterisation’, *Biometrika* pp. 677–690.
- Rao, C. R. (1975), ‘Simultaneous estimation of parameters in different linear models and applications to biometric problems’, *Biometrics* pp. 545–554.
- Tyler, e. a. (1988), ‘Evaluation of five different cochlear implant designs: audiologic assessment and predictors of performance’, *The Laryngoscope* **98**(10), 1100–1106.
- Verbeke, G. & Molenberghs, G. (2009), *Linear mixed models for longitudinal data*, Springer Science & Business Media.
- Ware, J. H. (1985), ‘Linear models for the analysis of longitudinal studies’, *The American Statistician* **39**(2), 95–101.
- Weiss, R. E. (2005), *Modeling longitudinal data*, Springer Science & Business Media.

-
- Zimmerman, D. L. & Núñez-Antón, V. A. (2009), *Antedependence models for longitudinal data*, CRC Press.
- Zimmerman, D. L., Núñez-antón, V. & El-Barmi, H. (1998), ‘Computational aspects of likelihood-based estimation of first-order antedependence models’, *Journal of Statistical Computation and Simulation* **60**(1), 67–84.
- Zimmerman, D. & Núñez-antón (1997), ‘Modelling Longitudinal and Spatially Correlated Data: Antedependence Models for Longitudinal Data’, *Lectures Notes in statistics* **12**(1), 63–76.