

UNIVERSITAT DE VALÈNCIA  
BIBLIOTECA CIÈNCIES

Nº Registr 8258  
DATA 16.XI.95  
SIGNATURA  
160.T.D  
Nº LIBLS: i18950826

.b Matemáticas.

25 cms.

b16763774



# **Análisis Bayesiano de muestras no aleatorias en Poblaciones Finitas**

Tesis presentada para aspirar al grado de Doctor en Matemáticas

**Presentada por:** María Begoña Font Belaire  
**Dirigida por:** María Jesús Bayarri García

septiembre 1995



UMI Number: U607172

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



UMI U607172

Published by ProQuest LLC 2014. Copyright in the Dissertation held by the Author.  
Microform Edition © ProQuest LLC.

All rights reserved. This work is protected against  
unauthorized copying under Title 17, United States Code.



ProQuest LLC  
789 East Eisenhower Parkway  
P.O. Box 1346  
Ann Arbor, MI 48106-1346

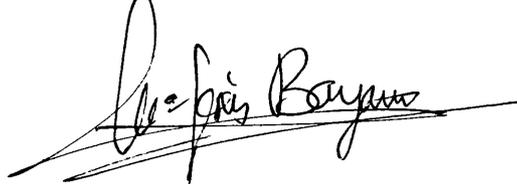


DÑA MARIA JESUS BAYARRI GARCIA, profesora titular de Estadística e Investigación Operativa de la Universitat de València.

CERTIFICA: que la presente memoria *Análisis Bayesiano de muestras no aleatorias en Poblaciones Finitas* ha sido realizada bajo mi dirección, en el Departamento de Estadística e Investigación Operativa por Dña. Begoña Font Belaire, y constituye su tesis para optar al grado de Doctor en Matemáticas.

Y para que conste, en cumplimiento de la legislación vigente, presento ante la Facultad de Matemáticas de la Universitat de València, a 1 de septiembre de 1995.

LA DIRECTORA

A handwritten signature in black ink, appearing to read 'María Jesús Bayarri García', written over a horizontal line.

María Jesús Bayarri García



**A mis padres  
por su cariño y apoyo**



# Prólogo

Durante estas últimas décadas se han ido haciendo muy corrientes las encuestas como procedimiento para obtener información acerca de las características físicas o sociológicas de determinadas poblaciones a partir de los datos de un número de elementos de la población (muestra), en general muy inferior al total. Y como resultado de este auge se han publicado numerosos trabajos sobre cómo inferir acerca de una característica global de la población a partir de una muestra, basados en dos aproximaciones al problema, la conocida como población fija y los modelos de superpoblación.

Esta tesis constituye una aproximación al problema basada en modelos de superpoblación tratados desde un punto de vista Bayesiano. Los contenidos de esta memoria están organizados en cinco capítulos y tres apéndices, cuyos contenidos describiremos brevemente a continuación.

**Capítulo 1** Presenta una recopilación de los trabajos publicados sobre poblaciones finitas, en la que se presentan los resultados más conocidos para las aproximaciones de población fija y modelos de superpoblación clásicos y se hace hincapié en los resultados basados en la aproximación mediante modelos de superpoblación Bayesianos.

**Capítulos 2 y 3** Se dedican a estudiar de manera exhaustiva un grupo bastante amplio de modelos de superpoblación Bayesianos (que generalizan algunos modelos estudiados por la bibliografía) indicando las situaciones de modelización en las que resultan adecuados.

**Capítulo 4** Se dedica a comparar los resultados inferenciales obtenidos cuando nos apoyamos en muestras aleatorias, estratificadas o cluster de la población.

**Capítulo 5** Propone tres modelos jerárquicos para rutas aleatorias que se revelan como una herramienta útil para analizar en qué medida una muestra obtenida por rutas puede ser empleada para establecer inferencias sobre una población como si se tratara de una muestra aleatoria de la misma.

**Apéndice A** Recoge las notaciones generales aplicadas en esta tesis.

**Apéndice B** Contiene algunos resultados sobre matrices, funciones y distribuciones, que han sido de utilidad para la demostración de los resultados presentados en esta tesis.

**Apéndice C** Presenta un resumen de los métodos computacionales, concretamente el algoritmo de Gibbs Sampling implementado mediante el programa BUGS, aplicados para el estudio de los modelos no conjugados planteados en esta tesis.

Las poblaciones simuladas para el estudio de los modelos de esta tesis y las muestras obtenidas a partir de las mismas aparecen recogidas en el disquette que acompaña esta tesis (léase el archivo LEEME para ver la organización de dicho disquette).

Quiero aprovechar estas líneas para agradecer el apoyo prestado en la realización de este trabajo por el Departamento de Estadística e Investigación Operativa de la Universitat de València y en especial por la Dra. María Jesús Bayarri sin cuyo apoyo, consejo y dirección no hubiera sido posible la realización de esta tesis.

# Índice

Prólogo	v
<b>1 Estimación en poblaciones finitas: una recopilación</b>	<b>1</b>
1.1 Introducción	1
1.2 Análisis de una Población Finita desde el punto de vista de Población Fija	4
1.2.1 Criterios para seleccionar el estimador	7
1.2.2 Estimadores de la media poblacional	8
1.2.3 Comentarios de la sección	11
1.3 Modelos de Superpoblación. Soluciones Clásicas	12
1.3.1 Criterios para seleccionar el estimador	15
1.3.2 Estimadores de la media poblacional	17
1.3.3 Comentarios de la sección	21
1.4 Modelos de Superpoblación. Soluciones Bayesianas	23
1.4.1 Modelos Bayesianos con distribuciones no totalmente especificadas. Estimadores Bayes lineales	26
1.4.2 Modelos Bayesianos para distribuciones específicas. Estimadores Bayes y Estimadores Bayes empíricos	31
1.4.3 Estimadores Bayes constreñidos	41
1.4.4 Modelos “estrictamente” Bayesianos	42
1.4.5 Comentarios de la sección	42
1.5 Comentarios finales	44
<b>2 Modelos de Superpoblación Bayesianos I: Modelos en una jerarquía</b>	<b>46</b>
2.1 Introducción	46
2.2 Modelos Bayesianos en una etapa	48
2.2.1 El modelo Bayes de Regresión Normal	49
2.2.2 El modelo Bayes de Regresión Normal-Gamma	51
2.2.3 El modelo Bayes de Regresión con cociente entre varianzas desconocido	53
2.2.4 Modelos Bayes en una etapa mínimo informativos	58
2.3 Modelos Bayesianos en dos etapas con una jerarquía	60
2.3.1 Modelo Bayes de Regresión en dos etapas con una jerarquía Normal	61
2.3.2 Modelo Bayes de Regresión en dos etapas con una jerarquía Normal- Gamma	64

