

**PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE**  
Facultad de Matemáticas  
Departamento de Estadística



**UNA PROPUESTA BAYESIANA PARA REALIZAR  
INFERENCIAS EN POBLACIONES  
HETEROCEDASTICAS  
CON OBSERVACIONES CORRELACIONADAS**

**TESIS PRESENTADA POR**

**Natalia Henríquez Cabezas**

**Como Requisito para  
obtener el grado de DOCTOR  
EN ESTADÍSTICA**

Profesor Guía: Reinaldo Arellano  
Profesor Co-tutor: Francisco Torres

**Noviembre de 2011.**

**Santiago. Chile.**

©2011 Natalia Henríquez. C.

**PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE**

**Facultad de Matemáticas  
Departamento de Estadística**

**Título de Tesis:** Una propuesta Bayesiana para realizar Inferencias en poblaciones heterocedásticas con observaciones correlacionadas.

**Autor:** Natalia Henríquez Cabezas.

**Fecha:** Noviembre de 2011.

Santiago. Chile.

**Jurado Externo 1:** \_\_\_\_\_

Marcia Branco.  
Universidad de Sao Paulo.

**Jurado Externo 2:** \_\_\_\_\_

Francisco Torres.  
Universidad de Santiago de Chile.

**Jurado Interno 1:** \_\_\_\_\_

Manuel Galea.  
P. Universidad Católica de Chile.

**Jurado Interno 2:** \_\_\_\_\_

Guido del Pino.  
P. Universidad Católica de Chile.

**Profesor Guía:** \_\_\_\_\_

Reinaldo Arellano.  
P. Universidad Católica de Chile.

*Dedicada a mi marido e hijos.*

*Natalia*

# Índice general

<b>AGRADECIMIENTOS</b>	<b>2</b>
<b><i>RESUMEN</i></b>	<b>4</b>
<b>1. Marco Teórico</b>	<b>6</b>
1.1. Introducción . . . . .	6
1.2. Enfoques Estadísticos . . . . .	11
1.3. Espacio Muestral y Paramétrico . . . . .	12
1.3.1. Permutabilidad . . . . .	13
1.4. Modelo Bayesiano . . . . .	13
1.4.1. Consistencia . . . . .	15
1.4.2. Distribuciones a Priori . . . . .	19
1.4.3. Distribuciones No Informativas . . . . .	19
1.4.4. Especificación Jerárquica a las distribuciones a Priori . . . . .	20
1.4.5. Métodos Computacionales . . . . .	22
1.5. Comparación de Modelos . . . . .	24
1.6. Objetivos de la Tesis . . . . .	27
1.6.1. Estructura de la tesis . . . . .	28
<b>2. Modelos Bayesianos Bajo Independencia</b>	<b>30</b>
2.1. Introducción . . . . .	30
2.2. Notación . . . . .	33
2.3. Modelo normal heterocedástico con varianzas conocidas . . . . .	34
2.4. Modelo normal homocedástico con varianza desconocida . . . . .	36
2.5. Modelo normal heterocedástico con varianzas desconocidas . . . . .	37
2.6. Modelo normal homocedástico con parámetros para la media aleatorios . . . . .	38
2.7. Modelo normal heterocedástico con variable auxiliar . . . . .	39

2.8. Modelo t-Student homocedástico . . . . .	40
2.9. Modelo t-Student heterocedástico . . . . .	42
2.10. Modelo slash homocedástico . . . . .	43
2.11. Modelo slash heterocedástico . . . . .	45
<b>3. Modelo Bayesiano Correlacionado</b>	<b>47</b>
3.1. Introducción . . . . .	47
3.2. Modelo normal heterocedástico . . . . .	50
3.3. Modelo t-Student heterocedástico . . . . .	52
3.4. Modelo slash heterocedástico . . . . .	53
<b>4. Simulaciones</b>	<b>56</b>
4.1. Modelos Independientes . . . . .	59
4.1.1. Estimaciones . . . . .	60
4.1.2. Densidades a Posteriori . . . . .	63
4.1.3. Calibración del procedimiento . . . . .	65
4.2. Modelos Correlacionados . . . . .	66
4.2.1. Estimaciones . . . . .	66
4.2.2. Densidades a Posteriori . . . . .	70
4.2.3. Calibración del procedimiento . . . . .	71
<b>Notas Finales</b>	<b>73</b>
Conclusiones Generales . . . . .	73
Proyecciones Futuras . . . . .	74
<b>Anexo</b>	<b>84</b>
A.1 Distribuciones Condicionales . . . . .	90
A.1.1 Modelos Independientes . . . . .	90
A.1.2 Modelos Correlacionados . . . . .	90
A.2 Tablas y Gráficos de convergencia . . . . .	90
A.2.1 Modelos Independientes . . . . .	90
A.2.2 Modelos Correlacionados . . . . .	90

## AGRADECIMIENTOS

## RESUMEN

*Es usual realizar inferencia con observaciones no correlacionadas, no obstante, en la vida real podemos encontrar datos que al parecer no poseen esta característica, por ejemplo, datos sísmicos, de esta manera es necesario investigar que procedimientos estadísticos pudiesen ayudar en esta idea y determinar si existe algún tipo de correlación significativa o no entre dichas observaciones.*

*El presente trabajo, aborda esta idea desde el punto de vista del enfoque Bayesiano, dando un procedimiento para establecer la existencia de correlación entre observaciones de subpoblaciones heterocedásticas, las cuales entre ellas están no correlacionadas.*

*Además nuestros datos podrían no ser modelados a través de una distribución normal, que es lo estándar, por lo cual los procedimientos que esta tesis presenta, son trabajados con los modelos normal, t-Student y slash. Esto se realizó con la incorporación de una variable aleatoria auxiliar  $u$ , al modelo de los datos bajo el esquema Bayesiano, de tal manera, que de acuerdo al modelo de probabilidad asociado a ella permite trabajar con mezcla de distribuciones normales, flexibilizando el modelamiento de estos, y trabajar con distribuciones de tipo t-Student's y slash. El artículo de Kenneth Lange and Janet Sinsheimer(1993)[37] abordan estos modelos y otros más analizando algunas propiedades. La inferencia se focaliza en la estimación de parámetros de medias, varianzas y covarianzas, de los estimadores de máxima verosimilitud(estimador de Bayes)*

*Los modelos mencionados, se abordan en el capítulo 2 bajo el supuesto de no correlación entre observaciones y entre subpoblaciones y en el capítulo 3 bajo una estructura de correlación al interior de cada subpoblación y no correlación fuera de ellas, obteniéndose las distribuciones condicionales completas en cada caso. En el capítulo 4 se realizan las simulaciones de los modelos abordados en los capítulos 2 y 3, para esto en el modelo slash hubo que incorporar una truncación estándar y para determinar la devianza de ésta, se utilizó una aproximaciones asintótica en el cálculo de la integral.*

*Para las estimaciones se ocuparon los estimadores de Bayes, y para determinar la calidad del ajuste de los modelos se utilizaron los criterios del DIC,  $pd$  y BIC. La convergencia de las cadenas se analizaron con apoyo del paquete BOA, de R-project versión 2.6, utilizando los test de Raftery and Lewis,*

*Heidleberger and Welch y Geweke. Todo esto se presenta en el capítulo 4 de simulaciones.*

*Además en el capítulo 1 se presenta el marco teórico, el que introduce los conceptos en forma resumida que se utilizaron en esta tesis.*



# Capítulo 1

## Marco Teórico

### 1.1. Introducción

*La motivación principal de este trabajo es la de realizar inferencia en super poblaciones finitas o infinitas, no existiendo correlación entre ellas, y con la característica de poseer una estructura de correlación al interior de éstas.*

*Por ejemplo, datos como mediciones sísmicas se consideran que están correlacionados por la zona (subpoblación), y datos relacionados con el procesamiento de imágenes, donde la gama de colores y el color suponen una incidencia en la calidad de la imagen hacen suponer una correlación entre las mediciones de estos. No obstante por sus características es natural trabajarlos como modelos jerárquicos o modelos en multiniveles.*

*Los métodos más utilizados en el análisis de variables numéricas continuas están en su mayoría diseñados en las que se registra una única medida por cada unidad de observación, por ejemplo en análisis de varianza y regresión. Es poco usual poseer réplicas de mediciones de un mismo individuo y además la misma cantidad para todos los individuos observados, que es el caso cuando se trabaja con datos longitudinales (temporales) [Cannon](2001)[10] o también en el caso de diseños balanceados [Cnaan](1997)[11]; no obstante*

podemos visualizar una cierta correlación entre las mediciones del individuo, por lo cual no pueden considerarse como observaciones independientes, supuesto básico para estimar un modelo de regresión clásico. Otra clase de estudios con observaciones correlacionadas son los datos agrupados, en los que existe un diseño jerárquico, por ejemplo datos agrupados en hospitales, puede ocurrir que los pacientes que se atienden en las distintas unidades tengan ciertas características muy similares entre ellos, es decir nuevamente bajo ciertos factores las observaciones no son independientes [Leary](2000)[40].

Estudios de crecimiento, en los que el sujeto se evalúa en diferentes edades o momentos, se consideran como investigaciones con datos correlacionados.

Modelos para el análisis de respuestas categóricas multivariadas correlacionadas han sido utilizados para medir la calidad de la docencia[Casero](2010)[12], en los cuales se visualizan multiniveles (al menos tres), este tipo de mediciones podrían estar correlacionadas de acuerdo a la apreciación que los estudiantes poseen a priori del profesor, y esta influir en la evaluación a posteriori del docente.

Además autores[Ita et al](1998)[33], hacen una introducción sobre el modelamiento con multiniveles en presencia de datos(encuestas) correlacionados y no correlacionados[Arijit et al](1982)[3], poniendo énfasis en estimadores sesgados e insesgados.

Esta teoría tiene como propósito flexibilizar el modelamiento bajo supuestos de independencia y/o idéntica distribución, en conjunto con el apoyo de los avances tecnológicos, la rapidez de procesamiento y generación de algoritmos estadísticos, es que ha permitido que variadas investigaciones se centren y se sigan desarrollando en el análisis y modelamiento de éstos, desde un enfoque paramétrico [Ortega-Basulto](2009)[48], [Bush](1996)[9], como no paramétrico[Casanova](2007)[13],[Orellana](2002)[49],[Ranganathan](2004)[2].

Diversos trabajos se han realizado bajo el contexto de familias paramétricas que contengan una amplia gama de distribuciones, las cuales posean algunas características como por ejemplo, simetría o asimetría, colas más o menos pesadas [Thomson](2005)[59], de tal manera de poder extender el modelamiento y tener algún grado de flexibilidad. Es así que este trabajo adoptará alguna de ellas y presentará algunos modelos bajo una estructura de correlación de manera que datos con las características mencionadas anteriormente puedan ser modelados.

Un concepto más general a la independencia de variables aleatorias es el de Permutabilidad, el cual establece que una sucesión de variables aleatorias se dirá permutable si: "toda permutación de  $n$  de ellas posee la misma distribución de probabilidad conjunta"; ambos conceptos son importantes de destacar y mencionar, a la vez indicar de que manera influyen en el análisis de los procedimientos (distribución de probabilidad a priori) y de los resultados obtenidos cuando se trabaja con muestras. Dichos conceptos se explican con mayor detalle más adelante. Investigaciones y resultados establecidos por de Finetti(1937)[16], Ericson (1969)[20], Draper et al. (1993)[17], entre otros en materias de Permutabilidad, permiten dar fluidez y respaldo a diferentes resultados en los que se relaciona teoría de muestreo y distribuciones asociadas.

En diferentes investigaciones desarrolladas en poblaciones finitas, es usual que se focalicen en estimaciones de medias, percentiles, varianzas, entre otras; todas estas desde el punto de vista de modelos lineales,[Bolfarine et al](1992)[7], [Bellhouse](1987)[6] las cuales presentan resultados de estimadores sesgados e insesgados; dichas investigaciones contemplan resultados equivalentes entre muestreo aleatorio simple y la noción de variables aleatorias permutables, lo cual nos permite realizar inferencia a partir de un buen diseño de muestreo [Ericson](1969)[20] y extender dichos resultados a poblaciones infinitas .

*En variados estudios las unidades de muestreo están agrupadas de forma natural tales como: colegios, empresas, hospitales u otras instituciones, es así que existen diversos sistemas que poseen una estructura de anidamiento, en los cuales las unidades de observación de un determinado nivel, se hallan agrupadas en otras de un nivel jerárquico superior. Por ejemplo, los estudiantes de un centro educacional agrupados en clases; estos pueden constituir el nivel inferior de la jerarquía, las clases el nivel intermedio y los centros educacionales el nivel superior.*

*Estos agrupamientos, además de representar una estructura específica de datos, constituyen unidades de análisis que tienen interés por sí mismas. Esto es, debido a la importancia que tiene considerar las variables medidas en los distintos niveles de la jerarquía es que permite entender la relevancia de los participantes en el estudio. De ahí, que el término multinivel [Goldstein](2002)[28] se utilice tanto para designar los niveles en que se pueden agrupar los datos y para caracterizar a las formas o técnicas utilizadas para modelar relaciones que se dan dentro y a través de los diferentes niveles.*

*Dicha estructura de jerarquización puede ser analizada utilizando modelos, ya sea desde el punto de vista clásico como Bayesiano, es en este último en el que se trabajará para realizar la estimaciones en poblaciones infinitas, bajo normalidad u otras distribuciones.*

*Por otra parte algunos autores [Ericson](1969)[20], dejan claro que el modelo básico en poblaciones finitas debe establecer dos hechos importantes: primero, que la información contenida en la verosimilitud de los datos (muestra cualquiera) está libre de toda subjetividad, y segundo, que la subjetividad sólo ingresa a través de la elección de la distribución de probabilidad a priori.*

*Prácticamente hasta la década de los 80, la construcción de modelos conjugados y el análisis asintótico, constituían las dos únicas formas de análisis*

Bayesiano. Así, junto con los desarrollos conceptuales en la formulación de modelos y el aumento de la tecnología computacional, se ha dado un nuevo impulso a los modelos bayesianos, cuya versatilidad para modelar situaciones más complejas no habían sido explotadas debido a resolución de procesos analíticos imposibles de resolver.

En los tiempos actuales se han implementado métodos de integración y de simulación; luego el marco Bayesiano parece ser el método a seguir para el análisis de modelos de este tipo, además el auge de las técnicas y/o métodos de simulación llamados en su generalización métodos MCMC, es decir, Métodos Montecarlo con Cadenas de Markov, se puede lograr en forma relativamente simple obtener muestras de una distribución objetivo que junto con los trabajos de Metropolis, permite dar fluidez y eficacia a la obtención de muestras de distribuciones a posteriori complejas, sobre todo cuando se trabaja con alta dimensionalidad.

Podemos encontrar algunas investigaciones de modelamiento, basadas en datos correlacionados en poblaciones finitas desde el punto de vista Bayesiano paramétrico, en el cual la variable de interés es continua, aplicando modelos normales y obteniéndose estimaciones usuales [Scott and Smith](1969)[55], como también, desde el punto de vista semiparamétrico se puede mencionar el trabajo desarrollado por Jara et al. (2006)[34], quienes abordaron distribuciones tipo mezcla de procesos Dirichlet en el análisis de respuestas binarias correlacionadas.

Otros estudios de modelamiento de respuestas binarias con mayor flexibilidad, se dan utilizando BRMUW (Bayesian Regression Model using WinBUGS)[Bazán](2010)[4]. Datos financieros [Mendoza](2006)[43], donde el interés se centra en pronosticar el comportamiento futuro de las pérdidas, son analizados bajo una estructura de correlación utilizando cuantiles bajo un enfoque Bayesiano y aplicando mezcla de normales. En el libro de Edward

W. Frees (2004)[21] se puede encontrar análisis de datos longitudinales, de panel y otros.

El objetivo de este capítulo es presentar el marco general, definiciones y propiedades necesarias para el desarrollo de este trabajo, bajo un enfoque Bayesiano Paramétrico.

En la sección 1.2 de este capítulo, se realiza una descripción de los distintos enfoques estadísticos, en la sección 1.3 se plantean los aspectos fundamentales de la Estadística Inferencial desde el punto de vista Bayesiano Paramétrico, dando un esbozo de la estructura que posee el diseño muestral y de que manera influye en nuestra forma de trabajar, ya sea en los procedimientos y/o de los resultados obtenidos, destacando los modelos abordados por Lindley y Smith (1972)[41]; en la tercera sección se mencionan algunas características de los modelos lineales y de que manera fueron abordados en este trabajo, en la cuarta sección se explican los procedimientos computacionales que se ocuparon y en la quinta sección se plantean algunos métodos de uso común de selección de modelos.

La sexta sección es dedicada al plan general de esta tesis, incluyendo sus objetivos.

## 1.2. Enfoques Estadísticos

El Objetivo del análisis estadístico es la inferencia, que intuitivamente se puede definir como el proceso de aproximar o estimar qué  $P \in \mathcal{P}$  genera los datos o bien, verificar o refutar alguna hipótesis acerca de la verdadera medida  $P$ .

Dependiendo de la naturaleza de la familia de medidas de probabilidad  $\mathcal{P}$ , es posible distinguir tres enfoques en la inferencia: el enfoque paramétrico, el no-paramétrico y el semi-paramétrico.

La inferencia paramétrica considera que cada medida de probabilidad de la familia, está indexada por un parámetro real  $\theta$ , escalar o vectorial. En tal caso escribimos:  $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ , donde  $\Theta$  es el llamado espacio de parámetros, el cual es finito-dimensional. Es usual suponer, en este caso, que a cada medida de probabilidad se le puede asociar una función de densidad (o cuantía), indexada por  $\theta$ . De esta forma surgen los diferentes modelos estadísticos paramétricos tales como los modelos Binomial, Poisson, Exponencial, Normal, *t*-Student, Slash, etc.

La inferencia no-paramétrica se refiere al caso cuando no se supone forma alguna para las medidas de probabilidad de la familia  $\mathcal{P}$ , en tal caso puede considerarse como conjunto de índices al conjunto de todas las funciones de distribución  $\mathcal{F}$  de interés, el cual es usual denotarlo por  $\wp = \{P_F : F \in \mathcal{F}\}$ , así el espacio de parámetros es infinito-dimensional.

Finalmente, se ha adoptado en llamar modelos estadísticos semi-paramétricos, a aquellos modelos que toman en consideración, tanto parámetros finito dimensionales, como infinitos dimensionales, considerando habitualmente una estructura jerárquica en su definición.

### 1.3. Espacio Muestral y Paramétrico

En esta etapa se presentan un conjunto de definiciones y resultados, que permitirá dar confiabilidad, solidez y fluidez a los resultados obtenidos, como por ejemplo ¿qué condiciones debe cumplir el espacio paramétrico para que una medida de probabilidad posea soporte?, ¿qué nos garantiza la existencia de la distribución a posteriori?, etc.

### 1.3.1. Permutabilidad

**Definición 1.3.1.** *Se dice que una secuencia de variables aleatorias son permutables si para toda permutación de  $n$  de ellas, la distribución de probabilidad conjunta es la misma.*

De Finetti (1935)[16] demostró, que una secuencia de variables aleatorias binarias son permutables, si y sólo si existe una función de distribución  $F$  o  $P$  sobre  $[0, 1]$ , tal que,  $\forall n \in \mathbb{N}$ ,

$$p(\mathbf{y}_n) = p(y_1, y_2, \dots, y_n) = \int_0^1 \theta^{t_n} \cdot (1 - \theta)^{n-t_n} dF(\theta), \text{ con } t_n = \sum y_i.$$

Hewitt - Savage (1955)[31] generalizaron este resultado de de Finetti a cualquier tipo de variables. No obstante, cabe mencionar que existen algunos estudios que permiten dar algunas caracterizaciones a los modelos Uniformes bajo el contexto de la permutabilidad [Esteves, Wechsler, Iglesias](2004)[19].

Dentro de este contexto podemos considerar muestras cualesquiera de tamaño  $n$  de unidades distintas, para realizar los procedimientos y obtener los modelos en cuestión.

Por lo expresado anteriormente denotaremos por  $\Theta$  al espacio paramétrico donde se asumirá que  $\Theta$  es un espacio métrico completo y separable con la  $\sigma$ -álgebra de Borel  $\mathcal{B}(\Theta)$ .

## 1.4. Modelo Bayesiano

Los modelos que se desarrollaron en esta tesis son modelos multivariantes con una estructura jerárquica Bayesiana, es así que se mencionarán algunas características y conceptos utilizados en ellos. No obstante en la partida, se presentan algunos fundamentos estadísticos que se emplearán como: modelo



Bayesiano,  $\sigma$ -álgebras asociadas, consistencia, modelos jerárquicos, distribuciones a Priori, distribuciones no informativas, etc.