2.3.3	Modelo Bayes de Regresión en dos etapas con una jerarquía con co-	67
	ciente entre varianzas desconocido . . . . .	
2.3.4	Modelos Bayes en dos etapas con una jerarquía mínimo informativos .	74
2.4	Ejemplos para modelos Bayesianos en una jerarquía . . . . .	77
2.4.1	Población 1A: modelos Bayesianos en una etapa . . . . .	80
2.4.2	Población 1B: modelos Bayesianos en dos etapas con una jerarquía . .	85
2.4.3	Conclusiones de la sección . . . . .	94
2.5	Comentarios finales: limitaciones de los modelos en una jerarquía . . . . .	95
<b>3</b>	<b>Modelos de Superpoblación Bayesianos II: Modelos en dos o más jerar-</b>	
	<b>quías</b> . . . . .	<b>96</b>
3.1	Introducción . . . . .	96
3.2	Modelos Bayesianos en dos etapas con dos jerarquías . . . . .	97
3.2.1	Modelo Bayes de Regresión en dos etapas con dos jerarquías normal .	97
3.2.2	Modelo Bayes de Regresión en dos etapas con dos jerarquías normal-	
	gamma . . . . .	102
3.2.3	Modelo Bayes de Regresión en dos etapas con dos jerarquías con co-	
	ciente entre varianzas desconocido . . . . .	107
3.2.4	Modelos Bayes en dos etapas con dos jerarquías mínimo informativos .	115
3.3	Ejemplos para modelos Bayesianos en dos etapas con dos jerarquías: otra	
	modelización de la población 1B . . . . .	119
3.4	Modelos Bayesianos $m$ -jerárquicos . . . . .	129
3.4.1	Modelo Bayes de Regresión $m$ -jerárquico Normal . . . . .	129
3.4.2	El modelo Bayes de Regresión $m$ -jerárquico Normal - Gamma . . . . .	132
3.4.3	Modelos Bayes $m$ -jerárquicos mínimo informativos . . . . .	135
3.5	Comentarios finales . . . . .	138
<b>4</b>	<b>Comparación Bayes de las inferencias obtenidas a partir de muestras</b>	
	<b>aleatorias, estratificadas y cluster</b> . . . . .	<b>139</b>
4.1	Introducción . . . . .	139
4.2	Fomulación del problema y análisis Bayes del modelo para muestras aleatorias	141
4.2.1	Formulación del problema . . . . .	141
4.2.2	Análisis Bayes del modelo para muestras aleatorias . . . . .	145
4.3	Análisis Bayes del modelo para muestreos en dos etapas y comparación entre	
	los tres procedimientos de muestreo . . . . .	146
4.3.1	Análisis Bayes del modelo para muestreos en dos etapas . . . . .	146
4.3.2	Comparación entre los tres tipos de muestreo . . . . .	151
4.4	Ejemplo: comparación numérica de los muestreos aleatorio, estratificado y	
	cluster . . . . .	156
4.4.1	Descripción de las poblaciones simuladas y obtención de las muestras	156
4.4.2	Descripción del modelo aplicado para el análisis de las poblaciones	
	simuladas . . . . .	157
4.4.3	Resultados obtenidos . . . . .	157
4.5	Comentarios finales . . . . .	158

<b>5</b>	<b>Análisis Bayesiano de Rutas Aleatorias</b>	<b>160</b>
5.1	Introducción . . . . .	160
5.2	Cinco modelos Bayes para describir una población . . . . .	161
5.2.1	Comentarios de los modelos Bayes para rutas aleatorias . . . . .	166
5.2.2	Modelos Bayes para muestreos aleatorios con varianzas desconocidas . . . . .	167
5.3	Análisis Bayesiano de los modelos propuestos . . . . .	169
5.3.1	Modelo Bayes de rutas aleatorias . . . . .	170
5.3.2	Modelo Bayes de rutas aleatorias con medias iguales . . . . .	181
5.3.3	Modelo Bayes para rutas aleatorias homocedásticas . . . . .	188
5.3.4	Modelos para rutas aleatorias mínimo informativos . . . . .	197
5.3.5	Modelo Bayes para muestras aleatorias con varianzas desconocidas . . . . .	198
5.4	Análisis empírico de los modelos Bayes para rutas . . . . .	200
5.4.1	Descripción de las poblaciones simuladas . . . . .	201
5.4.2	Determinación de las distribuciones iniciales y mecanismos aplicados para la realización de las inferencias . . . . .	208
5.4.3	Exposición y discursión de los resultados para la media no observada y media poblacionales . . . . .	211
5.4.4	Exposición y análisis de las inferencias sobre otras variables de los modelos de rutas aleatorias y sobre las distribuciones predictivas y finales dada la muestra y los tamaños de las rutas muestreadas . . . . .	216
5.5	Comentarios finales . . . . .	219
<b>A</b>	<b>Notación</b>	<b>226</b>
<b>B</b>	<b>Lemas aplicados</b>	<b>229</b>
B.1	Resultados sobre matrices . . . . .	229
B.2	Resultados sobre funciones . . . . .	229
B.3	Resultados sobre distribuciones . . . . .	230
B.4	Resultados sobre la obtención de distribuciones finales y marginales . . . . .	237
<b>C</b>	<b>Métodos computacionales aplicados</b>	<b>244</b>
C.1	Muestreo de la distribución por Gibbs Sampling . . . . .	245
C.1.1	El algoritmo Gibbs Sampling . . . . .	245
C.1.2	Convergencia del algoritmo Gibbs a la distribución final . . . . .	246
C.1.3	Obtención de una muestra de la distribución final, estimación de sus primeros momentos y representación gráfica de la misma . . . . .	247
C.1.4	Mecanismos para la simulación de variables aleatorias . . . . .	249
C.2	El programa BUGS (Bayesian Inference using Gibbs Sampling) . . . . .	250
C.2.1	Especificación del modelo . . . . .	251
C.2.2	Descripción del funcionamiento . . . . .	252
C.2.3	Diagnosis de la convergencia y obtención de resultados . . . . .	253
C.3	Aplicación del programa BUGS al análisis de los modelos Bayesianos no conjugados sobre poblaciones finitas estudiados en esta tesis . . . . .	254
	<b>Bibliografía</b>	<b>256</b>



# Capítulo 1

## Estimación en poblaciones finitas: una recopilación

### 1.1 Introducción

Desde la década de los sesenta hasta nuestros días el problema de cómo estimar una cantidad de interés de una población finita a partir del conocimiento de los valores de una o varias características de las unidades de una muestra de la misma, ha sido extensamente tratado desde diversos puntos de vista. El objetivo de este capítulo es hacer una recopilación de los resultados más conocidos de poblaciones finitas y en especial, aquellos basados en asumir un modelo de superpoblación y analizarlo desde un punto de vista Bayesiano (línea de trabajo en la que se haya incluida esta tesis).