### **Modelo Bayesiano**

**Definición 1.4.1.** *Un modelo estadístico es un trío formado por:  $(\mathcal{Y}, \mathcal{A}, \mathcal{P}_\theta)$  donde  $\mathcal{Y}$ , es el espacio de posibles observaciones,  $\mathcal{A}$  es la  $\sigma$ -álgebra asociada a  $\mathcal{Y}$  y  $\mathcal{P}_\theta$  es una familia de posibles medidas de probabilidad indexadas por un parámetro  $\theta$  y definidas sobre  $(\mathcal{Y}, \mathcal{A}) = \Omega$ .*

Una discusión de las  $\sigma$ -álgebras generadas por las distintas topologías asociadas puede ser encontrada en [Gaudard](1989)[22].

**Definición 1.4.2.** *El modelo Bayesiano en su forma más sencilla consiste de un parámetro o vector de parámetros desconocido  $\theta$  aleatorio, con  $\theta \in \Theta$ , donde  $\Theta$  es el espacio paramétrico de valores para  $\theta$ , el cual también tiene asociado una  $\sigma$ -álgebra y una distribución a priori  $\Pi$  de  $\Theta$ , perteneciente a alguna familia de distribuciones posibles, con una medida de probabilidad asociada, dada por el conocimiento o experiencia del experto.*

En términos prácticos, sea  $\mathbf{y} = (y_1, y_2, \dots, y_n)'$  un vector de  $n$  observaciones donde  $\mathbf{y} \sim P_\theta$ , la cual se interpreta como la distribución condicional de  $\mathbf{y}$  dado que el parámetro aleatorio toma el valor particular  $\theta$ , a su vez cada componente sigue una distribución de probabilidad  $p(y|\theta)$ , que depende de  $j$  parámetros que conforman el vector  $\theta = (\theta_1, \theta_2, \dots, \theta_j)'$ , el que asume una distribución de probabilidad  $\Pi(\theta)$  (distribución de probabilidad a priori), con estos elementos se pueden obtener:

i) la distribución de probabilidad conjunta de  $\theta$  e  $y$ :

$$\pi(y, \theta) = p(y|\theta) \cdot \pi(\theta) = \pi(\theta|y) \cdot p(y),$$

ii) la distribución de probabilidad marginal de  $y$ :

$$p(y) = \int \pi(y, \theta) d\theta = \int p(y|\theta) \cdot \pi(\theta) d\theta,$$

iii) la distribución de probabilidad a posteriori:

$$\pi(\theta|y) = \frac{p(y|\theta) \cdot \pi(\theta)}{p(y)}, \quad p(y) \neq 0.$$

Esta última ecuación es obtenida de forma sencilla a través del Teorema de Bayes.

De estas tres distribuciones de probabilidad, la herramienta principal del análisis Bayesiano es la tercera, la que permite realizar inferencia sobre  $\theta$ , con información de las observaciones resumidas a través de la función de verosimilitud  $p(y|\theta)$  o escrita de otra manera como  $L(\theta|y)$  [Robert](1994)[52], [Stigler](1986)[58]. Claro está, que para llegar a la distribución de probabilidad a posteriori, hubo que escoger y argumentar del porque de la distribución a priori utilizada; es aquí donde el análisis Bayesiano es cuestionado, aunque no se puede negar que su presencia resuelve muchos de los problemas y limitaciones presentes en el análisis clásico.

### 1.4.1. Consistencia

Además desde el punto de vista Bayesiano, si una muestra aleatoria de tamaño  $n$  es extraída de una población de tamaño  $N$ , se tiene,

$$p(y_{i(1)}, y_{i(2)}, \dots, y_{i(n)}|\mathbf{y}) = p(s) = 1/\binom{N}{n}.$$

y si se asume distribuciones a priori permutables, tendremos,

$$p(y_1, y_2, \dots, y_n) = p(y_{i(1)}, y_{i(2)}, \dots, y_{i(n)}), \quad \forall i \text{ permutación.}$$

Como la probabilidad del muestreo  $p(s)$ , no depende de los parámetros, la distribución de probabilidad a posteriori tampoco depende de ésta, además

la distribución a priori conjunta es de la forma  $\int \prod_{i=1}^N p(y_i|\theta) dF(\theta)$ , se tiene que, la distribución a posteriori es proporcional a,

$$\int_{\theta} \prod_{i \notin s} \prod_{i \in s} p(y_i|\theta) dF(\theta).$$

Por lo tanto bajo los argumentos de Permutabilidad se adopta como notación para una muestra aleatoria de tamaño  $n$ , a  $\mathbf{y}_n = (y_1, y_2, \dots, y_n)$ .

A través de la información del diseño, que esencialmente es una distribución generada por la aleatorización, hace posible evaluar los procedimientos de inferencia como estimación puntual y por intervalo.

La inferencia estadística en poblaciones finitas, se formula a través de un resultado central el cual manifiesta que no existe un estimador lineal insesgado y de varianza mínima uniformemente sobre el espacio muestral. El advenimiento de los modelos superpoblacionales, se da a raíz de una postulación formal como parte de los procesos de inferencia para poblaciones finitas.

Investigadores al respecto [Smith](1976,1984)[56],[Hartley](1975)[29], dan los principios y acotan las líneas de desarrollo de la teoría de muestreo bajo modelos superpoblacionales, en los cuales se asume que cualquier muestra de una población finita se puede considerar como una sub muestra de una superpoblación, es decir la población finita pasa a ser una muestra. Así las estimaciones que se realizan con la submuestra, se ocuparán para las predicciones en los valores faltantes para completar la muestra(población finita).

## Distribución a Posteriori y Consistencia

Supongamos que  $\Pi$  es una distribución a priori de  $\theta$ , es decir es una medida de probabilidad sobre  $(\Theta, \mathcal{B}(\Theta))$ , luego para cada  $n$ ,  $\Pi$  y las  $P_\theta$  's, definen una distribución conjunta de  $\theta$  e  $\mathbf{y}_n = (y_1, y_2, \dots, y_n)$ , la que se simboliza y se define como:  $\lambda_{n,\pi}(B, A) = \int_B P_\theta^n(A) d\Pi(\theta)$ , y cuya distribución marginal  $\lambda_n$  de  $\mathbf{y}_n = (y_1, y_2, \dots, y_n)$  es:

$$\lambda_n(A) = \lambda_{n,\pi}(\Theta, A).$$

**Definición 1.4.3.** La distribución condicional de  $\theta$  dado  $y_1, y_2, \dots, y_n$  es llamada la distribución a posteriori la cual se define como:

$\Pi(\cdot|\cdot) : \mathcal{B}(\Theta) \times \Omega_n \longrightarrow [0, 1]$  y se debe cumplir:

- (a) Para cada  $\omega \in \Omega_n$ ,  $\Pi(\cdot|\omega)$  es una medida de probabilidad sobre  $\mathcal{B}(\Theta)$ .
- (b) Para cada  $B \in \mathcal{B}(\Theta)$ ,  $\Pi(B|\cdot)$  es  $\mathcal{A}^n$  medible.

y

- (c) Para cada  $B \in \mathcal{B}(\Theta)$  y  $A \in \mathcal{A}^n$ ,  $\lambda_{n,\pi}(B \times A) = \int_A \Pi_{B|\omega} d\lambda_n(\omega)$

En el caso de espacios completos y separables la posteriori siempre existe y  $\Pi(\cdot|\omega)$ , es una función de  $y_1, y_2, \dots, y_n$ , así, la posteriori se denotará como:

$$\Pi(\cdot|y_1, y_2, \dots, y_n) \text{ o } \Pi(\cdot|\mathbf{y}_n),$$

y si las  $P_\theta$  's son todas dominadas por una medida  $\sigma$ -finita  $\mu$ , es decir,  $p_\theta = dP_\theta/d\mu$  entonces su función de densidad está dada por la expresión:

$$\frac{\prod_1^n p_\theta(y_i) \cdot \pi(\theta)}{\int_\Theta \prod_1^n p_\theta(y_i) \cdot \pi(\theta)}.$$

**Definición 1.4.4.** Para cada  $n$ , la secuencia  $\{\Pi(\cdot|\mathbf{y}_n)\}$  se dice consistente para  $\theta_0$ , si existe  $\Omega_0 \subset \Omega$  con  $P_{\theta_0}^\infty(\Omega_0) = 1$ , tal que si  $\omega \in \Omega_0$  entonces para cada vecindad  $U$  de  $\theta_0$ ,  $\{\Pi(\cdot|\mathbf{y}_n)\} \xrightarrow{D} \delta_{\theta_0}$  a.e  $P_{\theta_0}$ .

Así la posteriori es consistente para  $\theta_0$ , si con probabilidad 1, la posteriori se concentra alrededor de  $\theta_0$ .

Además si se tienen varias posibilidades de posteriori consistentes, el siguiente teorema nos permite trabajar con cualquiera de ellas para realizar las estimaciones respectiva.

**Teorema 1.4.1.** Supongamos que la familia  $P_\theta$ ,  $\theta \in \Theta$  es dominada por una medida  $\sigma$ -finita  $\mu$  y sea  $p_\theta$  la densidad de  $P_\theta$  y  $\theta_0$  un punto interior de  $\Theta$ , además sean  $\pi_1$  y  $\pi_2$  dos densidades con respecto a una medida  $\nu$ , las cuales son positivas y continuas para  $\theta_0$ .

Si  $\pi_i(\cdot|\mathbf{y}_n)$ ,  $i=1,2$  son ambas consistentes para  $\theta_0$  entonces

$$\lim_{n \rightarrow \infty} \int |\pi_1(\theta|\mathbf{y}_n) - \pi_2(\theta|\mathbf{y}_n)| d\nu(\theta) = 0 \text{ a.e } P_{\theta_0}.$$

En particular Aggarwal (1959,1966)[1], usó prioris Normales para la aplicación de Mínimax para obtener estimadores en poblaciones finitas.

Hill(1968)[32], consideró distribuciones a posteriori de medias y percentiles de poblaciones finitas e infinitas y posteriormente autores como, Cochran (1963)[15], Roderick (2006)[53] y Godambe(1955)[27], han utilizado distribuciones a priori en teoría de muestreo, en tal caso, el objetivo principal fue optimizar la estrategia del diseño y no enfocado en la utilidad de la distribución a posteriori.

### 1.4.2. Distribuciones a Priori

Los modelos Bayesianos requieren como se mencionó anteriormente de la especificación de una distribución a priori sobre el o los parámetros, este proceso puede resultar complicado en caso de múltiples parámetros, en particular en modelos de regresión. Diversas investigaciones y discusiones se basan en la búsqueda de ellas [Jeffreys](1939) [Raiffa y Schaifer](1961) [Morris](1983)[46].

Usualmente por conveniencia son generalmente usadas las prioris conjugadas (tanto distribución a priori como distribuciones condicionadas a posteriori y en el mejor de los casos la distribución a posteriori pertenecen a la misma familia), ya que en su mayoría, son analíticamente tratables, Morris (1983)[46] muestra que para las familias exponenciales, es posible escoger distribuciones a priori de estas características, las cuales por cierto no están exentas de subjetivismo por ejemplo parámetros subyacentes o hiperparámetros, pero de igual manera las hace muy adecuadas en la utilización de técnicas MCMC.

Hay elecciones de prioris completamente no subjetivas, las cuales también tienen su utilidad inferencial y de comparación, como son las distribuciones a priori No informativas.

### 1.4.3. Distribuciones No Informativas

En caso de que no se disponga ningún tipo de información respecto del parámetro de estudio, existen las distribuciones No informativas y su uso se remonta a la época de Laplace utilizando las distribuciones Uniformes. Críticas al respecto es que en algunos casos resultan distribuciones a posteriori impropias y que tales distribuciones a priori no son invariantes bajo reparametrizaciones [Robert](1992)[52]. Jeffreys(1961)[35] propone una aproximación en la que evita el tener que considerar la invarianza, incluso aunque a veces sea compatible con ella. Esta aproximación consiste en tomar una den-

sidad a priori proporcional a la raíz cuadrada de la matriz de información de Fisher; esta resulta ser invariante bajo transformaciones, es fácil de calcular en el caso univariado, Ibrahim y Laud (1995)[39] hacen un análisis de ésta en los parámetros de una regresión.

La distribución a priori como se indicó anteriormente, es una medida de probabilidad sobre un espacio paramétrico  $\Theta$ , además, la distribución a posteriori depende de las observaciones de  $y_1, y_2, \dots, y_n$ , en el caso que "n" tienda a infinito, ésta se relaciona directamente con la teoría de convergencia de medidas de probabilidad. Autores como Ghosh and Ramamoorthy (2003)[26], hacen una revisión extensa de este tema.

#### 1.4.4. Especificación Jerárquica a las distribuciones a Priori

Los modelos jerárquicos bayesianos, proporcionan una forma particular de modelar la información a priori, descomponiéndola en distintos niveles, lo que permite diferenciar entre elementos estructurales y subjetivos de esta información.

Una forma simple y útil de visualizar el modelo Bayesiano es en dos etapas o jerarquías, la primera corresponde al modelo observacional o verosimilitud y la segunda a la distribución a priori, es decir:

$$y|\theta \sim p(y|\theta). \tag{1.4.1}$$

$$\theta \sim \pi(\theta). \tag{1.4.2}$$

Generalmente, la distribución de la etapa (1.1). es una distribución conjugada, por la facilidad que tiene para ser tratada computacionalmente y porque en cierta manera, las etapas superiores compensan o ayudan en posibles errores de modelado en los niveles inferiores (da robustez a los estimadores), y/o

en los procesos computacionales. Pueden encontrarse numerosos ejemplos de modelos jerárquicos en las ciencias sociales [Draper](1995)[17], salud [Monsalve et al.](2010)[46], agricultura [Bazán et al.](2010)[4].

Aquí utilizaremos la notación  $y|\theta \sim f(y|\theta)$  para decir que la función de densidad condicional de  $y$  dado  $\theta$  es  $f(y|\theta)$ .

**Por ejemplo:**

Consideremos el modelo lineal jerárquico, que fue introducido por Lindley y Smith(1972)[41]; el cual establece que:

$$\begin{aligned} \mathbf{y}|\beta_1, \phi &\sim N_n(X_1\beta_1, \phi^{-1}I_n) \\ \beta_1|\beta_2 &\sim N_{k_1}(X_2\beta_2, C) \\ \beta_2 &\sim N_{k_2}(b, B) \\ \phi &\sim \text{Gamma}(n_0/2, n_0\sigma_0^2/2). \end{aligned}$$

Este modelo fue reformulado posteriormente por Mac Eachern (1994)[42], West y Turner (1994)[61], Escobar y West (1995)[18] y Müller, Erkanli y West(1996)[47], considerando componentes no paramétricas en la jerarquización, dándole un enfoque Semiparamétricos al modelamiento.

Es así que la distribución a priori  $\pi(\theta)$ , se puede descomponer en distribuciones condicionales de acuerdo a distintas etapas o niveles superiores como:

$$\pi(\theta|\theta_1), \pi(\theta_1|\theta_2), \dots, \pi(\theta_{n-1}|\theta_n) \text{ y la distribución marginal } \pi_{n+1}(\theta_n).$$

de tal forma que:

$$\pi(\theta) = \int_{\Theta_1 \times \dots \times \Theta_n} \pi(\theta|\theta_1), \pi(\theta_1|\theta_2), \dots, \pi_{n+1}(\theta_n) d\theta_1 \dots d\theta_n$$



(a los parámetros  $\theta_i$  se les denomina hiperparámetros de nivel  $i$ -ésimo)

Los modelos jerárquicos son preferidos por diferentes razones, entre ellas están:

1. En análisis de metapoblaciones, es decir en estudios de aquellas poblaciones que a su vez se subdividen en subpoblaciones.
2. En el ámbito no informativo, los modelos jerárquicos le dan robustez a las estimaciones [Berger](1985)[5] y en muchos casos se trabaja con distribuciones no informativas relacionadas con distribuciones conjugadas.
3. Un tercer caso, la formulación jerarquizada puede flexibilizar el procedimiento computacional.

### 1.4.5. Métodos Computacionales

La inferencia en los modelos Bayesianos radica sustancialmente en la información que la distribución a posteriori nos entregue, entre ellas podemos mencionar: medidas de tendencia central, medidas de variabilidad, intervalos de credibilidad, distribuciones marginales, factores de Bayes, etc; dichos resultados involucran cálculo de Integrales que tienen la misma dimensión de los parámetros a estimar, con lo cual la dificultad para obtener los resultados se vuelve compleja; es así que se han desarrollado mecanismos para simplificar estos procesos. Entre ellos destacaremos el algoritmo de Gibbs Sampling y el de Metropolis Hasting (casos particulares de MCMC).

Los Método Montecarlo con Cadenas de Markov (MCMC), propuesto por Geman y Geman (1984)[24] y difundido por Gelfand y Smith (1990)[23], tiene como objetivo trabajar con muestras pseudoaleatorias de distribuciones de probabilidad que cumplan con ciertas condiciones como el de obtener una

distribución estacionaria, es decir, la idea es garantizar que para  $n$  grande la distribución de los datos converge; autores como Hasting (1970)[30], Ross (1995), Geyer (1992)[25], Müller(1991), Chib y Greenberg (1995)[14], han estudiado y dado las condiciones para que esta distribución requerida exista. Así dado el parámetro  $\theta = (\theta_1, \dots, \theta_p)$ , el muestreador de Gibbs trabaja en forma iterativa, es decir, para cada  $j = 1, 2, \dots, p$  genera las distribuciones condicionales a posteriori de,

$$\theta_{j,n+1} \sim \theta_j \mid \mathbf{y}, \theta_{1,n+1}, \dots, \theta_{j-1,n+1}, \theta_{j+1,n+1}, \dots, \theta_{p,n+1}$$

El método de Gibbs debe su popularidad, al hecho que en muchos modelos estadísticos la distribución condicional a posteriori completa  $f(\theta_j \mid y, \theta_k, k \neq j)$  es posible de simular. Ocurre sin embargo, casos en donde esto no es posible y por ello se hace necesario contar con otros métodos MCMC alternativos. Posiblemente el más genérico de estos esquemas es el de Metropolis.

Las cadenas de Markov que son utilizadas en los esquemas MCMC poseen generalmente un espacio continuo de estados. Tierney (1994)[60] muestra que estos algoritmos convergen a una distribución ergódica estacionaria  $\pi(\theta) = f(\theta \mid y)$  sujeta a tres condiciones de regularidad: irreducibilidad, aperiodicidad e invarianza. La noción de irreducibilidad manifiesta que para cualquier estado  $\theta$  y cualquier conjunto de estados  $B$  con  $\pi(B) > 0$ , existe  $n \in \mathbb{N}$  tal que al cabo de  $n$  iteraciones la cadena pueda hacer una transición de  $\theta$  a  $B$  con probabilidad positiva. La invarianza se refiere por otro lado, a la propiedad de que si empezamos con un vector de estados generado por  $\pi$ , entonces futuras transiciones en la cadena dejarán la distribución marginal de  $\theta$  inalterada; es decir  $\theta_n \sim \pi$ , para cualquier  $n \in \mathbb{N}$ .

El muestreador de Gibbs y el esquema de Metropolis-Hasting son por construcción invariantes con respecto a la distribución a posteriori buscada. Lo que uno debe verificar entonces son la aperiodicidad e irreducibilidad de la cadena, siendo esta última la más crítica pues en ocasiones es posible encontrar un subconjunto de estados tales que cuando la cadena simulada entre

en ella sea improbable salir y el algoritmo por tanto se entrampe en ese punto sin llegar a converger. En la práctica más importante que establecer convergencias teóricas es reconocer la convergencia práctica; es decir juzgar cuantas transiciones  $M$  debe de ser suficientes como para obtener promedios ergódicos que estén cerca de la media deseada.

Una vez finalizada esta etapa comienza nuestro proceso de Inferencia para el o los parámetros o modelos en cuestión.

## 1.5. Comparación de Modelos

Si se desea comparar varios modelos para determinar cual de ellos posee mayor precisión en sus estimaciones o en su capacidad de predicción, existen variados métodos de comparación, dentro de ellos podemos mencionar: Criterio de Información Bayesiana (BIC), Criterio de Información de Akaike (AIC), Criterio de Información de Devianza (DIC), y Criterio de Mínima Pérdida Predictiva a Posteriori (D), entre otros, en particular se usarán el BIC, el DIC y  $pd$  para realizar las comparaciones de modelos y sus grados de complejidad.

### **Criterio de Información Bayesiana.**

El criterio de Información Bayesiana (BIC) y Criterio de Información de Akaike (AIC), son criterios que se utilizan en la selección de modelos para elegir el mejor entre un conjunto de modelos admisibles. Un modelo es mejor que otro si tiene un valor menor (AIC o BIC). El AIC se basa en la distancia de Kullback-Leibler en la teoría de la información y el BIC se basa en una verosimilitud integrada en la teoría Bayesiana. Si no aumenta la complejidad del modelo Bayesiano con el tamaño del conjunto de datos, es preferible el BIC, y en caso contrario el AIC [Bunham, and Anderson](1998)[8]. Otro criterio bastante utilizado es el factor de bayes, el cual realiza un promedio ponderado por la distribución de los parámetros. No obstante autores como Raftery y Kass (1995)[51] sugieren no utilizar el factor de Bayes y usar en su lugar el BIC, el que se calcula sin especificar ninguna distribución a priori, y el cual se define como:

$$BIC = -2\log(f(y|\theta)) + p \cdot \log(n),$$

donde  $f(y|\theta)$  es el valor del máximo de la función de verosimilitud,  $p$  es el número de variables explicativas del modelo (incluyendo las constantes) y  $n$  es el tamaño muestral.