Sea  $P = \{1, 2, \dots, N\}$  un conjunto de índices con los que numeramos cada uno de los elementos de una población, con  $N$ , el tamaño de la población, que asumiremos en los cuatro primeros capítulos de esta tesis conocido y en cuatro de los cinco modelos estudiados en el capítulo 5 desconocido, y sea  $\mathbf{y} = (y_1, y_2, \dots, y_N)^t$  el vector de características de la población donde  $y_i$  representa el valor desconocido de una característica del elemento  $i$ -ésimo de la población (algunos autores realizan la generalización de estudiar un vector de características para cada elemento de la población, generalización ésta que no trataremos en esta tesis). Considerada la *población finita* anterior, el objetivo usual es inferir acerca de una determinada función escalar  $g(\mathbf{y})$  del vector de características de la población ; para ello se recurre a seleccionar una muestra  $s = \{i_1, i_2, \dots, i_n\}$  de  $n$  elementos de la población en los que se mide la característica en estudio, que permite distinguir entre  $\mathbf{y}_s = (y_{i_1}, y_{i_2}, \dots, y_{i_n})^t$ , el vector poblacional observado, que es conocido y  $\mathbf{y}_u = (y_i : i \notin s)^t$  el vector poblacional no observado, y a la elaboración de un estimador de la función  $g(\mathbf{y})$ .

Un estimador  $\widehat{g(\mathbf{y}_s)}$  de  $g(\mathbf{y})$  es una función escalar de  $s$  y del vector poblacional observado asociado a la muestra  $s$ . Para algunas poblaciones, se conoce para cada elemento  $i$  de la población los valores de un  $p$ -vector  $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{ip})$  de características auxiliares que proporciona información acerca de la característica en estudio  $y_i$ ,  $i = 1, 2, \dots, N$ . En estos casos, denotaremos por  $\mathbf{X} = (\mathbf{x}_1^t, \mathbf{x}_2^t, \dots, \mathbf{x}_N^t)^t$ , a la matriz de características auxiliares



de la población y considerados los índices muestreados  $s$ , por  $\mathbf{X}_s = (\mathbf{x}_{i_1}^t, \mathbf{x}_{i_2}^t, \dots, \mathbf{x}_{i_n}^t)^t$ , a la matriz de características auxiliares correspondiente a elementos muestreados y por  $\mathbf{X}_u = (\mathbf{x}_i^t : i \notin s)^t$ , a la matriz de características auxiliares correspondiente a elementos no muestreados.

En numerosas ocasiones al analizar la población detectamos una *estructura en varias etapas* de la misma (ya sea natural en la población o bien introducida artificialmente por nuestros mecanismos de estudio), de modo que en el caso más sencillo (*estructura en dos etapas*), una unidad está formada por un conjunto de elementos (un ejemplo muy sencillo lo tenemos en las poblaciones finitas humanas clasificadas en unidades familiares formadas a su vez por elementos: padre, madre, hijos, ...). Por ser esta estructura en dos etapas muy extendida creemos conveniente el introducir la notación inicial al respecto:

- $P = \{1, 2, \dots, N\}$ , con  $N$  el número total de elementos de la población,
- $K$  el número de unidades de la población,  $K$  conocido,
- $N_i$  el número de elementos en la unidad  $i$ ,  $i = 1, \dots, K$ ,  $(\sum_{i=1}^K N_i = N)$ ,
- $\mathbf{y} = (\mathbf{y}_1^t, \mathbf{y}_2^t, \dots, \mathbf{y}_K^t)^t$  el vector de características de la población, donde  $\mathbf{y}_i = (y_{i1}, y_{i2}, \dots, y_{iN_i})^t$  es el vector de características de la unidad  $i$ ,  $i = 1, 2, \dots, K$  con  $y_{ij}$  el valor desconocido de la característica del elemento  $j$  de la unidad  $i$  de la población,  $j = 1, 2, \dots, N_i$ ,  $j = 1, 2, \dots, K$ ,
- y en poblaciones para las que se conoce para cada elemento  $(i, j)$  los valores de un  $p$ -vector  $\mathbf{x}_{ij} = (x_{ij1}, x_{ij2}, \dots, x_{ijp})$  de características auxiliares que proporciona información acerca de la característica en estudio  $y_{ij}$ ,  $j = 1, 2, \dots, N_i$ ,  $i = 1, 2, \dots, K$ , denotaremos por  $\mathbf{X}_i = (\mathbf{x}_{i1}^t, \mathbf{x}_{i2}^t, \dots, \mathbf{x}_{iN_i}^t)^t$ , a la matriz de características auxiliares de la unidad  $i$ ,  $i = 1, 2, \dots, K$ , y por  $\mathbf{X} = (\mathbf{X}_1^t, \mathbf{X}_2^t, \dots, \mathbf{X}_K^t)^t$  a la matriz de características auxiliares de toda la población.

Los tamaños de la población ( $N$ ) y de cada unidad ( $N_i$ ,  $i = 1, 2, \dots, K$ ), se asumirán en este capítulo y los capítulos 2, 3 y 4 conocidos, describiéndose en el capítulo 5 algunos modelos con tamaño de población y unidades desconocido.

Considerada la anterior estructura de población finita en dos etapas el objetivo de inferencia usual es estimar una función escalar  $g(\mathbf{y})$  del vector de características poblacional (idéntico objetivo que en la población finita no estructurada), o bien estimar una función escalar  $g_i(\mathbf{y}_i)$  del vector de características de la unidad  $i$ -ésima para  $i = 1, 2, \dots, K$ . En ambos casos a fin de lograr nuestro objetivo recurriremos a la obtención de una muestra de la población  $s$  de  $n$  elementos, que se obtiene, seleccionando  $k$  unidades entre las  $K$  existentes  $\{i_1, i_2, \dots, i_k\}$ , índices muestreados) y dentro de estas  $k$  unidades una muestra  $s_i = \{j_1, j_2, \dots, j_{n_i}\}$  de  $n_i$  elementos entre los  $N_i$  existentes,  $i \in \{i_1, i_2, \dots, i_k\}$  quedando la muestra  $s$  constituida por  $s = \{(i, j) : j \in s_i, i \in \{i_1, i_2, \dots, i_k\}\}$ . En particular distinguiremos dos situaciones, el caso  $k = K$  (situación habitual en los muestreos estratificados) y el caso  $n_i = N_i$ ,  $i \in \{i_1, i_2, \dots, i_k\}$  (habitual en muestreos estratificados). Dada la muestra  $s$  anterior denotaremos por  $\mathbf{y}_{s_i} = (y_{ij} : j \in s_i)^t$  al vector poblacional observado de la unidad  $i$ ,  $i \in \{i_1, i_2, \dots, i_k\}$  (que es conocido),  $\mathbf{y}_u = (y_{ij} : j \notin s_i)^t$  al vector poblacional no

observado de la unidad  $i$ ,  $i = 1, 2, \dots, K$ ,  $\mathbf{y}_s = (\mathbf{y}_{s_i}^t : i \in \{i_1, i_2, \dots, i_k\})^t$  al vector poblacional observado (que es conocido),  $\mathbf{y}_u = (\mathbf{y}_{u_i}^t : i = 1, 2, \dots, K)^t$  al vector poblacional no observado y las generalizaciones análogas referidas a la matriz de características auxiliares en el caso de existencia de información auxiliar en forma de covariables. A partir de los elementos muestreados se elaborará un estimador de la función del vector de características de la población o de la unidad  $i$ ,  $i = 1, 2, \dots, K$  que deseamos estimar, que se denotará tal y como señalábamos mediante el superíndice  $\hat{\phantom{x}}$  sobre la cantidad que estima.