### **Criterio de Información de Devianza**

El criterio de Información de la Desvianza (DIC) es una generalización de los modelos jerárquicos AIC (Criterio de Información de Akaike) y BIC , propuesto por Spiegelhalter et al. (2002)[57], este se utiliza cuando la distribución a posteriori se obtiene o determina vía simulaciones de cadenas de Markov (MCMC), por lo tanto se trata de una aproximación asintótica basándose en muestras grandes y pertinente cuando la distribución simulada

a posteriori es aproximadamente Normal y esta medida utiliza la "Devianza Bayesiana", cuya expresión está dada por:

$$D(\theta) = -2\log(f(y|\theta)) + 2\log(h(y)),$$

Se resume el ajuste del modelo por la esperanza posterior de la devianza  $\bar{D} = E_{\theta|y}[D]$ , como también una medida de complejidad definida por el número efectivo de parámetros  $p_D$ , definido como la devianza esperada menos la devianza evaluada esperada a posteriori, es decir:

$$p_D = E_{\theta|y}[D] - D(E_{\theta|y}[\theta]) = \bar{D} - D(\bar{\theta})$$

Es así que el Criterio de Información de la devianza queda expresado por:

$$DIC = \bar{D} + p_D = 2\bar{D} - D(\bar{\theta}),$$

cuyo cálculo se estima utilizando:

$$\hat{E}_{\theta|y}[D(\theta)] = \sum_{i=1}^N \frac{D(\theta^i)}{N}$$

En otras palabras el DIC se determina como: "dos veces la media de devianza menos la devianza de la media de los parámetros involucrados", una vez que se ha analizado la convergencia de la cadena.

Por cada iteración de dicha cadena, se evalúa la devianza en los valores obtenidos y luego se calcula el promedio de devianza, por otra parte, se realiza la estimación puntual de los parámetros a través del promedio muestral y se evalúa la devianza en estos valores, calculando finalmente el valor del DIC; el modelo con menor DIC, será nuestra elección.

## **Criterio de Mínima Pérdida Predictiva**

Otra alternativa además del DIC, es implementar el criterio de Mínima Pérdida Predictiva, el cual como lo indica Rodríguez(2009)[54] propuesto por Ibrahim, Laud(1995)[39]. Este criterio se basa en seleccionar el mejor modelo entre los admisibles, que posean menor  $D$ , mediante la generación de nuevas réplicas de los datos observados y se define como:

$$D = G + P \text{ donde, } G = \sum_{i=1}^N (\mu_i - y_{i,obs})^2 \text{ y } P = \sum_{i=1}^N \sigma_i^2,$$

$\mu_i$  es la media predictiva a posteriori ( $E(Y_{i,rep}|Y)$ ), y  $\sigma_i^2$  es la varianza predictiva.  $G$  corresponde a la capacidad predictiva y  $P$  es una componente de penalización por sobre ajuste.

Si el valor esperado de las predicciones a futuro es parecido a los datos que ya se han observado, simulaciones a partir de la predictiva serán similares a los datos, en promedio. Con esto se podría decir que el modelo representa o explica el mecanismo que generó esos datos observados. Por tal motivo se dice que el componente  $G$  cuantifica la capacidad del modelo de reproducir los datos observados[Rodriguez,J.](2009)[54], lo cual es deseable. En este caso también se escogerá el modelo con menor valor  $D$ .

De estas medidas ocuparemos el DIC, BIC y la complejidad  $p_D$  para comparar los modelos.

## **1.6. Objetivos de la Tesis**

Esta Tesis tiene como objetivo principal la estimación de parámetros considerando número finito de poblaciones aplicando modelos Jerárquicos Ba-

yesianos paramétricos para variables continuas; asociando una estructura de correlación entre las observaciones en cada población y no correlación entre ellas. El cual nos permite trabajar con una sólo muestra aleatoria sin la necesidad de tener réplicas de las mediciones o tener tamaños muestrales grandes para realizar el análisis inferencial, que es lo usual cuando se trabaja con análisis de varianzas o modelos de efectos aleatorios.

### **1.6.1. Estructura de la tesis**

*Esta tesis contempla un capítulo de introducción y Marco teórico, expuesto anteriormente y tres capítulos de resultados los cuales se describen a continuación:*

- 1. En el capítulo 1, se presentan el marco teorico y los fundamentos teóricos de los procesos que se desarrollaron.*
- 2. En el capítulo 2, Se trabaja con algunos modelos Bayesianos como: el modelo Normal, Normal-gamma (t-Student), Normal- Beta(/), bajo independencia, tanto entre las subpoblaciones como dentro de ellas, ampliando la jerarquización.*

*Se presentan los resultados de las distribuciones condicionales completas, bajo ciertas condiciones, las que nos permiten realizar las estimaciones respectivas vía simulaciones utilizando Gibbs Sampling.*

- 3. En el capítulo 3, de igual manera que en el capítulo 2 se trabajaron distribuciones similares, con algunos cambios en la aleatoriedad de algunos hiperprámetros, para no extender en forma abusiva la jerarquización,*

*pero se incorpora una componente aleatoria, la cual aportará en la obtención de correlación al interior de cada subpoblación y se asumirá no correlación entre las mismas.*

*No obstante la dificultad se hace notar en la obtención de las distribuciones condicionales completas, con una estructura de correlación, lo cual induce a tener que utilizar variables auxiliares para poder implementar Gibbs Sampling, lo que induce a la ampliación en la jerárquización del procedimiento Bayesiano.*

*En el capítulo 4, se realizan las simulaciones de modelos heterocedásticos bajo independencia y bajo correlación obtenidos en el capítulo 2 y 3. Además se realizan las evaluaciones de datos respectivos de los tres modelos involucrados normal, *t*-Student y slash, generados por representaciones estocásticas conocidas (mezcla de normales). Finalmente se realizan las comparaciones entre el total de modelos simulados, estableciendo un ranking junto con sus limitaciones.*



# Capítulo 2

## Modelos Bayesianos Bajo Independencia

### 2.1. Introducción

*El muestreo en varias etapas es frecuentemente usado en análisis de poblaciones humanas, Kish (1965)[38], describe un muestreo en tres etapas de extracción, en el cual una muestra de países es extraída en una primera etapa, una muestra de bloques (estratos) es extraída de la muestra de cada país como segunda etapa, y luego de cada bloque, se extrae una muestra como tercera etapa. Además, en cada etapa se utiliza un muestreo aleatorio simple, y se procede a realizar todas las estimaciones de las cantidades deseadas de acuerdo al interés del estudio.*

*En este capítulo el objetivo inicial es la estimación, es así que se destacan el vector de medias y la matriz de varianzas-covarianzas de los elementos de una población infinita (superpoblación), la cual a su vez está dividida en subpoblaciones.*

*En esta tesis se asume que el diseño del muestreo y la obtención de la muestra fue previo al análisis que sigue, de tal manera que contaremos con  $k$  subpo-*

blaciones de un total de  $K$  y de cada una de ellas fue extraída una cantidad  $n_i$ , completando un total de  $n$  observaciones muestrales de la superpoblación. Es decir,

$$\sum_{i=1}^k n_i = n.$$

Se utilizan algunos supuestos iniciales basados en un trabajo realizado por Scott Smith (1969)[55], cuyo objetivo es realizar estimaciones en poblaciones finitas. No obstante este trabajo se aborda desde el punto de vista de poblaciones infinitas, donde el muestreo se desarrolla en varias sub-etapas, con una estructura de correlación dentro de las subpoblaciones y no correlación fuera de ellas.

La variable de interés se denotará por  $\mathbf{y}_i$  con  $n_i$  elementos muestrales, es decir,  $y_{i1}, \dots, y_{in_i}$ , denotará una muestra de la  $i$ -ésima subpoblación.

Para especificar el modelo para la población de elementos, el diseño estará dado bajo los siguientes supuestos:

S1: Los elementos del  $i$ -ésimo cluster, están no correlacionados y tienen distribución  $F_i$ , con media  $\boldsymbol{\mu}_i$  y varianza  $\boldsymbol{\Sigma}_i$ .

S2: Las medias entre los cluster están no correlacionadas y tienen distribución  $G$  con media  $\nu$  y varianza  $\delta^2$ .

Estos supuestos son equivalentes a especificar una super población, desde la cual, una población finita ha sido extraída y puede ser expresada en términos del concepto de permutabilidad de de Finetti's [Ericson] (1968)[20].

En el modelo planteado se presenta la permutabilidad entre los elementos de cada subpoblación y entre las medias de los clusters.

*Por otra parte los estadísticos están en un continuo desafío por ampliar la familia de distribuciones, ya sea en, su forma, propiedades, espacios paramétricos, etc, de tal manera que estos puedan por ejemplo, modelar conjuntos de datos con valores extremos, que de ser así, se supondría que valores intermedios también debieran funcionar razonablemente, dando robustez a los procedimientos [Morgenthaler et al](1991)[45], y como consecuencia poder modelar datos reales con estas características.*

*Considerando lo anteriormente expuesto, en este trabajo la distribución  $F$  estará dada por una mezcla de distribuciones normales, para dar mayor flexibilidad al modelamiento.  $G$  será una distribución normal y el parámetro de localización de ésta última, en algunos casos será una constante conocida y en otros se asumirá como variable aleatoria con una distribución a priori normal, con media cero y varianza grande (mayor 0 igual a 1000). No obstante, el interés principal de esta tesis será estimar las medias y varianzas de cada subpoblación.*

*Aplicaciones mencionadas en el marco teórico aplican familia de modelos normales,  $t$ -Student y slash entre otras. En esta tesis se utilizan estos tres modelos, y distribuciones a priori estándar de acuerdo a las características de los parámetros de interés.*

## 2.2. Notación

Antes de presentar los modelos propuestos, se establecen a continuación algunas notaciones.

Los símbolos que estén ennegrecidos denotarán vectores y/o matrices y en caso de no ser así se referirá a elementos unidimensionales.

Dentro de los vectores o matrices se destacan: el vector o matriz formada sólo por valores 1's, los cuales se denotarán por  $\mathbf{1}_n$  de dimensión  $n$  y por  $\mathbf{1}_{n^2}$  de dimensión  $n^2$  respectivamente y a  $\mathbf{I}_n$  la matriz identidad de dimensión  $n^2$ . Además las variables aleatorias se denotarán con letras minúsculas.

Vectores que se utilizan en los diferentes modelos son: el vector de medias totales, el cual está formado por cada una de las medias de las subpoblaciones de tamaño  $n_i$  respectivamente y el vector de observaciones totales el que está compuesto por las observaciones de cada subpoblación, cuya notación está dada por:

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_1 \mathbf{1}_{n_1} \\ \mu_2 \mathbf{1}_{n_2} \\ \vdots \\ \mu_k \mathbf{1}_{n_k} \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_k \end{bmatrix}, \quad \mathbf{y}_i = \begin{bmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{in_i} \end{bmatrix}, \quad i = 1, \dots, k$$

Por lo tanto la notación abreviada para el vector de la media poblacional será  $\boldsymbol{\mu}$  y para el vector de las observaciones totales será  $\mathbf{y}$ , ambos de dimensión  $n \times 1$ .

Además  $\boldsymbol{\Sigma}_i$  indicará la matriz de varianzas covarianzas de la  $i$ -ésima subpoblación, definida por elementos distintos de cero en la diagonal y cero fuera de ella y por  $\boldsymbol{\Sigma}$  a la matriz de varianzas-covarianzas de la población completa,

es decir,  $\Sigma_i = \sigma_i^2 \mathbf{I}_{n_i}$ , las varianzas  $\sigma_i^2$  en algunos casos serán fijas, en otros casos aleatorias y definidas en el conjunto de los números reales positivos, no obstante en los procedimientos de las simulaciones se ocuparán varianzas aleatorias y  $\Sigma = (\delta_{ij} \Sigma_i)_{i,j=1,\dots,k}$ , de dimensión  $n^2$  por construcción será una matriz con elementos positivos en la diagonal y cero fuera de ella, donde,

$$\delta_{ij} = \begin{cases} 1 & \text{si } i = j \\ 0 & \text{si } i \neq j. \end{cases}$$

Los parámetros y variables auxiliares que se utilizan están definidos en  $\mathbb{R}$  tratándose de medias y en  $\mathbb{R}^+$  si se trata de varianzas y/o medidas de dispersión o en su defecto en subconjuntos continuos de estos.

### 2.3. Modelo normal heterocedástico con varianzas conocidas

En este modelo, se asumirá que las subpoblaciones son heterocedásticas, con varianzas conocidas. Los espacios tanto muestral y paramétrico, estarán dados por:  $(\mathbf{y}, \Theta) = (\mathbb{R}^n, \mathbb{R}^k)$ . Luego la representación Jerárquica del modelo está dada por:

$$\begin{aligned} \mathbf{y}_i | \mu_i, \Sigma_i &\stackrel{c.ind}{\sim} N(\mu_i \mathbf{1}_{n_i}, \Sigma_i) \\ \mu_i | \nu, \delta^2 &\stackrel{iid}{\sim} N(\nu, \delta^2) \\ \Pi(\nu) &= N(0, 1000), \end{aligned}$$

donde  $\mu_i \in \mathbb{R}$ ,  $\Sigma_i = \sigma_i^2 \cdot \mathbf{I}_{n_i}$ , de  $i = 1, 2, \dots, k$  y  $\delta$  constante conocida y real positiva.

De manera natural, el modelo propuesto puede ser analizado desde el punto

de vista de regresión, como un modelo de efectos aleatorios dado por:

$$\mathbf{y}_i = \mu_i \mathbf{1}_{n_i} + \boldsymbol{\xi}_i,$$

donde  $\boldsymbol{\xi}_i \sim N(\mathbf{0}, \boldsymbol{\Sigma}_i)$ .

No obstante como este trabajo fue diseñado desde el punto de vista Bayesiano por la flexibilidad que este posee de trabajar con observaciones correlacionadas, se adoptó, que los modelos bajo independencia también sean abordados bajo este mismo enfoque, quedando como trabajo futuro este tipo de análisis.

Aplicando teorema de Bayes se obtienen las siguientes distribuciones condicionales completas, las cuales se aplican en el mecanismos de Gibbs sampling en la generación de las cadenas para cada uno de los parámetros de interés y así realizar las estimaciones correspondientes de las medias. Tales distribuciones se explicitan a continuación:

**Algoritmo 2.1**

$$\boldsymbol{\mu}_i \mid \nu, \sigma_i, \delta^2, \mathbf{y}_i \sim N(A_i, B_i) \quad (2.3.1)$$

$$\nu \mid \boldsymbol{\mu}, \delta^2 \sim N\left(\bar{\boldsymbol{\mu}}, \frac{\delta^2}{k}\right), \quad (2.3.2)$$

con

$$A_i = \frac{n_i \cdot \delta^2 \cdot \bar{\mathbf{y}}_i + \nu \cdot \sigma_i^2}{n_i \cdot \delta^2 + \sigma_i^2}, \quad B_i = \frac{\sigma_i^2 \cdot \delta^2}{n_i \cdot \delta^2 + \sigma_i^2}, \quad \bar{\mathbf{y}}_i = \frac{\sum_{j=1}^{n_i} y_{ij}}{n_i}, \quad \text{y} \quad \bar{\boldsymbol{\mu}} = \frac{\sum_{i=1}^k \mu_i}{k}.$$

Donde  $\bar{\mathbf{y}}_i$  es el promedio muestral y  $\bar{\boldsymbol{\mu}}$  es el promedio de las medias de las  $k$  subpoblaciones seleccionadas.

## 2.4. Modelo normal homocedástico con varianza desconocida

En el modelo anterior se agrega una etapa a la jerarquización para flexibilizar el comportamiento de la varianza, pero con la restricción de que sea igual para todas las subpoblaciones (subpoblaciones homocedásticas). Una distribución a priori estándar para la varianza es una Gamma Invertida (IG) y se asumirá que  $(\mu_1, \dots, \mu_k)$  es independiente de  $\sigma^2$ , lo que será denotado como  $(\mu_1, \dots, \mu_k) \Pi \sigma^2$ .

Luego la representación jerárquica del modelamiento queda expresada como:

$$\begin{aligned} \mathbf{y}_i \mid \mu_i, \Sigma_i &\stackrel{c.ind}{\sim} N(\mu_i \mathbf{1}_{n_i}, \Sigma_i) \\ \mu_i \mid \nu, \delta^2 &\stackrel{iid}{\sim} N(\nu, \delta^2) \\ \sigma^2 \mid a, b &\sim IG(a, b) \\ \Pi(\nu) &= N(0, 1000), \end{aligned}$$

donde,  $\Sigma_i = \sigma^2 \mathbf{I}_{n_i}$ , de  $i = 1, \dots, k$  y  $a, b, \delta \in \mathbb{R}^+$  constantes cuales quiera y conocidas (hiperparámetros).

Aplicando teorema de Bayes se agrega la distribución condicional completa para  $\sigma^2$ , obteniéndose como resultados:

### Algoritmo 2.2

$$\mu_i \mid \mathbf{y}_i, \sigma^2, \nu, \delta^2 \sim N(A_i, B_i) \quad (2.4.1)$$

$$\nu \mid \boldsymbol{\mu}, \delta^2 \sim N\left(\bar{\boldsymbol{\mu}}, \frac{\delta^2}{k}\right) \quad (2.4.2)$$

$$\sigma^2 \mid \boldsymbol{\mu}, \mathbf{y}, \delta^2, a, b \sim IG\left(a + \frac{n}{2}, S\right), \quad (2.4.3)$$

con,

$$S = \frac{\left( \sum_{i=1}^k (n_i \cdot [\bar{y}_i^2 - 2\mu_i \cdot \bar{y}_i + \mu_i^2]) \right) + 2b}{2}, \quad \bar{y}_i^2 = \frac{\sum_{j=1}^{n_i} y_{ij}^2}{n_i},$$

y  $A_i, B_i$  dados en la sección 2.3 Algoritmo 2.1 y reemplazando  $\sigma_i^2$  por  $\sigma^2$ .

## 2.5. Modelo normal heterocedástico con varianzas desconocidas

De acuerdo al problema estadístico Inferencial que tengamos presente, es factible poder trabajar con poblaciones homocedásticas, no obstante parece más natural pensar en poblaciones con medias y varianzas distintas, es así que en esta etapa el modelamiento se asumirá varianzas distintas pero aleatorias, para esto se utilizará los símbolos  $a_i, b_i$ , para indicar los hiperparámetros de la distribución a priori para la varianza de la población  $i$ -ésima respectiva y además  $\mu_i \Pi \sigma_i^2, i = 1, \dots, k$ .

Luego la representación jerárquica del modelamiento queda expresada:

$$\begin{aligned} y_{ij} \mid \mu_i, \sigma_i^2 &\stackrel{c.iid}{\sim} N(\mu_i, \sigma_i^2) \\ \mu_i \mid \nu, \delta^2 &\stackrel{iid}{\sim} N(\nu, \delta^2) \\ \sigma_i^2 \mid a_i, b_i &\sim IG(a_i, b_i) \\ \Pi(\nu) &= N(0, 1000), \end{aligned}$$

donde  $a_i, b_i, \delta \in \mathbb{R}^+$  constantes cualesquiera y conocidas, de  $i = 1, \dots, k$ .

Así las distribuciones condicionales completas, están dadas por:



### Algoritmo 2.3

$$\mu_i | \mathbf{y}_i, \sigma_i^2, \delta^2, \sim N ( A_i, B_i) \quad (2.5.1)$$

$$\nu | \boldsymbol{\mu}, \delta^2, \sim N \left( \bar{\boldsymbol{\mu}}, \frac{\delta^2}{k} \right) \quad (2.5.2)$$

$$\sigma_i^2 | \mu_i, \mathbf{y}_i, \delta^2, a_i, b_i \sim IG \left( a_i + \frac{n_i}{2}, S^{(i)} \right), \quad (2.5.3)$$

con  $A_i$  y  $B_i$  dadas en la sección 2.3 algoritmo 2.1 y  $S^{(i)} = \frac{\sum_{j=1}^{n_i} (y_{ij} - \mu_i)^2}{2} + b_i$ .

## 2.6. Modelo normal homocedástico con parámetros para la media aleatorios

En ocasiones se tiene una idea muy vaga de la variabilidad que pueden poseer las medias de subpoblaciones que forman parte de las estimaciones de interés, por lo cual para realizar la inferencia respectiva se debe evaluar diferentes escenarios y ver su factibilidad. Con el objetivo de flexibilizar la varianza de las medias dada por  $\delta^2$ , ésta se asumirá aleatoria con distribución a priori  $IG(c,d)$ , con  $c$  y  $d$  constantes reales, conocidas y positivas.