La generalización de las notaciones anteriores para representar estructuras de población en más de dos etapas es obvia, por lo que no consideramos necesario abundar en esta línea descriptiva.

En esta tesis, realizaremos dos restricciones importantes sobre los conceptos de población finita señalados en los párrafos anteriores y sobre las funciones escalares a estimar. Estas restricciones son:

- respecto a la característica en estudio nos restringiremos a las características de población continuas (definidas en  $\mathfrak{R}$ ),
- respecto a las funciones escalares a estimar tendrán como consecuencia de la restricción anterior su imagen en  $\mathfrak{R}$ , restringiendo su estudio a la media poblacional ( $\bar{y}$ ) y a la media de la unidad  $i$ -ésima ( $\bar{y}_i$ ) en el caso de estructuras de población en dos etapas (las funciones más estudiadas en la bibliografía). Las definiciones de estas funciones son:

– para una población finita

$$\bar{y} = \frac{\sum_{i=1}^N y_i}{N}, \quad (\text{media poblacional}),$$

(denotaremos a su estimador por  $\hat{\bar{y}}$ )

– para una población finita estructurada en dos etapas

$$\begin{aligned} \bar{y} &= \frac{\sum_{i=1}^K \sum_{j=1}^{N_i} y_{ij}}{N} \quad (\text{media poblacional}), \\ \bar{y}_i &= \frac{\sum_{j=1}^{N_i} y_{ij}}{N_i} \quad (\text{media de la unidad } i\text{-ésima}), \quad i = 1, \dots, K, \end{aligned}$$

(denotaremos a sus estimadores respectivamente por  $\hat{\bar{y}}$  y  $\hat{\bar{y}}_i$  para  $i = 1, 2, \dots, K$ .)

Considerados los conceptos anteriores, en la literatura observamos planteamientos de resolución distintos que podemos clasificar según dos criterios:

1.-) **Atendiendo al procedimiento de selección de la muestra.**

- **Muestreo aleatorio.** Consiste en seleccionar una muestra de  $n$  elementos de la población de modo que todos los elementos tengan la misma posibilidad de ser seleccionados.

- **Muestreo no aleatorio.** Se produce cuando el muestreo no es aleatorio. Entre ellos tenemos: los muestreos proporcionales al tamaño, el muestreo al azar, conveniente o accidental, los muestreos estratificados, los muestreos cluster, las rutas aleatorias, la elección experta, los muestreos por cuotas de muestreo,...

## 2.-) Atendiendo al tratamiento probabilístico del problema.

- **Métodos de Población Fija.** El vector  $\mathbf{y}$  de características de la población y por tanto  $g(\mathbf{y})$  se consideran valores fijos (no aleatorios) aunque desconocidos (*Población Fija*) y se supone como única fuente de aleatoriedad en el problema la producida al introducir un procedimiento de muestreo aleatorio en la selección de los elementos que van a formar parte de la muestra.
- **Métodos basados en Modelos de Superpoblación.** Se basan en suponer el vector  $\mathbf{y}$  de características de la población en el instante del estudio como una realización de una distribución sobre un vector aleatorio  $\mathbf{Y}$ , lo que equivale a suponer que la población finita en estudio es una realización de una cierta población infinita o superpoblación, de aquí el nombre de estos modelos. Dentro de esta línea de trabajo podemos distinguir dos tendencias:
  - *Métodos basados en Modelos de Superpoblación usando métodos de inferencia clásicos.* Se supone un modelo de superpoblación y se realiza inferencia suponiendo los parámetros del modelo como constantes desconocidas de la población.
  - *Métodos basados en Modelos de Superpoblación usando métodos de inferencia Bayesianos.* Se supone un modelo de superpoblación y se realiza inferencia suponiendo los parámetros del modelo como variables aleatorias a las que se les asigna una distribución.

Este capítulo desarrollará las distintas contribuciones al problema de poblaciones finitas atendiendo a esta segunda clasificación; desarrollando en la sección 1.2 el tratamiento por métodos de población fija, en la sección 1.3 el tratamiento por métodos de superpoblación clásicos y en la sección 1.4 por métodos de superpoblación Bayesianos y planteando en la sección 1.5 una discusión general del capítulo destacando los resultados más importantes.

El objetivo de las secciones 1.2 y 1.3 es servir fundamentalmente de apoyo a la discusión de los modelos de superpoblación Bayesianos objeto de la sección 1.4 y los capítulos 2 y 3, y por lo tanto no pretende realizar una descripción exhaustiva de las otras dos aproximaciones.

## 1.2 Análisis de una Población Finita desde el punto de vista de Población Fija

Dada una población finita representada por el conjunto de índices  $P = \{1, 2, \dots, N\}$  y el vector de características  $\mathbf{y} = (y_1, y_2, \dots, y_N)^t$ , la aproximación de población fija al problema de estimar una función del vector de características de la población  $g(\mathbf{y})$ , asume como única fuente de aleatoriedad del problema la inducida por el mecanismo de muestreo aplicado

para seleccionar la muestra  $s$ . Esto es, si denotamos por  $S$  al conjunto de todas las posibles muestras de  $n$  elementos<sup>1</sup>, una aplicación  $p$  definida sobre  $S$  verificando que:

$$p(s) \geq 0, \quad s \in S$$

$$\sum_{s \in S} p(s) = 1.$$

Atendiendo, a la naturaleza de la aplicación  $p$  podemos destacar los siguientes procedimientos de muestreo:

- **Procedimiento de Muestreo Intencionado.** Es áquel procedimiento de muestreo en el que se selecciona una determinada muestra con probabilidad 1.
- **Procedimiento de Muestreo Informativo.** Son procedimientos de muestreo en los que la probabilidad asociada a incluir o no un elemento en la muestra depende del valor de la característica en estudio para ese elemento.
- **Procedimiento de Muestreo no Informativo.** Son procedimientos de muestreo en los que la probabilidad asociada a incluir o no un elemento en la muestra es independiente del valor de la característica en estudio.

Dentro de los procedimientos de muestreo podemos distinguir, entre los más habituales, los siguientes:

1. **Muestreo Aleatorio Simple sin reemplazamiento.** Consiste en seleccionar una muestra de  $n$  elementos *distintos* de la población de modo que todas las posibles muestras de  $n$  elementos *distintos* tengan la misma probabilidad de ser seleccionadas. Esto es,

$$p(s) = \binom{N}{n}^{-1}, \quad s \in S.$$

El muestreo aleatorio simple es el muestreo aleatorio más aplicado en la práctica por lo que se suele conocer simplemente como muestreo aleatorio o por sus siglas m.a.s.

2. **Muestreo Aleatorio Simple con reemplazamiento.** Consiste en seleccionar una muestra de  $n$  elementos (no necesariamente distintos) de la población con reemplazamiento de modo que todas las posibles muestras de  $n$  elementos tengan la misma probabilidad de ser seleccionadas.
3. **Muestreo con probabilidades desiguales sin reemplazamiento.** Consiste en seleccionar una muestra de  $n$  elementos *distintos*, elemento a elemento, de modo que:

---

<sup>1</sup>En esta tesis, nos restringiremos al estudio de procedimientos de muestreo con tamaño de la muestra fijo  $n$ . Esta hipótesis es muy habitual en la práctica (ver, por ejemplo, Cochran(77)[18]), pero admite lógicamente, la generalización basada en suponer mecanismos de muestreo en los que se pueden seleccionar muestras de distintos tamaños (ver, por ejemplo, Cassel, Särndal y Wretman(77)[17])



$$Pr(\text{elegir } i | s_c = \emptyset) = p_i, \quad i = 1, 2, \dots, N,$$

$$Pr(\text{elegir } i | \text{card}(s_c) = r - 1) = \frac{p_i}{1 - \sum_{i=1}^{r-1} p_i} \quad i = r, \dots, n,$$

donde  $0 \leq p_i \leq 1$ ,  $i = 1, \dots, N$  con  $\sum_{i=1}^N p_i = 1$ ,  $s_c$  representa a la muestra en construcción y  $\text{card}(s_c)$  el número de elementos de la misma. Como ejemplos de estos procedimientos de muestreo señalemos los muestreos con probabilidad proporcional al valor de una característica auxiliar positiva (conocida para todos los elementos de la población), los diseños muestrales con rechazo (*rejective sampling designs*), el diseño de Rao-Hartley-Cochran(62) [83] y los diseños de Hájek-Lahiri-Midzuno-Sen [53,66,73,107].

4. **Muestreo con probabilidades desiguales con reemplazamiento.** Consiste en seleccionar una muestra de  $n$  elementos de la población con reemplazamiento de modo que:

$$Pr(\text{elegir } i) = p_i, \quad i = 1, 2, \dots, N,$$

$$0 \leq p_i \leq 1, \quad i = 1, 2, \dots, N, \quad \sum_{i=1}^N p_i = 1.$$

Un ejemplo de este tipo de muestreo, es el muestreo con probabilidad proporcional a  $x_i$  ( $p_i = \frac{x_i}{\sum_{i=1}^N x_i}$ ,  $i = 1, 2, \dots, N$ ), donde  $x_i > 0$  es el valor (positivo) de una característica auxiliar conocida asociada a  $y_i$ ,  $i = 1, 2, \dots, N$ .

5. **Muestreos en dos etapas.** Se aplican cuando la población está estructurada en dos etapas o bien cuando hay razones económicas, administrativas o de estudio, para elegir este muestreo (en cuyo caso, la población adquiere una estructura en dos etapas sobrevenida). El procedimiento de muestreo (ya descrito en la sección 1.1.), consiste en la obtención en primer lugar de una muestra de unidades y en segundo lugar de una muestra de los elementos de las unidades seleccionadas para ser muestreadas. En general las dos fases de muestreo son independientes y los procedimientos empleados pueden ser iguales o distintos. Dentro de los procedimientos de muestreo en dos etapas señalemos dos casos particulares muy usados en la práctica:

- **El muestreo estratificado.** En el muestreo estratificado la población no está estructurada previamente, adquiriendo una estructura en dos etapas como resultado de una división de la misma realizada por el estadístico atendiendo a un criterio de tipo económico, administrativo o socio-económico. Una vez determinadas las unidades (*estratos*) se obtiene la muestra.

Habitualmente, al realizar la división de la población en estratos, se intenta que los estratos constituyan unidades (que pueden ser heterogéneas entre sí) con elementos (siempre, en relación a la característica en estudio) homogéneos entre sí, el número de estratos sea pequeño y el número de elementos dentro de cada

estratos grande, y la muestra se obtiene tomando  $k = K$  (esto es, muestreando todos los estratos).

- **El muestreo cluster.** El muestreo cluster se aplica cuando la población esta estructurada de manera natural o artificial (por ejemplo por divisiones administrativas, por barrios, ...) en dos etapas y queremos aprovechar esta estructura dada para obtener la muestra. Las unidades reciben en este caso el nombre de *clusters*, la división de la población viene dada, pero en cambio el coste de muestrear a los elementos de cada unidad suele ser pequeño, por lo que habitualmente la muestra se obtiene tomando  $n_i = N_i$ ,  $i \in \{i_1, i_2, \dots, i_k\}$ . En los muestreos cluster el número de clusters suele ser grande y el número de elemento dentro de cada cluster pequeño.

Cuando los procedimientos de selección en las dos fases son muestreos aleatorios sin reemplazamiento independientes estamos ante un muestreo aleatorio en dos etapas.

A continuación describiremos en las próximas subsecciones, los criterios aplicados para la selección de un estimador de la función en estudio y los estimadores habituales para la media poblacional y media de cada unidad. Cerrando esta sección con una reflexión acerca de la misma en su conjunto.

### 1.2.1 Criterios para seleccionar el estimador

Dada una población finita nuestro objetivo es estimar una función escalar  $g(\mathbf{y})$  del vector de características de la población, para lo cual obtenemos una muestra de la misma  $s$  a través de uno de los procedimientos de muestreo descritos, definiendo a continuación un estimador  $\widehat{g}(\mathbf{y}_s)$  de la función  $g(\mathbf{y})$  como una función de  $s$  y del vector de características muestral  $\mathbf{y}_s = (y_{i_1}, y_{i_2}, \dots, y_{i_n})^t$  asociado a  $s$ . Si denotamos por:

$$E_p(\widehat{g}(\mathbf{y}_s)) = \sum_{s \in S} \widehat{g}(\mathbf{y}_s) p(s),$$

$$Var_p(\widehat{g}(\mathbf{y}_s)) = E_p[(\widehat{g} - E_p(\widehat{g}))^2] = \sum_{s \in S} [\widehat{g}(\mathbf{y}_s) - E_p(\widehat{g})]^2 p(s),$$

$$ECM_p(\widehat{g}(\mathbf{y}_s)) = E_p[(\widehat{g} - g)^2] = \sum_{s \in S} [\widehat{g}(\mathbf{y}_s) - g(\mathbf{y})]^2 p(s),$$

los criterios más aplicados para la selección de un estimador  $\widehat{g}(\mathbf{y}_s)$  de  $g(\mathbf{y})$  desde el punto de vista de población fija son:

- **Estimadores Insesgados.** Diremos que  $\widehat{g}(\mathbf{y}_s)$  es un estimador insesgado de  $g(\mathbf{y})$  según el procedimiento de muestreo  $p$  si:

$$E_p(\widehat{g}(\mathbf{y}_s)) = g(\mathbf{y}).$$

- **Estimadores Lineales.** Diremos que  $\widehat{g}(\mathbf{y}_s)$  es un estimador lineal de  $g(\mathbf{y})$  si:

$$g(\widehat{\mathbf{y}}_s) = a_0 + \sum_{i \in s} a_i y_i,$$

donde  $\{a_i\}_{i=0}^n$  son constantes. Si  $a_0 = 0$  el estimador se conoce como *estimador lineal homogéneo*.