Luego la representación jerárquica del modelo queda expresada como:

$$\begin{aligned} \mathbf{y}_i | \mu_i, \sigma^2 &\stackrel{c.ind}{\sim} N ( \mu_i, \sigma^2 ) \\ \mu_i | \nu, \delta^2 &\stackrel{iid}{\sim} N ( \nu, \delta^2 ) \\ \sigma^2 | a, b &\sim IG( a, b ) \\ \delta^2 | c, d &\sim IG( c, d ) \\ \Pi(\nu) &= N(0, 1000), \end{aligned}$$

Luego las distribuciones condicionales completas están dadas por:

### Algoritmo 2.4

$$\mu_i | \mathbf{y}_i, \sigma_i^2, \nu, \delta^2 \sim N(A_i, B_i) \quad (2.6.1)$$

$$\nu | \boldsymbol{\mu}, \delta^2 \sim N\left(\bar{\boldsymbol{\mu}}, \frac{\delta^2}{k}\right) \quad (2.6.2)$$

$$\sigma^2 | \mathbf{y}, \boldsymbol{\mu}, a, b \sim IG\left(\frac{n}{2}, S\right) \quad (2.6.3)$$

$$\delta^2 | \boldsymbol{\mu}, c, d, \nu \sim IG\left(c + \frac{k}{2}, q\right), \quad (2.6.4)$$

donde,

$$q = \frac{\sum_{i=1}^k (\mu_i - \nu)^2}{2} + d,$$

$A_i, B_i, S$  dados en la sección 2.4.

## 2.7. Modelo normal heterocedástico con variable auxiliar

Uno de los fines principales de esta tesis es realizar inferencia en subpoblaciones con observaciones modeladas con distintas distribuciones. De esta manera, se introduce una variable auxiliar  $u$  la que nos permitirá trabajar con mezcla de normales, así, la forma de la variable aleatoria que modela los datos dados los parámetros de interés que son la media y la varianza, podrá ser representada estocásticamente como:

$$\mathbf{y}_i \stackrel{d}{=} u_i^{-\frac{1}{2}} \boldsymbol{\Sigma}_i^{\frac{1}{2}} \cdot \mathbf{z}_i + \boldsymbol{\mu}_i,$$

tal que  $u_i$  es independientes de  $\mathbf{z}_i$ ,  $u_i \sim G$  con  $G(0) = 0$ .

Sí,

$$\mathbf{y}_i | u_i, \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i, \stackrel{c.ind}{\sim} N(\boldsymbol{\mu}_i, u_i^{-1} \boldsymbol{\Sigma}_i)$$

y  $u_i = 1, \forall i = 1, \dots, k$ , entonces  $\mathbf{y}_i \sim N(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$ .

Luego la representación jerárquica del modelamiento queda expresada por:

$$\begin{aligned} y_{ij} | u_i, \mu_i, \sigma_i^2, & \stackrel{c.iid}{\sim} N(\mu_i, u_i^{-1} \sigma_i^2) \\ \mu_i | \nu, \delta^2 & \stackrel{iid}{\sim} N(\nu, \delta^2) \\ \sigma_i^2 | a_i, b_i & \sim IG(a_i, b_i), \end{aligned}$$

el cual es equivalente al de la sección 2.5, por lo tanto el algoritmo 2.5 posee las mismas etapas de éste modelamiento.

## 2.8. Modelo t-Student homocedástico

En esta sección la variable auxiliar  $u$ , asumirá una distribución de probabilidad a priori  $G(\varphi/2, \varphi/2)$ , con parámetro de forma y de escala iguales, con  $\varphi$  un real positivo mayor que dos.

Luego si,

$$\begin{aligned} y_{i1}, \dots, y_{in_i} | \mu_i, \sigma^2, u_i & \stackrel{c.iid}{\sim} N(\mu_i, u_i^{-1} \cdot \sigma^2) \\ u_i & \sim G(\varphi/2, \varphi/2), \end{aligned}$$

entonces el vector marginal  $\mathbf{y}_i$  distribuye:

$$\mathbf{y}_i | \mu_i, \Sigma_i, \varphi \stackrel{c.iid}{\sim} t_{n_i}(\mu_i \cdot \mathbf{1}_{n_i}, \Sigma_i, \varphi).$$

Con media  $\mu_i \cdot \mathbf{1}_{n_i}$ , y varianza  $\frac{\varphi}{\varphi-2} \Sigma_i$  con  $\varphi > 2$ , cuya expresión se utilizará para estimar la devianza y comparar los modelos.

Además para analizar la calibración de los procedimientos se evaluarán datos t-student, generados por su representación estocástica expresada anteriormente.

De acuerdo a los supuestos establecidos, la representación jerárquica del mo-

delo *t*-Student homocedástico está dado por:

$$\begin{aligned}
y_{ij} \mid \mu_i, \sigma^2, u_i &\stackrel{c.iid}{\sim} N(\mu_i, u_i^{-1}\sigma^2) \\
\mu_i \mid \nu, \delta^2 &\stackrel{iid}{\sim} N(\nu, \delta^2) \\
\sigma^2 \mid a, b &\sim IG(a, b) \\
u_i \mid \varphi &\stackrel{iid}{\sim} G\left(\frac{\varphi}{2}, \frac{\varphi}{2}\right) \\
\Pi(\nu) &= N(0, 1000),
\end{aligned}$$

donde,  $\delta^2$ ,  $a$ ,  $b$ ,  $\varphi$  son constantes reales conocidas y positivas, con  $\varphi > 2$ .

Así del modelamiento, se obtienen las siguientes distribuciones condicionales para el algoritmo Gibbs Sampling:

**Algoritmo 2.6**

$$\mu_i \mid \mathbf{y}_i, \sigma^2, u_i, \nu, \delta^2 \sim N(A_i, B_i) \quad (2.8.1)$$

$$\nu \mid \boldsymbol{\mu}, \delta^2, \sim N\left(\bar{\boldsymbol{\mu}}, \frac{\delta^2}{k}\right) \quad (2.8.2)$$

$$\sigma^2 \mid \mathbf{y}, \boldsymbol{\mu}, a, b, u \sim IG\left(a + \frac{n}{2}, S\right) \quad (2.8.3)$$

$$u_i \mid \mathbf{y}_i, \mu_i, \sigma, \varphi \sim G\left(\frac{n_i + \varphi}{2}, C_i\right), \quad (2.8.4)$$

donde,

$$B_i = \left[ u_i \sigma^{-2} \cdot n_i + \frac{1}{\delta^2} \right]^{-1}, \quad A_i = \left[ u_i \cdot \sigma^{-2} \cdot n_i \cdot \bar{\mathbf{y}}_i + \frac{\nu}{\delta^2} \right] \cdot B_i,$$

$$S = \frac{\sum_{i=1}^k u_i \cdot \sum_{j=1}^{n_i} (y_{ij} - \mu_i)^2}{2} + b, \quad C_i = \frac{\sigma^{-2} \|\mathbf{y}_i - \mu_i \mathbf{1}_{n_i}\|^2 + \varphi}{2},$$

$$\|\mathbf{y}_i - \mu_i \mathbf{1}_{n_i}\|^2 = \sum_{j=1}^{n_i} (y_{ij} - \mu_i)^2.$$

## 2.9. Modelo t-Student heterocedástico

De igual manera que en casos anteriores, se flexibiliza la varianza asumiéndolas distintas para cada subpoblación y modelando sus distribuciones a priori como IGamma con parámetros  $a_i, b_i$ , respectivamente. Así se obtiene el siguiente modelo jerárquico:

$$\begin{aligned} y_{ij} \mid \mu_i, \sigma_i^2, u_i &\stackrel{c.iid}{\sim} N(\mu_i, u_i^{-1} \sigma_i^2) \\ \mu_i \mid \nu, \delta^2 &\stackrel{iid}{\sim} N(\nu, \delta^2) \\ \sigma_i^2 \mid a_i, b_i &\sim IG(a_i, b_i) \\ u_i \mid \varphi &\stackrel{iid}{\sim} G\left(\frac{\varphi}{2}, \frac{\varphi}{2}\right). \end{aligned}$$

Donde,  $\delta^2, a_i, b_i, \varphi$  son constantes reales conocidas y positivas,  $\varphi > 2$ , de  $i = 1, 2, \dots, k$ .

Luego las distribuciones condicionales para el algoritmo Gibbs Sampling están dadas por:

### Algoritmo 2.7

$$\mu_i \mid \mathbf{y}_i, \sigma_i^2, u_i, \nu, \delta^2 \sim N(A_i, B_i) \quad (2.9.1)$$

$$\sigma_i^2 \mid \mathbf{y}_i, \mu_i, a_i, b_i, u_i \sim IG\left(a_i + \frac{n_i}{2}, S^{(i)}\right) \quad (2.9.2)$$

$$u_i \mid \mathbf{y}_i, \mu_i, \sigma_i^2, \varphi \sim G\left(\frac{n_i + \varphi}{2}, C_i\right). \quad (2.9.3)$$

Donde  $A_i, B_i, C_i$ , están especificadas en la sección anterior reemplazando  $\sigma^2$  por  $\sigma_i^2$  y

$$S^{(i)} = \frac{u_i \sum_{j=1}^{n_i} (y_{ij} - \mu_i)^2}{2} + b_i.$$

## 2.10. Modelo slash homocedástico

En esta etapa, la variable aleatoria  $u$  será modelada a priori por una  $Beta(\varphi, 1)$ , luego la marginal de  $\mathbf{y}_i$  dada la media, varianza y una constante real positiva  $\varphi$  distribuyese slash, es decir:

$$\mathbf{y}_i \mid \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i, \varphi \stackrel{c.ind}{\sim} slash(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i, \varphi).$$

Con media  $\boldsymbol{\mu}_i$ , y varianza  $\frac{\varphi}{\varphi-2} \boldsymbol{\Sigma}_i$ , la cual además se utilizará para estimar la devianza y comparar los modelos. En este caso la representación estocástica para generar datos slash, estará dada por el cociente de una normal y una uniforme elevada a una raíz de la forma  $1/(2 \cdot \varphi)$ , ambas variables aleatorias independientes entre sí, aplicando transformaciones lineales se tiene la slash multivariada y no centrada.

De manera formal tenemos que:

Sea el vector aleatorio  $\mathbf{y}_i = u_i^{-\frac{1}{2\cdot\varphi}} \boldsymbol{\Sigma}_i^{\frac{1}{2}} \cdot \mathbf{z}_i + \boldsymbol{\mu}_i$ , donde  $\mathbf{z}_i \sim N(\mathbf{0}, I_{n_i})$ , e independiente de la distribución uniforme estándar de dimensión  $n_i$ .

Luego

$$\mathbf{y}_i \mid \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i, \varphi \stackrel{c.ind}{\sim} slash(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i, \varphi),$$

tal que,  $E(\mathbf{y}_i) = \boldsymbol{\mu}_i$  y la  $V(\mathbf{y}_i) = \frac{\varphi}{(\varphi-1)} \boldsymbol{\Sigma}_i$ .

Para analizar el ajuste del modelo, se debe calcular la devianza y por ende una integral de la forma  $G_x(a, b) = \int_0^x g(u|a, b) du$ , la que aplicada al contexto queda expresada como  $G_1(\beta, s) = \int_0^1 u^{\beta-1} e^{-us} du$ . Dado que ésta no se puede resolver en forma analítica, es necesario recurrir a métodos numéricos o aproximaciones asintóticas para realizar las estimaciones, dentro de las

cuales, una de ellas establece:

$$G_1(\beta, s) = \int_0^1 u^{\beta-1} e^{-us} du \approx s^{-\beta} \gamma(\beta) \quad s \gg \beta,$$

en cuya expresión  $s$  está relacionada con la distancia de mahalanobis y  $\beta$  es una expresión que depende del tamaño muestral y del parámetro fijo  $\varphi$ , debiendo  $s$  ser mucho más grande que  $\beta$ .

Autores[Lange et al](1993)[37] hacen una revisión detallada de algunas de estas aproximaciones.

En el capítulo 4 se reportarán los estadísticos DIC, pd y BIC para analizar el ajuste y calibración de los modelos, como también los test mencionados para analizar la convergencia de las cadenas y las estimaciones respectivas.

Una publicación[Jing et al](2004)[36], hace una revisión de la distribución slash con algunas extensiones y propiedades.

Así, el modelo jerárquico slash homocedástico está dado por:

$$\begin{aligned} \mathbf{y}_i \mid \mu_i, \Sigma_i, u_i &\stackrel{c.ind}{\sim} N_{n_i}(\mu_i, u_i^{-1} \Sigma_i) \\ \mu_i \mid \nu, \delta^2 &\stackrel{iid}{\sim} N(\nu, \delta^2) \\ \sigma^2 \mid a, b &\sim IG(a, b) \\ u_i \mid \varphi &\stackrel{iid}{\sim} Beta(\varphi, 1) \\ \Pi(\nu) &= N(0, 1000). \end{aligned}$$

Donde,  $\Sigma_i = \sigma_i^2 \cdot \mathbf{I}_{n_i}$ , de  $i = 1, 2, \dots, k$  y  $\delta^2, a, b, \varphi$  son constantes reales conocidas y positivas, con  $\varphi > 2$ .

Las distribuciones condicionales obtenidas para el algoritmo Gibbs Sampling son:

### Algoritmo 2.8

$$\mu_i | \mathbf{y}_i, \sigma_i^2, u_i, \nu, \delta^2 \sim N(A_i, B_i) \quad (2.10.1)$$

$$\nu | \mu, \delta^2 \sim N\left(\bar{\mu}, \frac{\delta^2}{k}\right) \quad (2.10.2)$$

$$\sigma^2 | \mathbf{y}, \boldsymbol{\mu}, \mathbf{u} a, b, \sim IG\left(a + \frac{n}{2}, S\right) \quad (2.10.3)$$

$$u_i | \mathbf{y}, \mu_i, \sigma^2, \varphi \sim GT\left(\frac{n_i}{2} + \varphi, D_i\right). \quad (2.10.4)$$

Con,  $A_i, B_i, S$  dadas en el modelo de la sección 2.8,  $GT$  indica una gamma Truncada, es decir una gamma restringida al intervalo  $[0,1]$  y

$$D_i = \frac{\sigma^{-2} \|\mathbf{y}_i - \mu_i \mathbf{1}_{n_i}\|^2}{2}$$

En este caso como la distribución de base para generar los valores de  $u_i$  es una gamma, puede ocurrir que ciertos valores estén fuera del intervalo  $[0,1]$ , en tal caso se utilizará una truncación estándar, y así cumplir con el recorrido de  $u_i$ .

## 2.11. Modelo slash heterocedástico

Aquí se considera varianzas distintas para cada subpoblación con la misma forma y restricciones para los hiperparámetros que en la sección 2.9, salvo que la variable auxiliar distribuye  $Beta(\varphi,1)$ .

Luego la distribución condicional para el algoritmo de Gibbs Sampling que cambia es:

### Algoritmo 2.9

$$u_i | \mathbf{y}_i, \mu_i, \sigma_i, \varphi \sim GT\left(\frac{n_i + \varphi}{2}, \frac{\sigma_i^{-2} \cdot \|\mathbf{y}_i - \mu_i \cdot \mathbf{1}_{n_i}\|^2}{2}\right).$$



*Se observa que en los modelos homocedástico o heterocedásticos, cuando  $n_i$  se hace muy grande, las estimaciones de los valores esperados de los parámetros de interés son próximos a las medias muestrales y de la variable auxiliar es 1. Además, cuando se presente bastante desconocimiento respecto de la centralización de las prioris, en el caso de las varianzas y de la variable  $u$  se flexibilizarán los hiperparámetros formando prioris planas.*

*Los procedimientos para la obtención de las distribuciones condicionales completas, se encuentran en el anexo A.1.1 de modelos independientes.*

# Capítulo 3

## Modelo Bayesiano Correlacionado

### 3.1. Introducción

*En esta etapa se trabajarán los modelos propuestos en el capítulo anterior estableciendo la existencia de correlación al interior de cada subpoblación y no correlación entre ellas.*

*Para especificar el modelo para la población de elementos, el diseño estará dado bajo los siguientes supuestos:*

*S1: Los elementos del  $i$ -ésimo cluster, están correlacionados, con distribución  $F_i$ , de media  $\boldsymbol{\mu}_i$  y varianza  $\boldsymbol{\Sigma}_i$ .*

*S2: Las medias  $\mu_1, \dots, \mu_k$ , están no correlacionadas, con distribución  $G$ , de media  $\nu$  y varianza  $\delta^2$ .*

Tenemos entonces que  $E(\mathbf{y}_i) = \boldsymbol{\mu}_i$  y

$$\begin{aligned}
Cov(y_{ij}, y_{kl}) &= c_i + \sigma_i^2 & i = k, j = l, \\
&= c_i, & i = k, j \neq l \\
&= 0, & i \neq k
\end{aligned}$$

Donde  $c_i$  actúa como una medida de dispersión, la que asumirá un valor real positivo y por lo tanto nuestra matriz de varianzas-covarianzas queda de la forma:

$$\Sigma_n = \begin{bmatrix} \Sigma_1 & & & \\ & \Sigma_2 & & \\ & & \ddots & \\ & & & \Sigma_k \end{bmatrix}, \quad \Sigma_i = \sigma_i^2 \mathbf{I}_{n_i} + c_i \mathbf{J}_i, \quad \mathbf{J}_i = \mathbf{1}_{n_i} \mathbf{1}'_{n_i}.$$

Dicha estructura, presenta algunas dificultades para obtener las distribuciones a posteriori y desde luego realizar las estimaciones. Es así, que además de la variable auxiliar  $u$  del capítulo 2, se introduce una variable aleatoria  $w$ , para incluir el factor de covarianza en el modelamiento y poder aplicar el esquema de Gibbs Sampling.

La variable  $w$ , actuará como una componente aditiva a la media del modelo de los datos y en el modelamiento de ésta, se incluirá la covarianza  $c_i$  como una medida de dispersión, con el objetivo de poder realizar las estimaciones correspondientes bajo esta estructura.

Se asumirá que  $w_{ij} \parallel \mu_i, \forall i = 1, \dots, k$ , y  $\forall j = 1, \dots, n_i$ , además la variable aleatoria  $w_{ij}$  distribuirá normal con media 0 y varianza  $u_i^{-1} c_i$ , de  $i = 1, \dots, k$ ,  $j = 1, \dots, n_i$ . Para todos los efectos,  $\mathbf{w}_i$  denotará el vector de componente aditiva a los datos de dimensión  $n_i$ .

Así, la media de la distribución marginal para  $\mathbf{y}_i$  en los modelos normal, *t*-Student y slash, bajo una estructura de correlación está dada por:

$$E(\mathbf{y}_i) = E(E(\mathbf{y}_i | \boldsymbol{\mu}_i, \mathbf{w}_i, u_i)) = E(\boldsymbol{\mu}_i + \mathbf{w}_i) = \boldsymbol{\mu}_i,$$

cuya varianza en el caso de la distribución *t*-Student es:

$$\begin{aligned} V(\mathbf{y}_i) &= V(E(\mathbf{y}_i | \boldsymbol{\mu}_i, \mathbf{w}_i, u_i)) + E(V(\mathbf{y}_i | \boldsymbol{\mu}_i, \mathbf{w}_i, u_i)) \\ &= V(\boldsymbol{\mu}_i + \mathbf{w}_i) + E(u_i^{-1} \sigma_i^2 \mathbf{I}_{n_i}) \\ &= \frac{\varphi/2}{\varphi/2-1} \boldsymbol{\Sigma}_i, \end{aligned}$$

y en el caso de la distribución slash,

$$V(\mathbf{y}_i) = \frac{\varphi}{\varphi-1} \boldsymbol{\Sigma}_i.$$

Los procedimientos que se desarrollan en este capítulo, son similares a los obtenidos en el capítulo anterior, presentando los modelos normal, *t*-Student y slash, desde el punto de vista heterocedásticos por poseer mayor flexibilidad, no obstante las simulaciones que se presentan en el capítulo siguiente, son adaptables ya sea a modelos homocedásticos o heterocedásticos. Además se asumirá que las covarianzas de cada subpoblación respectivamente, son aleatorias con distribución  $IG(e_i, f_i)$ , con  $e_i, f_i > 0$ , cualesquiera y conocidas de  $i = 1, \dots, k$ .

**Condiciones para la covarianza, son las siguientes:**

1. Se asume que entre dos variables hay correlación positiva  $\rho$ , si existe una constante positiva  $\varrho$ , tal que:  $\varrho < \rho < 1$ . Un criterio estándar, para que exista correlación positiva significativa entre dos variables, es que ésta se encuentre entre 0,5 y 1 (a menos que se ocupe un criterio más amplio), luego entre las varianzas y la covarianza se debe cumplir que,  $0 < \sigma_i^2 < \frac{1-\varrho}{\varrho} \cdot c_i$ , para subpoblaciones heterocedásticas con covarianzas distintas o que el  $\max\{\sigma_i^2\} < \frac{1-\varrho}{\varrho} \cdot c$ , para covarianzas iguales entre

las subpoblaciones,  $\forall i = 1, \dots, k$ .