- **Estimadores Suficientes.** Un estimador  $g(\widehat{\mathbf{y}}_s)$  de  $g(\mathbf{y})$  es suficiente cuando es función de  $D' = (s', \mathbf{y}'_{s'})$  con  $D'$  una colección suficiente de datos y suficiente minimal cuando la colección es suficiente minimal. (Una colección  $D'$  es suficiente cuando la distribución de  $D = (s, \mathbf{y}_s) | D'$ ,  $\mathbf{y}$  sólo depende del vector de características de la población a través de  $D'$  y es suficiente minimal cuando  $D'$  es la de menor tamaño entre las colecciones suficientes de datos.) Basu y Ghosh(67)[4] justificaron que el dato consistente en los elementos identificados sin conservar el orden ni la multiplicidad al ser muestreadas es suficiente minimal para el vector  $\mathbf{y} = (y_1, \dots, y_N)^t$  de características de la población y por tanto es suficiente minimal para cualquier función  $g(\mathbf{y})$  del vector  $\mathbf{y}$ . El teorema de Rao-Blackwell adaptado al problema de población fija (ver por ejemplo, la adaptación de Ruiz(88)[100]), nos permite obtener una clase bastante amplia de estadísticos suficientes.
- **Estimadores Uniformemente de Mínima Varianza.** La clase de los estimadores insesgados y lineales es bastante amplia lo que condujo a la búsqueda del estimador que fuera óptimo entre ellos, éste es, según un criterio de pérdida cuadrática aquel que minimizara la varianza del estimador entre los estimadores insesgados (criterio equivalente a minimizar el error cuadrático medio del estimador respecto al procedimiento de muestreo) o entre los estimadores lineales; esto es, el estimador uniformemente de mínima varianza. Desgraciadamente, dichos estimadores no existen, así Godambe(55)[48] demuestra la inexistencia de estimadores de mínima varianza entre los estimadores lineales homogéneos y Godambe y Joshi(65)[51] entre los estimadores insesgados.

La no existencia del estimador uniforme de mínima varianza conduce, en general, a la selección de un estimador lineal e insesgado que intuitivamente represente la cantidad a estimar y cuya varianza sea suficientemente pequeña para que los resultados que obtengamos con este estimador sean buenos, otros investigadores (Ruiz(86,88) [99,102]), recurren a criterios de comparación experimentales entre estimadores mediante experimentos piloto o a la consideración adicional de un modelo de superpoblación como criterio adicional (Godambe(55)[48], Cochran(77)[18], Cumberland y Royall(81)[19] y Padmawar(81)[79]).

### 1.2.2 Estimadores de la media poblacional

Nos restringiremos en esa subsección a describir los estimadores más empleados para estimar la media poblacional cuando realizamos una aproximación de población fija al problema de estimación en poblaciones finitas. Los estimadores y sus propiedades más importantes son:

1. **La media muestral.** Dada una muestra  $s$  y el vector muestral  $y_s$ , asociado a la misma, se define como:

$$\hat{y}_{F1} = \bar{y}_s = \frac{\sum_{i \in s} y_i}{n}. \quad (1.1)$$

El estimador media muestral es un estimador suficiente, lineal e insesgado para un muestreo aleatorio sin reemplazamiento y lineal e insesgado para un muestreo aleatorio con reemplazamiento.

2. **Estimador de razón.** Se aplica en aquellas situaciones prácticas en las que asociado a cada elemento de la muestra conocemos el valor asociado de una característica auxiliar relacionada con la característica en estudio. Si denotamos por  $x_i$  al valor de una característica auxiliar asociada al elemento  $i$ -ésimo conocida para todos los elementos de la población, dada una muestra  $s$  de la población y  $y_s$ , el vector de características muestral asociado a la muestra  $s$  se define el estimador de razón como:

$$\hat{y}_{F2} = \frac{\bar{x} \bar{y}_s}{\bar{x}_s}, \quad (1.2)$$

donde  $\bar{y}_s$  es la media muestral definida en (1.1),  $\bar{x}_s = \frac{\sum_{i \in s} x_i}{n}$  y  $\bar{x} = \frac{\sum_{i=1}^N x_i}{N}$ . El estimador razón es lineal, suficiente y ligeramente sesgado con sesgo de orden  $\frac{1}{n}$  para un muestreo aleatorio sin reemplazamiento (ver Cochran(77)[18]), y es insesgado para los diseños de Hájek, Lahiri, Midzuno y Sen[53,66,73,107]

3. **El estimador de Hansen-Hurwitz(43)[57].** Se define dada una muestra  $s$  obtenida mediante un procedimiento de muestreo con reemplazamiento con vector de características muestral asociado  $y_s$ , como:

$$\hat{y}_{F3} = \sum_{i \in s} \frac{y_i}{N n p_i}, \quad (1.3)$$

donde  $p_i = Pr(\text{elegir } i)$  para  $i = 1, 2, \dots, N$ . El estimador de Hasen-Hurwitz es lineal e insesgado para todo modelo con probabilidades desiguales con reemplazamiento.

4. **El estimador de Horvitz-Thompson(52)[62].** Se define dada una muestra  $s$  obtenida mediante un procedimiento de muestreo sin reemplazamiento con vector de características muestral asociado  $y_s$ , como:

$$\hat{y}_{F4} = \sum_{i \in s} \frac{y_i}{N \alpha_i}, \quad (1.4)$$

donde  $\alpha_i = Pr(i \in s)$ . El estimador de Horvitz-Thompson es suficiente e insesgado para procedimientos de muestreo sin reemplazamiento con probabilidades de inclusión  $\alpha_i$  estrictamente positivas para  $i = 1, 2, \dots, N$ .

5. **El estimador de regresión.** Se aplica en las situaciones prácticas en las que asociado a cada elemento de la muestra conocemos el valor asociado de una característica auxiliar (situación del estimador razón); si denotamos por  $x_i$  el valor de la característica auxiliar asociada al elemento  $i$ -ésimo,  $s$  a la muestra e  $\mathbf{y}_s$  el vector de características asociado a la misma, el estimador de regresión se define como:

$$\widehat{y}_{F5} = \bar{y}_s + c(\bar{x} - \bar{x}_s), \quad (1.5)$$

donde  $c$  es una constante conocida que expresa la relación lineal entre la característica en estudio  $y_i$  y la característica auxiliar  $x_i$ ,  $i = 1, 2, \dots, N$ ,  $\bar{y}_s$  es la media muestral definida por (1.1),  $\bar{x}_s = \frac{\sum_{i \in s} x_i}{n}$  y  $\bar{x} = \frac{\sum_{i=1}^N x_i}{N}$ . El estimador de regresión es insesgado para un muestreo aleatorio sin reemplazamiento. Si la constante  $c$  es desconocida (lo más habitual), se recurre al estimador siguiente (también conocido como estimador de regresión):

$$\begin{aligned} \widehat{y}_{F6} &= \bar{y}_s + \widehat{c}(\bar{x} - \bar{x}_s), \\ \widehat{c} &= \frac{\sum_{i \in s} (y_i - \bar{y}_s)(x_i - \bar{x}_s)}{\sum_{i=1}^n (x_i - \bar{x}_s)^2}, \end{aligned} \quad (1.6)$$

donde  $\widehat{c}$  es un estimador de la cantidad en este caso desconocida  $c$ . El estimador  $\widehat{y}_{F6}$ , es ligeramente sesgado para un muestreo aleatorio sin reemplazamiento con sesgo proporcional a  $\frac{1}{n}$  (ver Cochran(77)[18]).