2. En el caso de querer ampliar el modelamiento respecto de la covarianza (pensando en el recorrido), es decir, el caso de desear analizar covarianza negativa, la restricción para éstas es:  $-\sigma_i^2/2 < c_i < -\frac{\rho}{1+\rho} \cdot \sigma_i^2$ , o  $\max\{-\sigma_i^2/2\} < c < \min\{-\frac{\rho}{1+\rho} \cdot \sigma_i^2\}$ ,  $\forall i = 1, \dots, k$ , con  $0 < \rho < 1$ . Se propone como solución para las estimaciones, el mecanismo MCMC.
3. Si la varianza es muy grande, la covarianza podría flexibilizar su rango, admitiendo la posibilidad de modelarla a través de una distribución normal en el esquema Bayesiano.

### 3.2. Modelo normal heterocedástico

Los modelos asociados para las distribuciones a priori de las medias y varianzas son similares al capítulo anterior como ya se mencionó, junto con la variable aleatoria  $u$  que en este caso se asume constante igual a 1 de igual manera que en la sección 2.7 y la variable aleatoria  $w$  que distribuye normal.

De esta manera la marginal de  $\mathbf{y}_i$  distribuye normal con media  $\boldsymbol{\mu}_i$  con matriz de varianzas-covarianzas definidas bajo esta nueva estructura de correlación. Luego la estructura de modelamiento jerarquizada queda de la forma:

$$\begin{aligned}
 y_{ij} | \mu_i, w_{ij}, \sigma_i^2, u_i, c_i &\stackrel{c.iid}{\sim} N(\mu_i + w_{ij}, u_i^{-1} \sigma_i^2) \\
 w_{ij} | u_i, c_i &\stackrel{c.iid}{\sim} N(0, u_i^{-1} c_i) \\
 \mu_i | \nu, \delta^2 &\stackrel{iid}{\sim} N(\nu, \delta^2) \\
 c_i | e_i, f_i &\sim IG(e_i, f_i) \\
 \sigma_i^2 | a_i, b_i &\sim IG(a_i, b_i).
 \end{aligned}$$

Donde  $a_i, b_i, \delta, e_i, f_i, u_i \in \mathbb{R}^+$ ,  $\nu \in \mathbb{R}$ , constantes conocidas.

De igual manera este modelo puede analizarse como un modelo de efectos aleatorios dado por:

$$\mathbf{y}_i = \mu_i \mathbf{1}_{n_i} + \mathbf{w}_i + \boldsymbol{\xi}_i,$$

donde  $\boldsymbol{\xi}_i \sim N(\mathbf{0}, \boldsymbol{\Sigma}_i)$ , presentado en el capítulo anterior y  $\mathbf{w}_i$  bajo las especificaciones expuestas anteriormente, lo cual queda para un trabajo futuro.

Luego las distribuciones condicionales obtenidas para el algoritmo Gibbs Sampling son las siguientes:

### Algoritmo 3.1

$$\mu_i | \mathbf{y}_i, \sigma_i^2, \nu, \delta^2, w_i, u_i \sim N(A_i, B_i) \quad (3.2.1)$$

$$w_{ij} | y_{ij}, \mu_i, u_i, c_i, \sigma_i^2 \sim N(\kappa_{ij}, \tau_{ij}) \quad (3.2.2)$$

$$\sigma_i^2 | \mathbf{y}_i, a_i, b_i, u_i, \mu_i, \mathbf{w}_i \sim IG\left(\frac{n_i}{2} + a_i, S^{(i)}\right) \quad (3.2.3)$$

$$c_i | u_i, w_i \sim IG\left(e_i + \frac{n_i}{2}, H_i\right). \quad (3.2.4)$$

Con,

$$A_i = \frac{n_i \cdot \delta^2 \cdot u_i \cdot (\bar{\mathbf{y}}_i - \bar{\mathbf{w}}_i) + \nu \cdot \sigma_i^2}{n_i \cdot u_i \cdot \delta^2 + \sigma_i^2}, \quad B_i = \frac{\sigma_i^2 \cdot \delta^2}{u_i \cdot n_i \cdot \delta^2 + \sigma_i^2}$$

$$\kappa_{ij} = \frac{(y_{ij} - \mu_i) \cdot c_i}{c_i + \sigma_i^2}, \quad \tau_{ij} = \frac{\sigma_i^2 \cdot c_i}{u_i \cdot (c_i + \sigma_i^2)}$$

$$S^{(i)} = b_i + \frac{u_i \cdot \sum_{j=1}^{n_i} (y_{ij} - (\mu_i + w_{ij}))^2}{2}.$$

$$H_i = f_i + 0,5 \cdot u_i \cdot \|\mathbf{w}_i\|^2$$

### 3.3. Modelo t-Student heterocedástico

En este modelo se asume la aleatoriedad de la variable auxiliar  $u$ , a través de un modelo Gamma utilizado en las secciones 2.8 y 2.9 del capítulo 2, y la variable  $w$  especificada en la sección anterior, de esta manera se obtiene una distribución t-Student para la marginal  $\mathbf{y}_i$  la que se expresa de la siguiente manera:

$$\mathbf{y}_i \mid \mu_i, \Sigma_i, \varphi, c_i \stackrel{c.ind}{\sim} t(\mu_i \mathbf{1}_{n_i}, \Sigma_i, \varphi).$$

Donde  $\Sigma_i = c_i \mathbf{1}_{n_i} + \sigma_i^2 \mathbf{I}_{n_i}$ .

Luego el modelo jerarquizado bajo una estrucutra de correlación está dado por

$$\begin{aligned} y_{ij} \mid u_i, \mu_i, w_{ij}, \sigma_i^2, c_i &\stackrel{c.iid}{\sim} N(\mu_i + w_{ij}, u_i^{-1} \sigma_i^2) \\ w_{ij} \mid u_i, c_i &\stackrel{c.ind}{\sim} N(0, u_i^{-1} c_i) \\ \mu_i \mid \nu, \delta^2 &\stackrel{iid}{\sim} N(\nu, \delta^2) \\ u_i \mid \varphi &\stackrel{iid}{\sim} G(\varphi/2, \varphi/2) \\ c_i \mid e_i, f_i &\sim IG(e_i, f_i) \\ \sigma_i^2 \mid a_i, b_i &\sim IG(a_i, b_i). \end{aligned}$$

Con  $a_i, b_i, e_i, f_i, \delta, \varphi \in \mathbb{R}^+, \nu \in \mathbb{R}$  constantes conocidas.

Por lo tanto las distribuciones Condicionales completas obtenidas para el algoritmo Gibbs Sampling son:

#### Algoritmo 3.2

$$\mu_i \mid \mathbf{y}_i, \sigma_i^2, \nu, \delta^2, \mathbf{w}_i, u_i \sim N(A_i, B_i) \quad (3.3.1)$$

$$w_{ij} \mid y_{ij}, \mu_i, u_i, c_i, \sigma_i^2 \sim N(\kappa_{ij}, \tau_{ij}) \quad (3.3.2)$$

$$\sigma_i^2 \mid \mathbf{y}_i, a_i, b_i, u_i, \mu_i, \mathbf{w}_i \sim IG\left(\frac{n_i}{2} + a_i, S^{(i)}\right) \quad (3.3.3)$$

$$u_i \mid \mathbf{y}_i, \mu_i, \mathbf{w}_i, \varphi, c_i \sim G\left(\frac{\varphi}{2} + n_i, B1_i\right) \quad (3.3.4)$$

$$c_i \mid u_i, \mathbf{w}_i \sim IG\left(\frac{n_i}{2} + e_i, H_i\right). \quad (3.3.5)$$

Con,

$$\begin{aligned}
A_i &= \frac{n_i \cdot \delta^2 \cdot u_i \cdot (\bar{\mathbf{y}}_i - \bar{\mathbf{w}}_i) + \nu \cdot \sigma_i^2}{n_i \cdot u_i \cdot \delta^2 + \sigma_i^2}, & B_i &= \frac{\sigma_i^2 \cdot \delta^2}{u_i \cdot n_i \cdot \delta^2 + \sigma_i^2} \\
\kappa_{ij} &= \frac{(y_{ij} - \mu_i) \cdot c_i}{c_i + \sigma_i^2}, & \tau_{ij} &= \frac{\sigma_i^2 \cdot c_i}{u_i \cdot (c_i + \sigma_i^2)} \\
S^{(i)} &= b_i + \frac{1}{2} \cdot (u_i \sum_{j=1}^{n_i} (y_{ij} - (\mu_i + w_{ij}))^2), \\
B1_i &= \frac{\frac{\|\mathbf{y}_i - (\mu_i \mathbf{1}_{n_i} + \mathbf{w}_i)\|^2}{\sigma_i^2} + \frac{\|\mathbf{w}_i\|^2}{c_i} + \varphi}{2}, & H_i &= \frac{u_i \|\mathbf{w}_i\|^2 + 2f_i}{2}.
\end{aligned}$$

### 3.4. Modelo slash heterocedástico

En esta etapa a diferencia del modelo anterior, el modelo especificado para la v.a  $u$  está dado por una  $Beta(\varphi, 1)$ . Respectivamente si integramos la conjunta del modelo de los datos, la variable  $w$  y la variable  $u$  dados los parámetros de interés respecto de  $w$  y  $u$ , obtenemos que el vector aleatorio marginal  $\mathbf{y}_i$  distribuye:

$$\mathbf{y}_i \mid \mu_i, \Sigma_i, \varphi, c \stackrel{c.ind}{\sim} slash(\mu_i \mathbf{1}_{n_i}, \Sigma_i, \varphi).$$

Donde  $\Sigma_i = c_i \mathbf{1}_{n_i} + \sigma_i^2 \mathbf{I}_{n_i}$ .

Luego el modelo jerarquizado está dado por:

$$\begin{aligned}
y_{ij} \mid u_i, \mu_i, w_{ij}, \sigma_i^2, c_i &\stackrel{c.iid}{\sim} N(\mu_i + w_{ij}, u_i^{-1} \sigma_i^2) \\
w_{ij} \mid u_i, c_i &\stackrel{c.iid}{\sim} N(0, u_i^{-1} c_i) \\
\mu_i \mid \nu, \delta^2 &\stackrel{iid}{\sim} N(\nu, \delta^2) \\
u_i \mid \varphi &\stackrel{iid}{\sim} beta(\varphi, 1) \\
c_i \mid e_i, f_i &\stackrel{c.ind}{\sim} IG(e_i, f_i) \\
\sigma_i^2 \mid a_i, b_i &\stackrel{ind}{\sim} IG(a_i, b_i).
\end{aligned}$$



Con  $a_i, b_i, e_i, f_i, \varphi \in \mathbb{R}^+$  constantes conocidas.

Así las distribuciones condicionales completas obtenidas para el algoritmo Gibbs Sampling son:

**Algoritmo 3.3**

$$\mu_i | \mathbf{y}_i, \sigma_i^2, \nu, \delta^2, \mathbf{w}_i, u_i \sim N(A_i, B_i) \quad (3.4.1)$$

$$w_{ij} | y_{ij}, \mu_i, u_i, c_i, \sigma_i^2 \sim N(\kappa_{ij}, \tau_{ij}) \quad (3.4.2)$$

$$\sigma_i^2 | \{\mathbf{y}_i, a_i, b_i, u_i, \mu_i, \mathbf{w}_i \sim IG\left(\frac{n_i}{2} + a_i, S^{(i)}\right) \quad (3.4.3)$$

$$u_i | \mathbf{y}_i, \mu_i, \mathbf{w}_i, \varphi, c_i \sim GT(\varphi + n_i, B1_i) \quad (3.4.4)$$

$$c_i | u_i, \mathbf{w}_i \sim IG\left(e_i + \frac{n_i}{2}, H_i\right). \quad (3.4.5)$$

Donde,

$$A_i = \frac{n_i \cdot \delta^2 \cdot u_i \cdot (\bar{\mathbf{y}}_i - \bar{\mathbf{w}}_i) + \nu \cdot \sigma_i^2}{n_i \cdot u_i \cdot \delta^2 + \sigma_i^2}, \quad B_i = \frac{\sigma_i^2 \cdot \delta^2}{u_i \cdot n_i \cdot \delta^2 + \sigma_i^2}$$

$$\kappa_{ij} = \frac{(y_{ij} - \mu_i) \cdot c_i}{c_i + \sigma_i^2}, \quad \tau_{ij} = \frac{\sigma_i^2 c_i}{u_i (c_i + \sigma_i^2)}$$

$$S^{(i)} = b_i + \frac{u_i \cdot \sum_{j=1}^{n_i} (y_{ij} - (\mu_i + w_{ij}))^2}{2}$$

$$B1_i = \frac{\frac{\|\mathbf{y}_i - (\mu_i \mathbf{1}_{n_i} + \mathbf{w}_i)\|^2}{\sigma_i^2} + \frac{\|\mathbf{w}_i\|^2}{c_i}}{2}, \quad H_i = f_i + \frac{u_i \|\mathbf{w}_i\|^2}{2}.$$

De igual manera se establece que cuando los tamaños muestrales aumentan las estimaciones de los valores esperados de los parámetros de interés, en este caso las medias, varianzas y covarianzas se concentran alrededor de las medias muestrales.

Los procedimientos para la obtención de las distribuciones condicionales completas, se encuentran en el anexo A.1.2 en modelos correlacionados.

*Si la correlación es no significativa los modelos son equivalentes a los obtenidos en el capítulo anterior.*

# Capítulo 4

## Simulaciones

### *Introducción*

*El objetivo de este capítulo, es apoyarnos de procesos de simulación para ejecutar nuestros procedimientos y así poder inspeccionar las condiciones que debe poseer el sistema para realizar estimaciones apropiadas.*

*Las simulaciones son comunes hacerlas bajo varios escenarios, las que nos permiten analizar comportamientos de los elementos involucrados, medir la efectividad de los procesos respecto de condiciones dadas y evaluar su factibilidad. Estas permiten medir el impacto de cambios en el proceso, mejora el conocimiento de este y enriquece la toma de decisiones, sobre todo cuando se les quiere aplicar a situaciones reales.*

*Dentro de los elementos claves que ayudan, en el desarrollo de un modelo de simulación, se encuentran:*

- 1. Tamaño de corridas adecuadas.*
- 2. Especificaciones claras de las variables involucradas.*
- 3. Relaciones adecuadas entre los elementos del sistema.*

4. *Análisis constante del comportamiento de los resultados obtenidos.*
5. *Realizar los ajustes necesarios dentro de las especificaciones establecidas para comprobar la mejora de resultados.*
6. *Reportar debilidades y fortalezas del sistema.*

*En la ejecución de las simulaciones se consideraron algunos aspectos, dentro de los cuales se destacan:*

1. *Se realizaron variadas simulaciones, con condiciones iniciales arbitrarias con dos subpoblaciones.*
2. *Los valores de los hiperparámetros de las prioris fueron similares para establecer posibles comparaciones.*
3. *Las simulaciones se realizaron para cada uno de los modelos con tamaños muestrales distintos.*
4. *En general el número de iteraciones fue alto para garantizar convergencia de las cadenas.*
5. *Se eliminó un porcentaje inicial de la cadena, asumiendo un máximo dentro del análisis como criterio estándar(en algunos casos pudo ser menos), para disminuir el efecto de condiciones iniciales.*
6. *Se presentan los test de Diagnóstico de Covergencia estándar como: Geweke(1992), Raftery Lewis(1992), y Heidelberg Welch(1983), que nos proporciona el boa de R-project(versión 2.6), , para analizar el comportamiento de las cadenas y la convergencia de las medias y así entregar las estimaciones respectivas de los parámetros de interés. ver anexo A.2.1 para los modelos independientes y anexo A.2.2 para los modelos bajo correlación.*

7. Una vez eliminado una cantidad inicial, del porcentaje restante de la cadena se realizó un muestreo sistemáticos, para obtener la convergencia de éstas, utilizando los test de Raftery and Lewis, Heidleberger and Welch y Geweke. No obstante los resultados que se presentan en las tablas, están dados en general con muestreo cada 10, no obstante el test de Raftery and Welch en algunos casos falla la convergencia de las varianzas.
8. La calibración de cada uno de los pocedimientos, se realizó con 100 simulaciones.

## DEVIANZA

Para los modelos de las secciones 2.3 a la sección 2.7 la forma de la Devianza para cada población está dada por:

$$D(\theta; \mathbf{y}_i) = \log(2\pi\sigma_i^2) + \sum_{i=1}^k \sum_{j=1}^{n_i} \left( \frac{y_{ij} - \mu_i}{\sigma_i} \right)^2.$$

Para los modelo de las secciones 2.8 y 2.9 es:

$$D(\theta; \mathbf{y}_i) = -2\log \left( \frac{\gamma \left( \frac{\varphi + n_i}{2} \right)}{(\sigma^2 \pi \varphi)^{n_i/2} \gamma(\varphi/2)} \right) + (\varphi + n_i) \log \left[ 1 + \frac{\lambda}{\varphi} \right].$$

$$\text{Donde } \lambda = \frac{\sum_{j=1}^{n_i} (y_{ij} - \mu_i)^2}{\sigma^2}$$

Para el modelo 2.10 y 2.11 es:

$$D(\theta; \mathbf{y}_i) \approx \log(2\pi\sigma_i^2) + 2(\varphi + n_i/2) \log \left( \frac{\sum_{j=1}^{n_i} (y_{ij} - \mu_i)^2}{2\sigma^2} \right) - 2\log \gamma \left( \varphi + \frac{n_i}{2} \right) \sum_{j=1}^{n_i} \left( \frac{y_{ij} - \mu_i}{\sigma_i} \right)^2.$$

En el caso de los modelos correlacionados, para estimar las devianzas se ocupan las verosimilitudes Multivariadas, con la estructura de matriz de varianzas covarianzas explicitada anteriormente.

## 4.1. Modelos Independientes

Para los modelos independientes los valores de hiperparámetros que se optaron para la simulación, fueron:  $\varphi = 10$ ,  $a_1 = a_2 = 5$ ,  $b_1 = 7$ ,  $b_2 = 6$ ,  $\nu = 10$ ,  $\delta^2 = 100$ . Con estos valores, la priori de las medias, quedó centrada en 10 con varianza 100, las prioris para las varianzas de las poblaciones 1 y 2, quedaron centradas en 1.75 y 1.5 con varianzas de 1.02 y 0.75 respectivamente.

Los valores de parámetros utilizados, para realizar las comparaciones versus las estimaciones fueron:  $\mu_1 = -99,4423$ ,  $\mu_2 = 60,9660$ ,  $\sigma_1^2 = 1$ ,  $\sigma_2^2 = 1,2$ .

En los modelos Independientes se realizaron 30 veces 50 simulaciones para medir la calibración de estos (con 10000 iteraciones) con distintos tamaños muestrales, cuyo reporte final, se obtuvo con,  $n_i = 30, 60, 100$  observaciones, que se indica en el cuadro 4.5. Además bajo pérdida cuadrática, las estimaciones para los parámetros, junto con el error, están dados por la media y la varianza de la distribución a posteriori (estimada), que se obtienen de las cadenas, junto con el resultado de su intervalo de credibilidad.

Las estimaciones que se presentan a continuación se realizaron con 100000 a 150000 iteraciones, de ellas se eliminaron el 20 por ciento inicial y se realizó un muestreo sistemático de las cadenas usando MCMC para realizar las estimaciones y/o analizar la convergencia de estas. Los largos que se emplearon fueron cada 10, 20 o 25 de ellas, valores extremos para que cada una de ellas presentara convergencia. Obtenida la muestra de la iteración testeada se realizaron las estimaciones.

#### 4.1.1. Estimaciones

<b>Criterio</b>	<b>Numero de Obs.</b>	<b>Normal</b>	<b>t-Student</b>	<b>Slash</b>
<b>DIC</b>	30	<b>170,5973</b>	186,3018	197,5202
	60	<b>346,0262</b>	400,3599	381,6638
	100	<b>573,8066</b>	642,2035	604,5395
<b>Pd</b>	30	3,1314	<b>1,241</b>	12,3919
	60	3,8069	<b>0,7341</b>	11,4786
	100	3,69491	<b>0,4998</b>	10,0226
<b>BIC</b>	30	<b>188,3003</b>	207,7842	200,3674
	60	<b>362,3782</b>	425,6301	386,3375
	100	<b>591,5931</b>	653,3115	623,8129

Cuadro 4.1: Ajuste de los Modelos Ind.

*En los modelos Independientes, para los distintos tamaños muestrales pequeños, el que mejor se ajusta es el modelo Normal. No obstante para tamaños muestrales grandes los datos normales tienden a ajustar de igual manera a los tres modelos, lo cual es esperable dada las características de la distribución normal. Dentro de estos tres modelos el que presenta menor complejidad es el modelo t-Student [Cuadro 4.1].*

Modelo Normal						
Parámetro		Nº de Obs.	Media	Desv.est	IC	
$\mu_1$		30	-99,106685	0,13416	-99,3680122	-98,839921
	-99,44233	60	-99,5563574	0,11117	-99,7789014	-99,343478
		100	-99,4756108	0,08150	-99,6710499	-99,306156
$\mu_2$		30	60,9333244	0,20336	60,5442969	61,345614
	60,96605	60	60,7690305	0,13472	60,5060262	61,038509
		100	60,9381461	0,10139	60,727318	61,154434
$\sigma_1^2$		30	0,5446681	0,09524	0,3748543	0,737403
	1	60	0,7386262	0,10762	0,5410284	0,9477302
		100	0,6925688	0,08225	0,6819915	1,155964
$\sigma_2^2$		30	1,2104712	0,22711	0,7963064	1,676811
	1,2	60	1,067701	0,16133	0,7787134	1,402957
		100	1,028564	0,13078	0,9181382	1,585036

Cuadro 4.2: Estimaciones con Modelo Normal

A mayor tamaño muestral, en la mayoría de los casos el error disminuye y las estimaciones mejoran, obteniéndose menor amplitud en los intervalos de credibilidad [Cuadro 4.2], [Cuadro 4.3], [Cuadro 4.4].

Modelo T-student						
Parámetro		Nº de Obs.	Media	Desv.est	IC	
$\mu_1$		30	-99,2633353	0,127438456	-99,5240903	-99,015602
	-99,44233	60	-99,373233	0,077265866	-99,5291837	-99,227343
		100	-99,475404	0,088414478	-99,6423434	-99,293704
$\mu_2$		30	60,8641464	0,166189681	60,5399535	61,200805
	60,96605	60	61,061092	0,0989714	60,8658511	61,25798
		100	60,980657	0,092039633	60,7970204	61,161044
$\sigma_1^2$		30	0,7688306	0,217308352	0,4161809	1,202117
	1	60	1,176835	0,312293324	0,6327922	1,808733
		100	1,376123	0,39592596	0,6719182	1,877684
$\sigma_2^2$		30	1,4569299	0,470822026	0,6671733	2,435565
	1,2	60	1,590704	0,475130675	0,769715	2,557025
		100	1,424064	0,422127737	0,6673381	2,22171

Cuadro 4.3: Estimaciones con Modelo t-Student



En las distintas simulaciones se observó, que Medias extremas y varianzas grandes, datos normales se confunden con procedimiento T-student.

Modelo Slash						
Parámetro		Nº de Obs.	Media	Desv.est	IC	
$\mu_1$		30	-99,5565122	0,36842	-100,2175297	-98,800685
	-99,44233	60	-99,200334	0,24960	-99,6667285	-98,698223
		100	-99,4417419	0,21235	-99,8696718	-99,034813
$\mu_2$		30	60,6515631	0,39436	59,8280633	61,404171
	60,96605	60	60,947192	0,32132	60,3336567	61,620178
		100	60,9825106	0,23545	60,51273	61,478681
$\sigma_1^2$		30	0,7685184	0,27627	0,3877812	1,334343
	1	60	1,11347	0,20592	0,7549718	1,537146
		100	0,8168201	0,12109	0,5939067	1,053417
$\sigma_2^2$		30	0,8786015	0,38172	0,4140839	1,708231
	1,2	60	1,126841	0,22361	0,7736548	1,570799
		100	0,9117777	0,37730	0,5099988	1,808601

Cuadro 4.4: Estimaciones con Modelo Slash

Medias extremas y varianzas pequeñas datos normales se identifican con el procedimiento Normal.

El modelo normal independiente, pasa todos los test de convergencia, para los tamaños muestrales pre-establecidos, ver en Anexo A.2 [Cuadro 11].

El test de Geweke, en casi todos los casos indica convergencia, no obstante en los modelos t-Student y Slash independientes, los test de Heildeberg and Welch y Raftery, no siempre se presenta tan natural o simple la convergencia; en algunos casos es necesario incrementar el número de iteraciones, a su vez eliminar un mayor número de iteraciones iniciales y/o cambiar el periodo o salto al realizar el muestreo de las cadenas, ver en Anexo A.2 [Cuadro 12], [Cuadro 13].

En la calibración de los procedimientos, se observa que datos normales en un alto porcentaje se identifican con el procedimiento normal, en este caso mayor al 88 por ciento [Cuadro 4.5]. Una debilidad de este proceso, es que en particular datos *t*-Student con tamaño muestral de  $n_i = 60$  se confunden con el procedimiento normal.

### 4.1.2. Densidades a Posteriori

Los siguientes gráficos que se presentan, fueron obtenidos con 100 observaciones para cada una de las subpoblaciones.

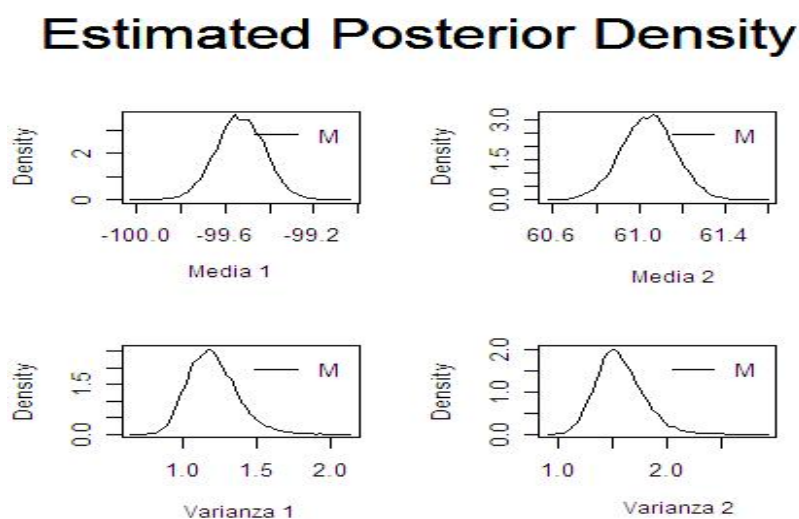


Figura 4.1: D.Posteriori en el Modelo Normal

La forma de las densidades a posteriori, tanto para las medias como para las varianzas, son similares en los modelos normales y *t*-Student respectivamente. No obstante en el modelo Slash al obtenerse una menor precisión en la estimación y un incremento en el error, la forma de la densidad a posteriori de la varianza de la población 2, presenta una diferencia con los otros dos modelos [Figura 4.1], [Figura 4.2], [Figura 4.3].

## Estimated Posterior Density

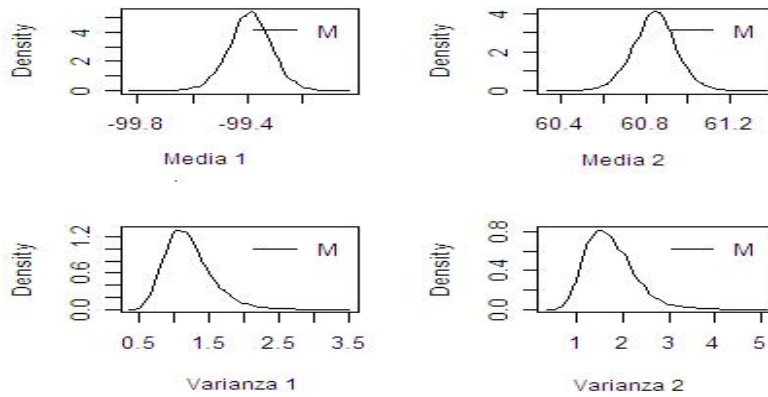


Figura 4.2: D.Posteriori en el Modelo t-Student

## Estimated Posterior Density

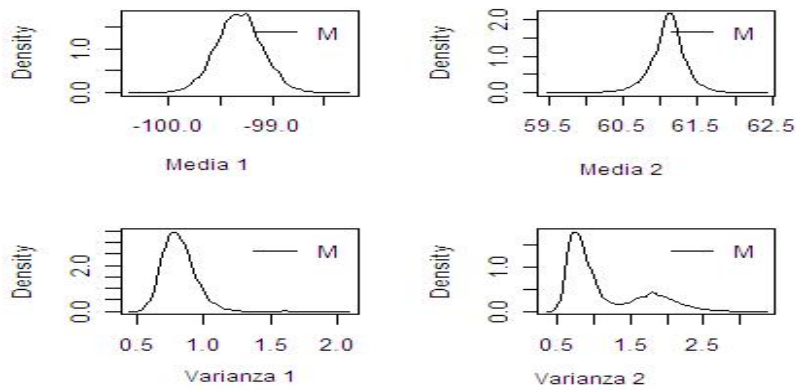


Figura 4.3: D.Posteriori en el Modelo Slash

### 4.1.3. Calibración del procedimiento

Datos	Procesos	Modelos Independientes		
		30	60	100
Normales	Normal v/s T-student	88 - 100	90 - 100	90 - 100
	Normal v/s Slash	85 - 99	88 - 100	98 - 100
T-student	Normal v/s T-student	80 - 100	50 - 90	85 - 100
Slash	Normal v/s Slash	85 - 98	83 - 97	83 - 100

Cuadro 4.5: Calibración de M.Independientes

Los modelos *Independientes*, presentan un mejor ajuste y una menor complejidad, indistintamente del número de observaciones (mayores a 30). No obstante en general, de los modelos correlacionados, el modelo *t-Student* con 100 observaciones o más, obtiene la menor complejidad.

Con tamaños muestrales pequeños, independiente del valor del hiperparámetro  $\varphi$ , datos tanto normales como *t-Student* se identifican con procedimiento *t-Student*.

En el cuadro 4.5 de las 50 simulaciones (30 veces) para cada tamaño muestral, datos Normales con tamaño muestral 30, al procesarlos con el modelo Normal y modelo *t-Student* estos, se identifican con el modelo Normal entre un 88 y 100 por ciento de las veces y a medida que aumenta el tamaño muestral el porcentaje aumenta a un 90 y 100 por ciento. Si se considera un porcentaje sobre el 60 por ciento como una buena calibración, estos pasan la prueba. En cambio datos *Slash* con tamaño muestral de 30 al procesarlos con los modelos Normal y *Slash* estos se identifican con este entre un 85 y 98 por ciento, hasta llegar a un 83 y un 100 por ciento aumentando su variabilidad respecto del Normal, no obstante de igual manera se considera satisfactoria la calibración.

## 4.2. Modelos Correlacionados

### 4.2.1. Estimaciones

*Dado que los procesos desarrollados en el capítulo anterior, se asumieron con covarianza positiva, y considerando las relaciones con las varianzas respectivas, se consideró un valor extremo inferior para  $\rho$  de 0.38. Es así que los valores de los hiperparámetros que se utilizaron en esta etapa, en algunos casos son similares al caso Independiente, con algunas variaciones para mejorar las estimaciones de los mismos, con una estructura de correlación.*

*Los valores utilizados para los Hiperparámetros de la distribución a priori de las medias fue de  $\delta^2 = 100$ ,  $\nu = 0$ ,  $e_1 = e_2 = 4$ ,  $f_1 = f_2 = 3$ ,  $a_1 = a_2 = b_1 = b_2 = 5$ ,  $\varphi = 4$ , por lo tanto la priori de las medias quedó centrada en 0 con varianza 100 y las prioris de las varianzas quedaron centradas en 1.25 con varianza 0.52.*

*Los valores de parámetros utilizados para realizar las comparaciones respectivas fueron:  $\mu_1 = -99,4423$ ,  $\mu_2 = 60,9660$ ,  $\sigma_1^2 = 0,9$ ,  $\sigma_2^2 = 0,95$ ,  $c_1 = c_2 = 0,6$ .*

*Se trabajó con datos normales para ver el ajuste de los modelos Bayesianos Normales, *t*-Student y slash.*

*Se realizaron 50 simulaciones (30 veces) con 10000 iteraciones para medir la calibración de estos, cuyo reporte se observa en la tabla 4.10.*

Criterio	Número de obs.	Normal	t-Student	Slash
DIC	30	160,6636	170,093	160,0225
	60	352,63	317,0633	350,087
	100	560,4263	565,3868	564,798
Pd	30	2,7537	0,6632	1,7052
	60	6,4596	0,5783	0,825
	100	9,6978	0,5052	0,7244
BIC	30	191,1049	204,7153	191,5609
	60	375,66	351,8552	384,3857
	100	596,9794	600,325	599,2978

Cuadro 4.6: Ajuste de los M.Correlacionados.

*En los modelos correlacionados, se observa que a tamaños muestrales pequeños respecto del ajuste, no existe mayor diferencia entre modelo normal y modelo slash. Presenta mejor ajuste el modelo Normal a medida que aumenta el tamaño muestral, no obstante no se presenta mayor diferencia entre el DIC y el BIC, entre estos tres. Al igual que en los modelos independientes acá también presenta menor complejidad el modelo t-Student [cuadro 4.6].*

Modelo Normal						
Parámetro		Nº de obs.	Media	Desv.est	IC	
$\mu_1$		30	-99,3877044	0,45398866	-100,2560714	-98,504286
	-99,4423273	60	-99,3523233	0,334586806	-100,0424554	-98,713297
		100	-99,4423273	0,251272083	-99,8365625	-98,8625908
$\mu_2$		30	61,3499335	0,59097806	60,1529033	62,4933898
	60,9660504	60	61,1933865	0,409530573	60,3979072	62,0214361
		100	60,9660504	0,298129737	60,2258098	61,3790401
$\sigma_1^2$		30	0,6599319	0,496876053	0,4092369	1,3483149
	0,9	60	0,7312971	0,380649159	0,4110076	1,0607111
		100	0,9160923	0,290553642	0,3919102	0,8756341
$\sigma_2^2$		30	0,6331947	0,475496425	0,4042124	1,433638
	0,95	60	0,7552028	0,216245069	0,4104446	1,1931813
		100	0,9689479	0,135116394	0,3835611	0,9529592
$c_1$		30	0,4849789	0,269277069	0,2336315	0,9357693
	0,7	60	0,5316998	0,213632441	0,2369089	0,7571468
		100	0,6403724	0,23204144	0,2103819	0,6054983
$c_2$		30	0,4664187	0,325223062	0,2498717	1,1012962
	0,6	60	0,5496373	0,222625066	0,2540786	0,9050174
		100	0,6752505	0,135596645	0,2434916	0,7466204

Cuadro 4.7: Estimaciones con Modelo Normal

Se observa nuevamente que en el modelo normal, a mayor muestral, mejoran las estimaciones, disminuye el error y la amplitud del intervalo de credibilidad. Cabe mencionar que existen algunos resultados no deseables pero esperados en los modelos *t-Student* y *slash*, con número de observaciones bajas, en el caso de las varianzas y covarianzas en los cuales, se produce una mayor diferencia con el valor verdadero, de igual manera aumenta su precisión [cuadro 4.7], [cuadro 4.8], [cuadro 4.9].



Modelo T-student						
Parámetro	Nº de obs.	Media	Desv.est	IC		
$\mu_1$		30	-99,1823711	0,641901616	-100,3735173	-97,8819423
	-99,4423273	60	-98,9803076	0,420805623	-99,7901049	-98,1523699
		100	-99,404674	0,259079766	-99,9504046	-98,9303442
$\mu_2$		30	59,7709879	0,628226425	58,5638597	61,0310074
	60,9660504	60	60,8220986	0,388038498	60,024791	61,5408557
		100	60,5782231	0,340136698	59,9068715	61,2510049
$\sigma_1^2$		30	1,0972895	1,261788849	0,2386109	2,347862
	0,9	60	0,9729415	0,420100321	0,3488181	1,785174
		100	0,9160923	0,380649156	0,3348848	1,656621
$\sigma_2^2$		30	1,0598901	0,491689129	0,3648663	1,934941
	0,95	60	0,958282	0,40782983	0,35908	1,755087
		100	0,9689479	0,475496425	0,3270357	1,861962
$c_1$		30	0,682589	0,343030319	0,2244977	1,349511
	0,7	60	0,6793544	0,307296713	0,2287144	1,273935
		100	0,6403724	0,269277069	0,2177027	1,163145
$c_2$		30	0,7389347	0,344017732	0,2380541	1,391009
	0,6	60	0,6819921	0,305641718	0,240162	1,313654
		100	0,6752505	0,325223062	0,2173254	1,322658

Cuadro 4.8: Estimaciones con Modelo t-Student

Modelo Slash						
Parámetro	Nº de obs.	Media	Desv.est	IC		
$\mu_1$		30	-98,9044124	1,200327963	-100,7013121	-96,1793628
	-99,4423273	60	-99,3370909	0,493016383	-100,2343337	-98,4008593
		100	-99,6675969	0,363483155	-100,3434144	-98,9758289
$\mu_2$		30	61,4203289	1,084188768	59,0784753	63,3865383
	60,9660504	60	61,3381981	0,446207717	60,4538272	62,2160573
		100	61,7508032	0,410868302	60,934814	62,4943143
$\sigma_1^2$		30	0,4352705	0,28224686	0,2967604	1,2374038
	0,9	60	0,5513363	0,180480276	0,2441414	0,9029326
		100	0,7178619	0,148743874	0,3379621	0,90349
$\sigma_2^2$		30	0,5951107	0,881801264	0,3118787	1,5145882
	0,95	60	0,5318914	0,203086509	0,2895154	1,0405312
		100	0,8628064	0,205654467	0,2443104	0,9245549
$c_1$		30	0,5130177	0,195679841	0,2156859	0,8882537
	0,7	60	0,3954976	0,131732077	0,1989911	0,7680023
		100	0,32777	0,110293109	0,1738227	0,7319258
$c_2$		30	0,3979216	0,25845487	0,2273909	1,0979921
	0,6	60	0,4346079	0,146166823	0,1956235	0,7084513
		100	0,6066564	0,147570526	0,1784186	0,7092763

Cuadro 4.9: Estimaciones con Modelo Slash



El modelo normal correlacionado pasa todos los test de convergencia, no así los modelos *t*-Student y slash los cuales para tamaños muestrales pequeños falla el test de Raftery and Lewis, en el caso de varianzas y covarianzas. Estos parámetros en general son sensibles en sus estimaciones y por consiguiente en establecer su convergencia el los procedimientos que los estiman.

#### 4.2.2. Densidades a Posteriori

El siguiente gráfico es obtenido con 100 observaciones.

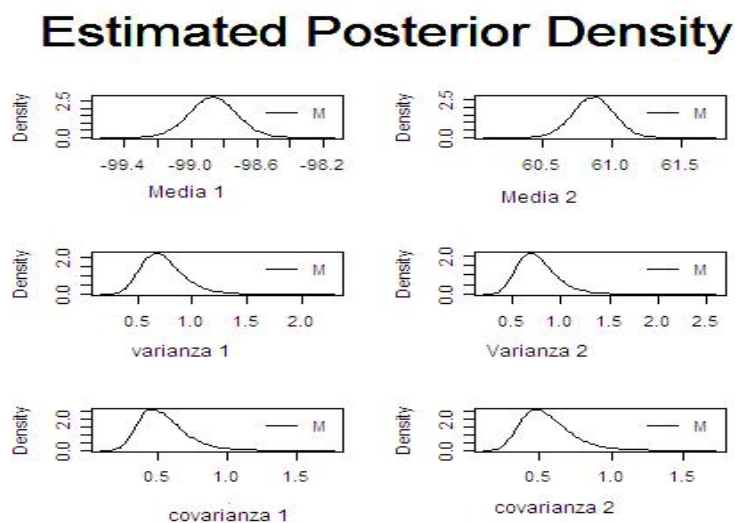


Figura 4.4: Densidades a Posteriori en el Modelo Normal

La forma de las distribuciones a posteriori estimadas, son de la misma familia que en el caso de los modelos Independientes, respectivamente para los parámetros de comparación.

### 4.2.3. Calibración del procedimiento

		Modelos Correlacionados		
Datos	Procesos	30	60	100
Normales	Normal v/s T-student	60 - 83	60 - 80	80 - 100
	Normal v/s Slash	62 - 95	90 - 100	90 - 100
T-student	Normal v/s T-student	55 - 93	60 - 100	70 - 100
Slash	Normal v/s Slash	45 - 70	55 - 85	61 - 100

Cuadro 4.10: Calibración Modelos Correlacionados

*En comparación con los modelos independientes, la calibración de datos normales disminuye inferiormente de un 80 por ciento a un 60 por ciento aproximadamente para número de observaciones pequeñas, mejorando al aumentar el tamaño muestral en los caso normales y t-Student y levemente en datos slash. Esto podría considerarse como una debilidad del proceso, no obstante la presencia de correlación ha de influir en la calidad de ajuste por la complejidad del sistema en si, y dado que para tamaños muestrales grandes (sobre 100) la calibración es sobre el 60 por ciento, esta se manifiesta como aceptable.*

#### *Conclusiones Generales del Capítulo.*

- 1. La realización de las simulación permite llevar a terreno resultados teóricos y poder contrastarlos con lo obtenido, no siempre siendo muy simple su aplicabilidad.*
- 2. Las simulaciones permite visualizar distintos escenarios y/o poder corregir condiciones iniciales que se contraponen al sistema, es decir, condiciones que podrían darse en forma arbitraria de las prioris, además*

un número elevado en el proceso de jerarquización del procedimiento Bayesiano, podría también hacer colapsar el sistema, por la dificultad de controlar un gran número de variables o eventos involucrados. Esto incide en la reformulación del diseño inicial y dar una propuesta final.