6. **El estimador de Ruiz(88)[101].** Se define dada una muestra  $s$  obtenida mediante un procedimiento de muestreo definido por la distribución de probabilidades asociada a la variable discreta  $e_i$  con vector de características muestral asociado  $\mathbf{y}_s$  como:

$$\widehat{y}_{F7} = \sum_{i \in s} \frac{y_i}{NE(e_i)} e_i, \quad (1.7)$$

donde la variable aleatoria  $e_i$  toma valores en el conjunto  $\{0, 1, 2, \dots, n\}$  y representa el número de veces que aparece el elemento  $i$  en la muestra para  $i = 1, \dots, N$  y  $E(e_i)$  es el valor esperado de la variable  $e_i$  para  $i = 1, \dots, N$ . El estimador de Ruiz es lineal e insesgado para el muestreo definido por las variables  $e_i$ ,  $i = 1, \dots, N$  y coincide para algunos procedimientos de muestreo con los estimadores de Hansen-Hurwitz(43)[57], Horwitz-Thompson(52)[62] y Sanchez-Crespo(77)[103] y para algunos procedimientos presenta menor varianza que los estimadores antes citados (recordemos la inexistencia de un estimador uniformemente de mínima varianza).

7. **Estimadores para poblaciones muestreadas en dos etapas.** Se aplican para estimar la media poblacional o la media de la unidad  $i$ -ésima, con  $i = 1, 2, \dots, K$ , y suelen ser versiones combinadas de los estimadores anteriores. Destaquemos los trabajos de Cochran(77)[18], Sundberg(83)[114] y Rueda y Arcos(94)[98]. Cochran(77)[18] proporciona estimadores insesgados o ligeramente sesgados para la media poblacional y la

media de la unidad  $i$ -ésima para muestreo aleatorio estratificado sin reemplazamiento (con  $k = K$ ), muestreo aleatorio por clusters sin reemplazamiento (con  $n_i = N_i$ ,  $i \in \{i_1, i_2, \dots, i_k\}$ ) y en general para muestreos en dos etapas con procedimientos de selección de unidades y de elementos independientes. Recientemente, y siguiendo métodos de trabajo distintos, Sundberg(83)[114] estudia estimadores para modelos de muestreo de dos etapas desde una aproximación predictiva, condicionada a los resultados del muestreo, comparándolos con los estimadores razón y Horwitz-Thompson (bajo la hipótesis de un muestreo proporcional al tamaño muestral para una población en dos etapas, que consiste en seleccionar las unidades según un procedimiento de muestreo proporcional al tamaño y los elementos de las unidades seleccionadas mediante un muestreo aleatorio sin reemplazamiento) y Rueda y Arcos(94)[98] presentan un estimador para muestreos estratificados consistente en una combinación lineal de los estimadores de razón en cada estrato.

### 1.2.3 Comentarios de la sección

La aproximación al problema de estimación en poblaciones finitas desde la metodología de población fija adolece de dos problemas importantes:

1.- La ausencia de un estimador óptimo entre los estimadores lineales y los insesgados según el criterio de estimadores uniformemente de mínima varianza, que ha conducido a la elaboración de otros criterios, como son:

- el criterio de admisibilidad, un estimador es admisible cuando no existe otro con menor  $ECM_p$ , se trata de reducir la clase de estimadores a los estimadores admisibles y reducir el problema de selección del estimador óptimo a uno de ellos.
- el criterio minimax, que sustituye la condición de uniformidad en el criterio de minimizar  $ECM_p$ , por otra más débil.
- el criterio de suficiencia lineal aplicado por Godambe(65)[49], que permite deducir la existencia de un único estimador insesgado de la media poblacional en la familia de los estimadores lineales “suficientes” (el criterio de suficiencia es distinto al criterio de suficiencia habitual que definíamos en la subsección de criterios) y homogéneos,

que podemos encontrar perfectamente explicados y referenciados en la tesis de Murgui(82)[76] [capítulo 1] y en Cassel, Särndal y Wretman(77)[17][capítulo 3]. El defecto de estos criterios es la no obtención de un estimador óptimo y en el caso de obtención del mismo su optimalidad restringida a una clase reducida de estimadores (que puede no ser la más adecuada para un problema de estimación en poblaciones finitas dado), lo que conduce a que el problema de obtener el estimador óptimo dentro de esta metodología siga abierto en la actualidad (ejemplo de ello son las contribuciones de Ruiz(86,88)[99,100,101,102]) y la no existencia de un criterio de selección unanimemente defendido por todos.

2.- La estimación en poblaciones finitas dentro de la metodología de población fija se basa en el procedimiento de muestreo elegido y por lo tanto su validez depende de que la muestra haya sido obtenida de acuerdo con dicho procedimiento de muestreo y a que sea representativa de la población en estudio (ver el ejemplo de *Jumbo* de Basu(71)[3]). En la práctica

es difícil obtener una muestra aleatoria con o sin reemplazamiento de una población recurriéndose a practicar procedimientos aproximativos, así por ejemplo, para obtener una muestra aleatoria con reemplazamiento se recurre a establecer listas numeradas de los elementos de la población y generar  $n$  números aleatorios, los elementos correspondientes a estos números serán los elementos a muestrear (procedimiento que coincide con el muestreo aleatorio con reemplazamiento cuando la lista está aleatoriamente ordenada) y en el caso de no disponer de censos a los muestreos por rutas aleatorias. (Ver Kish(65)[65] y Hess(85)[58] como referencias de este tipo de procedimientos para ajustar los procedimientos de muestreo teóricos a la situación práctica de muestreo de una población real.)

Estas dos objeciones a la aproximación de población fija pueden superarse con la introducción de un modelo de superpoblación tal y como veremos en la próxima sección que trata sobre la aproximación basada en modelos de superpoblación analizados desde un punto de vista clásico.

Destaquemos como referencias fundamentales para el estudio de una población finita desde el punto de vista de población fija (algunas de ellas ya referenciadas) por presentar un planteamiento global del tema, las siguientes: los artículos de Smith(76)[110], Särndal(78)[104] y Hansen, Dalenius y Tepping(85)[56] y los libros de Cochran(77)[18], Cassel, Särndal y Wretman(77)[17] [capítulos 1,2 y 3] y Murgui(82)[76][capítulo 1].