3. *Las estimaciones en ambos escenarios, en lo que se refiere a los modelos Independientes versus modelos correlacionado son buenas, el primero gana en su exactitud y facilita la convergencia de las cadenas. No obstante cada uno posee su beneficio de acuerdo al problema que se tenga en cuestión.*
  
4. *Los cuadros de autocorrelación y correlación (en el anexo 3), nos indican que a medida que aumenta el tamaño del salto disminuye la dependencia, y entre las variables de interés a mayor tamaño muestral disminuye la correlación.*

## **CONCLUSIONES GENERALES**

*Una conclusión principal que se puede establecer, es que el uso de técnicas Bayesianas permiten resolver con cierta comodidad un problema complejo. En el presente trabajo, se realizó un análisis bayesiano para poder obtener estimaciones de parámetros desconocidos, en una primera etapa bajo el supuesto básico de no correlación y como segunda etapa bajo el supuesto de correlación entre observaciones de subpoblaciones independientes entre ellas, no obstante, se realizaron ambos casos para establecer comparaciones.*

*En todos los modelos expuestos, se utilizaron distribuciones a priori usuales (familia conjugada) para los parámetros de interés, como también para las variables utilizadas para realizar el proceso, dentro de las cuales están las distribuciones normales para modelar medias y la distribución gamma o Igamma para modelar varianzas.*

*En el modelo slash hubo que aplicar truncación estándar en la implementación de Gibbs Sampling y aproximaciones asintóticas en la obtención de la devianza para medir el ajuste, esta última fue extraída del artículo de Kenneth Lange and Janet S. Sinsheimer (1993). Los cuales se ajustaron de buena manera a los procedimientos.*

*Los modelos asociados a los datos fueron, el Normal, *t*-Student y slash, dando un grado de flexibilidad al proceso. Se obtuvieron las funciones de verosimilitud y las ecuaciones para aplicar Gibbs Sampling, esto último dada la dificultad analítica en obtener las distribuciones a posteriori de los parámetros. De esta manera se pudo desarrollar la implementación de los algoritmos establecidos y así mediante el muestreo de las densidades a posteriori generadas por las cadenas vía simulación, obtener las estimaciones de los estimadores de máxima verosimilitud(estimadores de Bayes).*

*Las estimaciones obtenidas mediante los supuestos de independencia y bajo correlación son igualmente satisfactorias, claro está que los modelos normales bajo independencia, tienden a ser más óptimas, principalmente cuando el tamaño muestral aumenta.*

*Los resultados de calibración de los modelos independientes resultaron adecuados, no obstante en los modelos bajo correlación dicha calibración se presenta menos eficiente. Cabe hacer notar que el objetivo es detectar correlación y/o estimar correlaciones entre observaciones, por lo cual el mecanismo se vuelve más complejo en su implementación y por tanto las simulaciones deben ser más trabajosas en el proceso.*

*Dentro de un proceso de Inferencia usual, aplicada a datos reales, uno podría estar interesado en determinar la independencia de las observaciones (no obstante depende de la investigación en si), claro está, que para hacer análisis de correlación entre observaciones en el método clásico, habría que tener muchas réplicas de ellas, lo cual no es usual por diferentes motivos. Así el enfoque Bayesiano nos permite determinar o analizar la existencia o no de dicha correlación con la misma muestra de observaciones.*

## **POYECCIONES FUTURAS**

*Bajo el mismo esquema incluir un modelo normal contaminado.*

*La génesis de esta tesis, estuvo motivada en poder realizar inferencia sobre poblaciones finitas, con una estructura de correlación bajo el enfoque Bayesiano. Es así, que queda como trabajo futuro realizar una propuesta al respecto de modelamiento e inferencia sobre el tamaño muestral bajo esta estructura.*

*Aplicar los procedimientos dados en esta tesis en paralelo con el método de aceptación y rechazo de MCMC a conjuntos de datos reales y realizar las comparaciones con sus fortalezas y debilidades.*

*Trabajar los modelos propuestos desde el punto de vista de modelos de efectos aleatorios y ampliar los resultados hacia modelos lineales de tal manera de poder realizar predicciones bajo una estructura de correlción entre observaciones.*

*Presentar una propuesta de modelamiento con covarianza negativa.*

*Realizar un análisis de sensibilidad de los modelos, utilizando diferentes tipos de prioris.*

# Bibliografía

- [1] Aggarwal. *Sample surveys: inference and analysis, Volumen 2.*
  
- [2] Ananth Ranganathan.(20th September 2004). *The Dirichlet Process Mixture (DPM) Model.*
  
- [3] Arijit Chaudhuri and Raghunth Arnab.(1982). *On Unbiased Variance-Estimation with Various Multi-stage Sampling Strategies. The Indian Journal of Statistics, Vol.44, Series B, Pt.1. pp.92-101. Indian Statistical Institute.*
  
- [4] J. Bazán, C. Bayes(2010). *Inferencia Bayesiana en Modelos de Regresion Binaria Usando BRMUW. Reporte de Investigación. Serie B. Nro 25. Departamento de Ciencias.PUCP.*
  
- [5] Berger, James O.(1985). *Statistical Decision Theory and Bayesian Analysis. Segunda Edición, Springer-Verlag.*
  
- [6] Bellhouse. D. R. (1987). *Model- Based Estimation in Finite Population Sampling. The American Statistician, Vol. 41, N°4(Nov., 1987). pp. 260-262.*
  
- [7] Bolfarine, H., Zacks, S. (1992). *Prediction Theory For Finite Populations- A Case Study Approach. Springer-verlag.*

- [8] *Burnham, K.P., Anderson, D.R 1998. Model selection and inference. Springer.*
- [9] *Bush. C. and MacEachern. S.(1996). A semiparametric Bayesian model for randomised block designs. Biometrika 83(2):275-285.*
- [10] *Cannon MJ, Warner L, Taddei JA, Kleinbaum DG.(2001). What can go wrong when you assume that correlated data are independent: an illustration from the evaluation of a childhood health intervention in Brazil. Statistics in Medicine 20(9-10):1461-7.*
- [11] *Cnaan, A., Laird, NM., Slasor P.(1997). Using the general linear mixed model to analyse unbalanced repeated measures and longitudinal data. Statistics in Medicine 16(20):2349-2380.*
- [12] *Caseros Antonio.(2010). Factores Modulares de la Percepción de la Calidad Docente. Journal of Educational Research, Assessment and Evaluation, Vol.71.Nº2.pp 90-94.*
- [13] *Casanova Laudien María Paz.(2005). Análisis Bayesiano Semiparamétrico del problema de calibración en Modelos de Regresión Elípticos. Tesis de Doctorado en Estadística. Pontificia Universidad Católica de Chile.*
- [14] *Chib, S., y Greenberg, E. (1995). Understanding the Metropolis-Hasting algorithm. Ann.Math.Statist. 49: 327-335.*
- [15] *Cochran, W. G. Sampling Techniques. New York: John Wiley and Sons, 1963.*
- [16] *De Finetti,B.(1937).La prévision: ses lois logiques, ses sources subjectives. Ann. Inst. H. Poincaré, 7, 1-68.*



(1938) *Sur la condition d'equivalence partielle. Act. Sci. Ind.*, 739

- [17] Draper. David, James S. Hodges, Colin L. Mallows, Daryl Pregibon. *Exchangeability and Data Analysis. Journal of the Royal Statistical Society. Series A (Statistics in Society)*, Vol. 156, N°1.(1993).pp.9-37.
  
- [18] Escobar, M. D. y West, M. (1995). *Bayesian density estimation and inference using mixtures. Journal of American Statistical Association* 85, 378-409.
  
- [19] Esteves, L.G. Wechsler, S. and Iglesias. P.L. *Some characterizations of uniform models. Statistics Probability Letters, Volume 69, Issue 4, October 2004, Pages 397-404.*
  
- [20] Ericson, W.A.(1969.a). *Subjective Bayesian Models in Sampling Finite Populations. Journal of the Royal Statistical Society, Series B, (Methodological)*, Vol.31, N°2. pp 195-233.
  
- [21] Edward W. Frees. (2004). *Longitudinal and Panel Data. Analysis and Applications in the Social Sciences. University of Wisconsin-Madison.*
  
- [22] Gaudard, M. and Hadwin, D. (1989). *Sigma-algebras on spaces of probability measures. Scand. J. Statist.*, 16(2):169-175.
  
- [23] Gelfand, A., y Smith, A. (1990). *Sampling-Based Approaches to Calculating Marginal Densities. Journal of the American Statistical Association*, 85, pp. 398-409.
  
- [24] Geman, S., y Geman, D. (1984). *Stochastic relaxation, Gibbs distribution, and the bayesian restoration of image. IEEE transactions on*

*pattern analysis and machine intelligence*, 6:721-741.

- [25] Geyer, CS. (1992). *Practical Markov Chain Monte Carlo*. *Stat.Sci.* 7:473-511.
- [26] Ghosh, and Ramamoorthy, M.(2003). *Bayesian Nonparametrics*, Springer, New York.
- [27] Godambe, V.P. *A unified theory of sampling from finite populations*. *Journal of the Royal Statistical Society (B)*, 17 (1955) 268-278.
- [28] Goldstein,H.(2002). *Multilevel statistical models (3rd ed.)*. London: Arnold.
- [29] Hartley.H.O. and R.L Sielken(1975): *A superpopulation viewpoint for finite population sampling*. *Biometrics*, 31,411-422.
- [30] Hasting, wk. (1970). *Monte Carlo Sampling Methods Using Markov Chain and Their Applications*. *Biometrika*, 57:97-109.
- [31] Hewitt, E. and Savage, L.J. (1995): *Symetric measures on Cartesian products*. *Trans. American Statistician*, 30, 188-9.
- [32] Hill,B.M.(1968). *Posterior distribution of percentiles: Bayes'theorem for sampling from a population*. *Journal of the American Statistical Association*,63,677-691.
- [33] Ita Kreft and Jaude Leeuw(jun 18,1998). *Introduction to multilevel Modeling*.

- [34] Jara. A, García.M.J, Lesaffre. E.(October 2006). A Dirichlet process mixture model for the analysis of correlated binary responses. *Computational Statistics & Data Analysis* 51(2007) 5402-5415.
- [35] Jeffreys, H. (1961).*Theory of probability*. Oxford University Press, Oxford, U.K.
- [36] Jing Wang, Marc G.Genton.(2004). The multivariate skew-slash distribution. *Journal of statistical Planning and Inference*. 136(2006) 209-220.
- [37] Kenneth Lange and Janet S.Sinsheimer.(1993). *Normal/independent Distributions and Their Applications in Robust Regression*. American Statistical Association, Institute of Mathematical Statistics and Interface Foundation of North America. Volumen 2, Number 2, Pages 175-198.
- [38] Kish, L. *Survey Sampling*. New York: John Wiley and Sons,1965.
- [39] Laud, P.W., E Ibrahim, J. G. (1995). Predictive model selection. *J. Roy. Statist. Soc.Ser.B*, 57, pp. 247-262.
- [40] Leary AC, Donnan PT, MacDonald TM, Murphy MB. (2000). The influence of physial activity on the variability of ambulatory blood pressure. *American Journal of the Hypertension*. 13(10):1067-73.
- [41] Lindley, D.V. y Smith, A.F.M.(1972).Bayes estimates for the linear model(with discussion). *Journal of the Royal Statistical Society B* 34, 1-42.
- [42] MacEachern, S.(1994). Estimating normal means a conjugate style Dirchlet process prior. *Communications in Statistics: Simulation and*

*Computation* 23, 727-741.

- [43] Mendoza, M. & Nieto-Barajas, L. E.(2006). *Bayesian solvency analysis with autocorrelated observations. Applied Stochastic Models in Business and Industry* 22:169-180.
  
- [44] Monsalve, G. N. (2005). *Modelaje Bayesiano Espacio-Temporal de Factores Asociados con la Incidencia del Dengue en el área de Maracay. Tesis de Maestría en Estadística de la Universidad Simón Bolívar.*
  
- [45] Morgenthaler, S., Tukey, J. W. (1991). *Configural Polysampling: A Route to Practical Robustness.* Wiley, New York.
  
- [46] Morris, H. Degroot. *Probabilidad y Estadística. Segunda edición.* Addison Wesley.
  
- [47] Müller, P., Erkanli, A. y West, M. (1996). *Bayesian curve fitting using multivariate normal mixtures. Biometrika* 83, 67-79.
  
- [48] Ortega Irizo, Fco. Javier y Basulto Santos, Jesús. (2009). *Estimación Bayesiana en Modelos de Producción con Frontera determinista.*
  
- [49] Orellana Zapata Yasna. (2007). *Análisis Bayesiano No paramétrico utilizando Procesos Skew Dirichlet. Tesis de Doctorado en Estadística. Pontificia Universidad Católica de Chile.*
  
- [50] Quintana, F. and Iglesias, P. (2003). *Bayesian clustering and product partition models. . Pontificia Universidad Católica de Chile, Santiago Chile. J.R. Statist. Soc. B (2003) 65, Part 2, pp.557-574.*

- [51] *Rasftery, A.E., Kass, R. (1995). Bayes Factors. Journal of the American Statistical Association, Vol.90, No.430, pp.773-795.*
- [52] *Robert, C.P. (1992). L'analyse Statistique Bayésienne. Economica, Paris France. pp 109-111, 336.*
- [53] *Roderick J. Little and Hui Zheng. (April 5, 2006). The Bayesian Approach to the Analysis of Finite Population Surveys. Department of Biostatistics, Universidad of Michigan.*
- [54] *Rodriguez, J. (2009). Modelaje de la vulnerabilidad de la población Venezolana a eventos de lluvia extrema. Tesis de maestría en Estadística de la Universidad Simón Bolívar, Departamento de Cómputo Científico y Estadística.*
- [55] *Scott, A.J. and Smith, T.M.F. Estimation in Multi-Stage Surveys. Journal of the American Statistical Association, Vol.64, N° 327. (Sep., 1969), pp. 830-840.*
- [56] *Smith, T.M.F (1976): The foundations of survey sampling a review. Journal of The Royal Statistics Society, Series A(139), 183-204. (1984): Present position and potential developments- Some personal views about sample surveys. Journal of The Royal Statistics Society, Series A(147), 208-221.*
- [57] *Spielgelhalter, D. J., Best, N., Carlin, B. P., y Van Der Linde, A. (2002). Bayesian measures of model complexity and fit (with discussion). J Roy. Statist. Soc., Ser. B, 64, pp. 583-639.*
- [58] *Stigler, S.M. (1986). The history of statistics the measurement of uncertainty before 1900. Harvard University Press, Cambridge, MA.*

- [59] Thomson, R. 1986. *Estimation of realized heritability in a selected population using mixed model methods*. *Génét. Sél. Evol.* 18(4): 475-484.
- [60] Tierney. (1994). *Markov Chains for Exploring Posterior Distributions (with discussion)*. *Annals of Statistics*, 22,1701-1762.
- [61] West, M. y Turner, D. A. (1994). *Deconvolution of mixtures in analysis of neural synaptic transmission*. *Statistician* 43, 31-43.

## Anexo.

### A.1 Distribuciones condicionales.

#### A.1.1 Modelos Independientes (capítulo 2).

Distribuciones condicionales del Modelo Normal Heterocedástico (2.6).

$$\begin{aligned}
 & \Pi[\mu_i \mid y_i, \sigma_i, u_i, \nu, \delta^2] \\
 & \propto \Pi[y_i \mid \mu_i] \cdot \Pi[\mu_i \mid \nu, \delta^2] \\
 & \propto \prod_{j=1}^{n_i} \exp \left\{ -\frac{1}{2} \left[ \frac{y_{ij} - \mu_i}{\sigma_i} \right]^2 \right\} \cdot \exp \left\{ -\frac{1}{2} \left[ \frac{\mu_i - \nu}{\delta} \right]^2 \right\} \\
 & \propto \exp \left\{ -\frac{1}{2} \left\{ \frac{[n_i \cdot \mu_i^2 - 2\mu_i \cdot n_i \cdot \bar{y}_i]}{\sigma_i^2} + \frac{\mu_i^2 - 2\mu_i \cdot \nu}{\delta^2} \right\} \right\} \\
 & \propto \exp \left\{ -\frac{1}{2} \left\{ \left( \frac{n_i}{\sigma_i^2} + \frac{1}{\delta^2} \right) \left[ \mu_i - \left\{ \frac{\frac{n_i \cdot \bar{y}_i + \nu}{\sigma_i^2 + \delta^2}}{\frac{n_i}{\sigma_i^2} + \frac{1}{\delta^2}} \right\} \right]^2 \right\} \right\} \\
 & \therefore \mu_i \mid y_i, \sigma_i^2, \nu, \delta^2 \sim N(A_i, B_i).
 \end{aligned}$$

$$\begin{aligned}
 & \Pi[\sigma_i^2 \mid \mu_i, y_i, \nu, \delta^2, a_i, b_i] \\
 & \propto \frac{1}{\sqrt{(\sigma_i^2)^{n_i}}} \cdot \prod_{j=1}^{n_i} \exp \left\{ -\frac{1}{2} \cdot u_i \cdot \left[ \frac{y_{ij} - \mu_i}{\sigma_i} \right]^2 \right\} \cdot (\sigma_i^2)^{-(a_i+1)} \cdot \exp(-b_i/\sigma_i^2) \\
 & \propto (\sigma_i^2)^{-(a_i+1+n_i/2)} \cdot \exp \left\{ -\frac{1}{2} u_i \cdot (\sum_{j=1}^{n_i} (y_{ij} - \mu_i)^2 + 2b_i) / \sigma_i^2 \right\} \\
 & \therefore \sigma_i^2 \mid \mu_i, y_i, a_i, b_i \sim IG(a_i + n_i/2, B_i).
 \end{aligned}$$

Distribuciones condicionales del Modelo T-student (2.8).

$$\begin{aligned}
& \Pi[\mu_i \mid y_i, \sigma_i, u_i, \nu, \delta^2] \\
& \propto \Pi[y_i \mid \mu_i] \cdot \Pi[\mu_i \mid \nu, \delta^2] \\
& \propto \prod_{j=1}^{n_i} \exp \left\{ -\frac{1}{2} \left[ \frac{y_{ij} - \mu_i}{\sigma_i} \right]^2 \right\} \cdot \exp \left\{ -\frac{1}{2} \left[ \frac{\mu_i - \nu}{\delta} \right]^2 \right\} \\
& \propto \exp \left\{ -\frac{1}{2} \left\{ \frac{[n_i \mu_i^2 - 2\mu_i \cdot n_i \cdot \bar{y}_i]}{\sigma_i^2} + \frac{\mu_i^2 - 2\mu_i \cdot \nu}{\delta^2} \right\} \right\} \\
& \propto \exp \left\{ -\frac{1}{2} \left\{ \left( \frac{n_i}{\sigma_i^2} + \frac{1}{\delta^2} \right) \left[ \mu_i - \left\{ \frac{\frac{n_i \cdot \bar{y}_i}{\sigma_i^2} + \frac{\nu}{\delta^2}}{\frac{n_i}{\sigma_i^2} + \frac{1}{\delta^2}} \right\} \right]^2 \right\} \right\} \\
& \therefore \mu_i \mid y_i, \sigma_i^2, \nu, \delta^2 \sim N(A_i, B_i).
\end{aligned}$$

$$\begin{aligned}
& \Pi[\sigma_i^2 \mid \mu_i, y_i, \nu, \delta^2, a_i, b_i] \\
& \propto \frac{1}{\sqrt{(\sigma_i^2)^{n_i}}} \cdot \prod_{j=1}^{n_i} \exp \left\{ -\frac{1}{2} \cdot u_i \cdot \left[ \frac{y_{ij} - \mu_i}{\sigma_i} \right]^2 \right\} \cdot (\sigma_i^2)^{-(a_i+1)} \cdot \exp(-b_i/\sigma_i^2) \\
& \propto (\sigma_i^2)^{-(a_i+1+n_i/2)} \cdot \exp \left\{ -\frac{1}{2} u_i \cdot (\sum_{j=1}^{n_i} (y_{ij} - \mu_i)^2 + 2b_i) / \sigma_i^2 \right\} \\
& \therefore \sigma_i^2 \mid \mu_i, y_i, a_i, b_i \sim IG(a_i + n_i/2, E_i).
\end{aligned}$$

Con,

$$E_i = \frac{u_i \cdot \|y_i - \mu_i \cdot 1_{n_i}\|^2}{2} + b_i, \quad \|y_i - \mu_i \cdot 1_{n_i}\|^2 = \sum_{j=1}^{n_i} (y_{ij} - \mu_i)^2$$



$$\begin{aligned}
& \Pi[u_i \mid \mu_i, y_i, \sigma_i^2, \varphi] \\
& \propto \frac{1}{\sqrt{(u_i^{-1})^{n_i}}} \cdot \prod_{j=1}^{n_i} \exp \left\{ \frac{-1}{2} \cdot u_i \cdot \left[ \frac{y_{ij} - \mu_i}{\sigma_i} \right]^2 \right\} \cdot u_i^{\varphi/2-1} \cdot \exp(-\varphi \cdot u_i/2) \\
& \propto u_i^{(n_i+\varphi)/2-1} \cdot \exp \left\{ -\frac{1}{2} u_i \cdot \left( \sum_{j=1}^{n_i} (y_{ij} - \mu_i)^2 / \sigma_i^2 + \varphi \right) \right\} \\
& \therefore u_i \parallel \mu_i, y_i, \sigma_i^2, \varphi \sim G \left( (\varphi + n_i)/2, \frac{\sigma_i^{-2} \cdot \|y_i - \mu_i \cdot \mathbf{1}_{n_i}\|^2 + \varphi}{2} \right).
\end{aligned}$$

*Distribuciones condicionales del Modelo Slash (2.10).*

$$\begin{aligned}
& \Pi[\mu_i \mid y_i, \sigma_i, u_i, \nu, \delta^2] \\
& \propto \Pi[y_i \mid \mu_i] \cdot \Pi[\mu_i \mid \nu, \delta^2] \\
& \propto \prod_{j=1}^{n_i} \exp \left\{ -\frac{1}{2} \left[ \frac{y_{ij} - \mu_i}{\sigma_i} \right]^2 \right\} \cdot \exp \left\{ \frac{-1}{2} \left[ \frac{\mu_i - \nu}{\delta} \right]^2 \right\} \\
& \propto \exp \left\{ -\frac{1}{2} \left\{ \frac{[n_i \cdot \mu_i^2 - 2\mu_i \cdot n_i \cdot \bar{y}_i]}{\sigma_i^2} + \frac{\mu_i^2 - 2\mu_i \cdot \nu}{\delta^2} \right\} \right\} \\
& \propto \exp \left\{ -\frac{1}{2} \left\{ \left( \frac{n_i}{\sigma_i^2} + \frac{1}{\delta^2} \right) \left[ \mu_i - \left\{ \frac{\frac{n_i \cdot \bar{y}_i + \nu}{\sigma_i^2 + \delta^2}}{\frac{n_i}{\sigma_i^2} + \frac{1}{\delta^2}} \right\} \right]^2 \right\} \right\} \\
& \therefore \mu_i \mid y_i, \sigma_i^2, \nu, \delta^2 \sim N(A_i, B_i).
\end{aligned}$$

$$\begin{aligned}
& \Pi[\sigma_i^2 \mid \mu_i, y_i, \nu, \delta^2, a_i, b_i] \\
& \propto \frac{1}{\sqrt{(\sigma_i^2)^{n_i}}} \cdot \prod_{j=1}^{n_i} \exp \left\{ \frac{-1}{2} \cdot u_i \cdot \left[ \frac{y_{ij} - \mu_i}{\sigma_i} \right]^2 \right\} \cdot (\sigma_i^2)^{-(a_i+1)} \cdot \exp(-b_i/\sigma_i^2) \\
& \propto (\sigma_i^2)^{-(a_i+1+n_i/2)} \cdot \exp \left\{ -\frac{1}{2} u_i \cdot (\sum_{j=1}^{n_i} (y_{ij} - \mu_i)^2 + 2b_i) / \sigma_i^2 \right\} \\
& \therefore \sigma_i^2 \mid \mu_i, y_i, a_i, b_i \sim IG(a_i + n_i/2, B\hat{i}).
\end{aligned}$$

$$\begin{aligned}
& \Pi[u_i \mid \mu_i, y_i, \sigma_i^2, \varphi] \\
& \propto \frac{1}{\sqrt{(u_i^{-1})^{n_i}}} \cdot \prod_{j=1}^{n_i} \exp \left\{ \frac{-1}{2} \cdot u_i \cdot \left[ \frac{y_{ij} - \mu_i}{\sigma_i} \right]^2 \right\} \cdot u_i^{\varphi-1} \\
& \propto u_i^{(n_i/2+\varphi-1)} \cdot \exp \left\{ -\frac{1}{2} u_i \cdot \sum_{j=1}^{n_i} (y_{ij} - \mu_i)^2 / \sigma_i^2 \right\} \\
& \therefore u_i \mid \mu_i, y_i, \sigma_i^2, \varphi \sim GT \left( \varphi + n_i/2, \frac{\sigma_i^{-2} \cdot \|y_i - \mu_i \cdot 1_{n_i}\|^2}{2} \right).
\end{aligned}$$

### A.1.2 Modelos Correlacionados (capítulo 3).

Distribución condicional del Modelo Normal Heterocedástico (3.1).

$$\begin{aligned}
& \Pi[\mu_i \mid y_i, \sigma_i, u_i, c_i, \nu, \delta^2] \\
& \propto \Pi[y_i \mid \mu_i, \sigma_i^2] \cdot \Pi[\mu_i \mid \nu, \delta^2] \\
& \propto \exp \left\{ -\frac{1}{2} [y_i - (\mu_i \cdot 1_{n_i} + w_i)]^t \cdot [u_i^{-1} \sigma_i^2 I_{n_i}]^{-1} \cdot [y_i - (\mu_i \cdot 1_{n_i} + w_i)] \right\} \exp \left\{ \frac{-1}{2\delta^2} (\mu_i - \nu)^2 \right\} \\
& \propto \exp \left\{ -\frac{1}{2} u_i \sigma_i^{-2} [\mu_i^2 \cdot 1_{n_i}^t \cdot 1_{n_i} - 2\mu_i (1_{n_i}^t \cdot (y_i - w_i))] \right\} \cdot \exp \left\{ \frac{-1}{2\delta^2} (\mu_i^2 - 2\mu_i \nu) \right\} \\
& \propto \exp \left\{ -\frac{1}{2} [u_i \sigma_i^{-2} \cdot 1_{n_i}^t \cdot 1_{n_i} + \frac{1}{\delta^2}] - 2\mu_i (u_i \sigma_i^{-2} (1_{n_i}^t (y_i - w_i)) + \frac{\nu}{\delta^2}) \right\} \\
& \therefore \mu_i \mid y_i, \sigma_i^2, w_i, u_i, \nu, \delta^2 \sim N(A_i, B_i).
\end{aligned}$$

Donde,

$$A_i = \frac{n_i \cdot \delta^2 \cdot u_i \cdot (\bar{y}_i - \bar{w}_i) + \nu \cdot \sigma_i^2}{n_i \cdot u_i \cdot \delta^2 + \sigma_i^2}, \quad B_i = \frac{\sigma_i^2 \cdot \delta^2}{u_i \cdot n_i \cdot \delta^2 + \sigma_i^2}$$

$$\Pi[w_{ij} \mid y_{ij}, \mu_i, \sigma_i, u_i, c_i]$$

$$\propto \exp \left\{ -\frac{1}{2} u_i \cdot \sigma_i^{-2} (y_{ij} - (\mu_i + w_{ij}))^2 \right\} \cdot \exp \left\{ -\frac{u_i w_{ij}^2}{2c_i} \right\}$$

$$\propto \exp \left\{ -\frac{1}{2} (w_{ij}^2 (u_i \sigma_i^{-2} + u_i/c_i) - 2w_{ij} (u_i \sigma_i^{-2} (y_{ij} - \mu_i))) \right\}$$

$$\therefore w_{ij} \mid y_{ij}, \sigma_i^2, u_i, \sim N(\kappa_{ij}, \tau_{ij}).$$

Con,

$$\kappa_{ij} = \frac{(y_{ij} - \mu_i) \cdot c_i}{c_i + \sigma_i^2}, \quad \tau_{ij} = \frac{\sigma_i^2 \cdot c_i}{u_i \cdot (c_i + \sigma_i^2)}$$

$$\Pi[\sigma_i^2 \mid \mu_i, y_i, w_i, a_i, b_i]$$

$$\propto \frac{1}{\sqrt{(\sigma_i^2)^{n_i}}} \cdot \prod_{j=1}^{n_i} \exp \left\{ \frac{-1}{2} \cdot u_i \cdot \left[ \frac{y_{ij} - (\mu_i + w_{ij})}{\sigma_i} \right]^2 \right\} \cdot (\sigma_i^2)^{-(a_i+1)} \cdot \exp(-b_i/\sigma_i^2)$$

$$\propto (\sigma_i^2)^{-(a_i+1+n_i/2)} \cdot \exp \left\{ -\frac{1}{2} u_i \cdot (\sum_{j=1}^{n_i} (y_{ij} - (\mu_i + w_{ij})))^2 + 2b_i \right\} / \sigma_i^2$$

$$\therefore \sigma_i^2 \mid \mu_i, y_i, w_i, a_i, b_i \sim IG(a_i + n_i/2, B_i).$$

$$\begin{aligned}
& \Pi[c_i | w_i, e_i, f_i, u_i] \\
& \propto \frac{1}{\sqrt{(c_i)^{n_i}}} \cdot \exp \left\{ \frac{-1}{2} \cdot \frac{u_i \|w_i\|^2}{c_i} \right\} \cdot c_i^{-(e_i+1)} \cdot \exp \left\{ -\frac{f_i}{c_i} \right\} \\
& \therefore c_i | w_i, u_i, e_i, f_i \sim IG(e_i + n_i/2, f_i + 0,5 \cdot u_i \|w_i\|^2).
\end{aligned}$$

*Distribución condicional del Modelo T-student (3.2).*

*Las distribuciones condicionales son de la misma forma que el modelo anterior sólo se agrega la variable aleatoria  $u_i$ .*

$$\begin{aligned}
& \Pi[u_i | y_i, \mu_i, w_i, \varphi] \\
& \propto \frac{1}{\sqrt{(u_i^{-1})^{n_i}}} \cdot \exp \left\{ -u_i \cdot \left[ \frac{\sum_{j=1}^{n_i} (y_{ij} - (\mu_i + w_{ij}))^2}{2\sigma_i^2} \right] \right\} \cdot \frac{1}{\sqrt{(u_i^{-1})^{n_i}}} \cdot \exp \left\{ -\frac{u_i \|w_i\|^2}{2c_i} \right\} \cdot u_i^{\varphi/2-1} \cdot \exp \left\{ -\frac{u_i \cdot \varphi}{2} \right\} \\
& \propto u_i^{(\varphi/2+n_i-1)} \cdot \exp \left\{ -\frac{u_i}{2} \cdot \left[ \frac{\sum_{j=1}^{n_i} (y_{ij} - (\mu_i + w_{ij}))^2}{\sigma_i^2} + \frac{\|w_i\|^2}{c_i} + \varphi \right] \right\} \\
& \therefore u_i | y_i, \mu_i, w_i, \varphi \sim GT \left( n_i + \varphi/2, \frac{1}{2} \cdot \left[ \frac{\sum_{j=1}^{n_i} (y_{ij} - (\mu_i + w_{ij}))^2}{\sigma_i^2} + \frac{\|w_i\|^2}{c_i} + \varphi \right] \right).
\end{aligned}$$

*Distribución condicional del Modelo Slash (3.3).*

*Las distribuciones condicionales son de la misma forma que el modelo anterior sólo se agrega una forma distinta para la variable aleatoria  $u_i$ , dado que ahora distribuye Beta.*

$$\begin{aligned}
& \Pi[u_i | y_i, \mu_i, w_i, \varphi] \\
& \propto \frac{1}{\sqrt{(u_i^{-1})^{n_i}}} \cdot \exp \left\{ -u_i \cdot \left[ \frac{\sum_{j=1}^{n_i} (y_{ij} - (\mu_i + w_{ij}))^2}{2\sigma_i^2} \right] \right\} \cdot \frac{1}{\sqrt{(u_i^{-1})^{n_i}}} \cdot \exp \left\{ -\frac{u_i \|w_i\|^2}{2c_i} \right\} \cdot u_i^{\varphi-1} \\
& \propto u_i^{(\varphi+n_i-1)} \cdot \exp \left\{ -\frac{u_i}{2} \cdot \left[ \frac{\sum_{j=1}^{n_i} (y_{ij} - (\mu_i + w_{ij}))^2}{\sigma_i^2} + \frac{\|w_i\|^2}{c_i} \right] \right\} \\
& \therefore u_i | y_i, \mu_i, w_i, \varphi \sim IG \left( n_i + \varphi, \frac{1}{2} \cdot \left[ \frac{\sum_{j=1}^{n_i} (y_{ij} - (\mu_i + w_{ij}))^2}{\sigma_i^2} + \frac{\|w_i\|^2}{c_i} \right] \right).
\end{aligned}$$

## A.2 Convergencias (capítulo 4).

### A.2.1 Modelos Independientes.

#### Tabla y Gráfico.

Como se observa en el cuadro 1, los tres test indican convergencia de las cadenas con tamaño muestral de datos de 30 observaciones, en el modelo Normal. De igual manera en la figura 1, el gráfico muestra que al muestrear de las cadenas con largo 10, 20, 30, y 40 la autocorrelación disminuye (o no es significativa).

#### Convergencia

Modelo Normal			
Convergencia con tamaño muestral $n_i = 30, 60, 100$			
Parámetro	RAFTERY AND LEWIS(f. de Ind)	HEIDLEBERGER AND WELCH(test)	GEWEKE (p-value)
$\mu_1$	pasa	pasa	0.7121097
	pasa	pasa	0.2556318
	pasa	pasa	0.2556318
$\mu_2$	pasa	pasa	0.219858
	pasa	pasa	0.5068565
	pasa	pasa	0.5068565
$\sigma_1^2$	pasa	pasa	0.1493023
	pasa	pasa	0.07715655
	pasa	pasa	0.07715655
$\sigma_2^2$	pasa	pasa	0.97108743
	pasa	pasa	0.5106960
	pasa	pasa	0.5106960

Cuadro 11: T.Convergencia

Modelo T-student			
Convergencia con tamaño muestral $n_i = 30, 60, 100$			
Parámetro	RAFTERY AND LEWIS(f. de Ind)	HEIDLEBERGER AND WELCH(test)	GEWEKE (p-value)
$\mu_1$	pasa	pasa	0.3864864
	pasa	pasa	0.94547763
	pasa	pasa	0.1935353
$\mu_2$	pasa	pasa	0.1955869
	pasa	pasa	0.5927890
	pasa	pasa	0.98652846
$\sigma_1^2$	pasa	pasa	0.7734136
	no pasa	pasa	0.08866585
	no pasa	pasa	0.4657585
$\sigma_2^2$	no pasa	pasa	0.8814455
	no pasa	pasa	0.3004784
	no pasa	pasa	0.2037630

Cuadro 12: T.Convergencia

<u>Modelo Slash</u>			
Convergencia con tamaño muestral $n_i = 30, 60, 100$			
Parámetro	RAFTERY AND LEWIS(f. de Ind)	HEIDLEBERGER AND WELCH(test)	GEWEKE (p-value)
$\mu_1$	pasa	pasa	0.9933788
	pasa	pasa	0.1667637
	pasa	pasa	0.7895374
$\mu_2$	pasa	pasa	0.7991444
	pasa	pasa	0.3140531
	pasa	pasa	0.5853741
$\sigma_1^2$	no pasa	pasa	0.4821030
	pasa	pasa	0.9986424
	pasa	pasa	0.2142107
$\sigma_2^2$	pasa	pasa	0.2792106
	no pasa	pasa	0.3892408
	pasa	pasa	0.4516853

Cuadro 13: T.Convergencia

HEIDLEBERGER AND WELCH STATIONARITY AND INTERVAL HALFWIDTH TESTS:									
=====									
Halfwidth test accuracy = 0.1									
Chain: M1									
	Stationarity Test	Keep	Discard	C-von-M	Halfwidth Test	Mean	Halfwidth		
Media 1	passed	4000	0	0.1115906	passed	-99.1066850	Media 1 0.004484772		
Media 2	passed	4000	0	0.0981895	passed	60.9333244	Media 2 0.005930112		
Varia 1	passed	4000	0	0.1131593	passed	0.5446681	Varia 1 0.002832889		
Varia 2	passed	4000	0	0.1011837	passed	1.2104712	Varia 2 0.007057295		
RAFTERY AND LEWIS CONVERGENCE DIAGNOSTIC:									
				Thin	Burn-in	Total	Lower Bound	Dependence	Factor
Quantile = 0.025	Media 1	1	2	3790	3746	1.0117459			
Accuracy = +/- 0.005	Media 2	1	2	3711	3746	0.9906567			
Probability = 0.95	Varia 1	1	2	3711	3746	0.9906567			
Chain: M1	Varia 2	1	2	3635	3746	0.9703684			
GEWEKE CONVERGENCE DIAGNOSTIC:									
Fraction in first window = 0.1									
Fraction in last window = 0.5									
Chain: M1									
	Media 1	Media 2	Varia 1	Varia 2					
Z-Score	-0.3690242	1.226906	1.4420003	0.03624446					
p-value	0.7121097	0.219858	0.1493023	0.97108743					

Cuadro 14: Test de convergencia en el Modelo Normal



## Sampler Lag-Autocorrelations

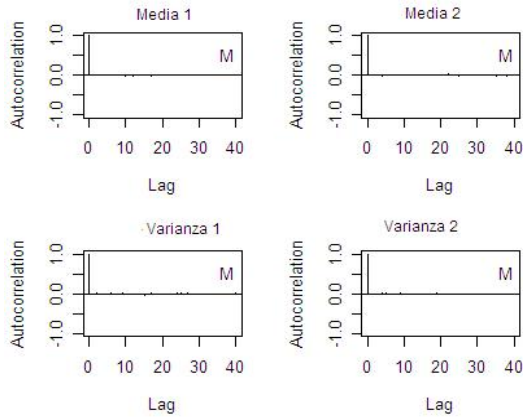


Figura 5: Gráfico de autocorrelaciones en el Modelo Normal

### A.2.2 Modelos Correlacionados.

Tablas y Gráfico.

convergencia

<b>Modelo Normal.</b>			
<b>Convergencias con tamaño muestral</b>			
<b><math>n_i = 30, 60, 100</math></b>			
<b>Parámetro</b>	<b>RAFTERY AND LEWIS(f. de Ind)</b>	<b>HEIDLEBERGER AND WELCH(test)</b>	<b>GEWEKE (p-value)</b>
$\mu_1$	pasa	pasa	0.3179406
	pasa	pasa	0.2267473
	pasa	pasa	0.1150825
$\mu_2$	pasa	pasa	0.06349525
	pasa	pasa	0.4249357
	pasa	pasa	0.2538732
$\sigma_1^2$	pasa	pasa	0.1824621
	pasa	pasa	0.7967284
	pasa	pasa	0.93176485
$\sigma_2^2$	pasa	pasa	0.1654755
	pasa	pasa	0.5800649
	pasa	pasa	0.1892235
$c_1$	pasa	pasa	0.5225430
	pasa	pasa	0.7797671
	pasa	pasa	0.3663374
$c_2$	pasa	pasa	0.7389219
	pasa	pasa	0.6329745
	pasa	pasa	0.95026029

Cuadro 15: Convergencia en el Modelo Normal

<b>Modelo Slash</b>			
<b>Convergencias con tamaño muestral <math>n_i = 30, 60, 100</math></b>			
<b>Parámetro</b>	<b>RAFTERY AND LEWIS(f. de Ind)</b>	<b>HEIDLEBERGER AND WELCH(test)</b>	<b>GEWEKE (p-value)</b>
$\mu_1$	pasa	pasa	0.3387888
	pasa	pasa	0.6065286
	pasa	pasa	0.8437012
$\mu_2$	pasa	pasa	0.7610131
	pasa	pasa	0.6672558
	pasa	pasa	0.254999
$\sigma_1^2$	pasa	pasa	0.4541123
	pasa	pasa	0.6798954
	pasa	pasa	0.3305433
$\sigma_2^2$	pasa	pasa	0.9341066
	pasa	pasa	0.1493284
	pasa	pasa	0.4184906
$c_1$	pasa	pasa	0.9341066
	pasa	pasa	0.09103674
	pasa	pasa	0.7329114
$c_2$	No pasa	pasa	0.8642549
	pasa	pasa	0.3264349
	pasa	pasa	0.398907

Cuadro 17: Convergencia en el Modelo Slash

<b>Modelo T-student</b>			
<b>Convergencias con tamaño muestral <math>n_i = 30, 60, 100</math></b>			
<b>Parámetro</b>	<b>RAFTERY AND LEWIS(f. de Ind)</b>	<b>HEIDLEBERGER AND WELCH(test)</b>	<b>GEWEKE (p-value)</b>
$\mu_1$	pasa	pasa	0.94751484
	pasa	pasa	0.3482938
	pasa	pasa	0.9202863
$\mu_2$	pasa	pasa	0.2420177
	pasa	pasa	0.1852611
	pasa	pasa	0.6800999
$\sigma_1^2$	No pasa	pasa	0.7375292
	No pasa	pasa	0.1548646
	No pasa	pasa	0.127637
$\sigma_2^2$	No pasa	pasa	0.5861935
	No pasa	pasa	0.6984626
	No pasa	pasa	0.2079626
$c_1$	pasa	pasa	0.95440717
	pasa	pasa	0.95065449
	pasa	pasa	0.8580590
$c_2$	pasa	pasa	0.8166286
	pasa	pasa	0.7899073
	pasa	pasa	0.3634813

Cuadro 16: Convergencia en el Modelo T-student