### 1.3 Modelos de Superpoblación. Soluciones Clásicas

Dada una población finita representada por un conjunto de índices  $P = \{1, 2, \dots, N\}$  y el vector de características  $\mathbf{y} = (y_1, y_2, \dots, y_N)^t$ , la aproximación por modelos de superpoblación asume que el vector de características de la población es una realización de un vector aleatorio  $\mathbf{Y} = (Y_1, Y_2, \dots, Y_N)^t$  que se distribuye según la distribución  $\xi$  (a  $\xi$  se le llama *modelo de superpoblación*). De este modo, la función a estimar es función de un vector aleatorio  $g(\mathbf{Y})$  (y como tal aleatoria) y la media poblacional  $\bar{Y}$ , es una variable aleatoria dada por:

$$\bar{Y} = \frac{\sum_{i=1}^N Y_i}{N}.$$

Para estimar la función  $g(\mathbf{Y})$  se recurrirá a la obtención de una muestra  $s = \{i_1, i_2, \dots, i_n\}$  de tamaño fijo  $n$ , para cuyos índices se conocerá la realización del vector poblacional  $\mathbf{Y}$  de modo que  $Y_i = y_i$  para  $i \in s$ . A partir de  $s$  y del vector de características muestral  $\mathbf{y}_s = (y_{i_1}, y_{i_2}, \dots, y_{i_n})^t$  (los índices del vector  $\mathbf{y}_s$  son, en general, relevantes), se construirá el estimador  $g(\mathbf{y}_s)$  de la función  $g(\mathbf{Y})$ , que en el caso de la media poblacional  $\bar{Y}$  denotaremos por  $\hat{\bar{Y}}$ . En las próximas subsecciones describiremos los criterios aplicados más habitualmente desde un *punto de vista clásico*, para la selección del estimador óptimo y los resultados correspondientes en el caso de la estimación de la media poblacional; respecto a los criterios de selección del estimador señalar como avance que se pueden clasificar en dos bloques: los que hemos llamado *criterios de población fija - superpoblación*, que asumen en el proceso de estimación la existencia de dos fuentes de aleatoriedad: la del procedimiento de muestreo y

la del modelo de superpoblación  $\xi$  y los *criterios predictivos* que consideran en el proceso de estimación, la muestra como dada.

Notemos que a lo largo del párrafo anterior, hemos realizado una readaptación de la notación en poblaciones finitas presentada en la sección 1.1. de modo que a partir de ahora y siempre que asumamos un modelo de superpoblación  $\xi$  (en la sección 1.4. trataremos también la aproximación por modelos de superpoblación aunque desde el punto de vista Bayesiano), denotaremos por mayúsculas a las cantidades aleatorias y con minúsculas a los valores realizados de las mismas (que sólo serán conocidos para los elementos muestreados). Este cambio de notación se aplicará también a las notaciones descritas en la sección 1.1. para poblaciones con estructuras en varias etapas. Además, para un modelo de superpoblación dado,  $\xi$  denotaremos por:  $E_\xi(\cdot)$  al operador esperanza según la distribución  $\xi$ ,  $V_\xi(\cdot)$  al operador varianza según  $\xi$  y  $Cov_\xi(\cdot, \cdot)$  al operador covarianza según  $\xi$ .

Los modelos de superpoblación más estudiados pueden clasificarse (atendiendo a las características de la población en:

- **Modelo Clásico Básico.** Pertenecen a este modelo la clase de distribuciones de superpoblación  $\xi$  tales que:

$$Y_i = \mu + e_i, \quad i = 1, \dots, N,$$

donde  $\mu$  es escalar desconocido y  $\mathbf{e} = (e_1, \dots, e_N)^t$  es vector aleatorio de media  $\mathbf{0}_N$  y matriz de varianza-covarianzas  $\mathbf{V}$  desconocida ( $\mathbf{0}_N$   $N$ -vector de ceros y  $\mathbf{V}$  matriz  $N \times N$  definida positiva).

- **Modelo Clásico de Regresión.** Pertenecen a este modelo la clase de distribuciones de superpoblación  $\xi$  para las cuales, asociada a cada variable  $Y_i$  tenemos una información auxiliar representada por un  $p$ -vector  $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})$  para  $i = 1, 2, \dots, N$  conocido, cumpliéndose que:

$$Y_i = \mathbf{x}_i \boldsymbol{\beta} + e_i, \quad i = 1, 2, \dots, N,$$

donde  $\mathbf{X} = (\mathbf{x}_1^t, \mathbf{x}_2^t, \dots, \mathbf{x}_N^t)^t$ ,  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^t$  es vector de parámetros del modelo desconocido y  $\mathbf{e} = (e_1, e_2, \dots, e_N)^t$  es vector aleatorio de media  $\mathbf{0}_N$  y matriz de varianza-covarianzas  $\mathbf{V}$  desconocida ( $\mathbf{0}_N$   $N$ -vector de ceros y  $\mathbf{V}$  matriz  $N \times N$  definida positiva). Dos casos particulares de este modelo son: el modelo clásico básico tomando  $p = 1$ ,  $\mathbf{x}_i = x_i = 1$  para  $i = 1, 2, \dots, N$  y  $\boldsymbol{\beta} = \mu$  y el *modelo clásico simplificado de regresión* que consiste en tomar  $p = 1$ ,  $\mathbf{x}_i = x_i$  para  $i = 1, 2, \dots, N$ ,  $\boldsymbol{\beta} = \beta$  y  $\mathbf{V} = \text{diag}_{i=1, \dots, N} \{\sigma^2 v(x_i)\}$  con  $\beta$  y  $\sigma^2$  desconocidas y  $v(\cdot)$  una función con imagen en los reales positivos conocida ( $v : \mathfrak{R} \rightarrow \mathfrak{R}^+$ ).

- **Modelo Clásico Básico en dos etapas.** Pertenecen a este modelo la clase de distribuciones  $\xi$  tales que para  $j = 1, \dots, N_i$ ,  $i = 1, \dots, K$  :

$$E_\xi(Y_{ij}) = \mu_i,$$

$$Cov_\xi(Y_{ij}, Y_{lm}) = \begin{cases} \sigma_i^2, & i = l, j = m, \\ \rho_i \sigma_i^2, & i = l, j \neq m, \\ 0, & i \neq l, \end{cases}$$

donde  $\mu_i$ ,  $\sigma_i^2$  ( $\sigma_i^2 > 0 \forall i$ ) y  $\rho_i$  son cantidades desconocidas para  $i = 1, 2, \dots, K$ . Usualmente, se asume que las unidades  $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{iN_i})^t$ ,  $i = 1, \dots, K$  constituyen vectores aleatorios independientes según  $\xi$  (*modelo clásico básico independiente en dos etapas*), incorrelados según  $\xi$  (*modelo clásico básico incorrelado en dos etapas*) o intercambiables (veremos seguidamente el concepto) según  $\xi$  (*modelo clásico básico intercambiable en dos etapas*).

- **Modelo Clásico de Regresión Simplificado en dos etapas.** Pertenecen a este tipo de modelo la clase de distribuciones de superpoblación  $\xi$  para las cuales, asociada a cada variable  $Y_{ij}$  tenemos una información auxiliar representada por una cantidad  $x_{ij}$  conocida para  $j = 1, 2, \dots, N_i$ ,  $i = 1, 2, \dots, K$ , de modo que para  $j = 1, \dots, N_i$   $i = 1, \dots, K$ :

$$\begin{aligned} E_{\xi}(Y_{ij}) &= \mu_i x_{ij}, \\ Cov_{\xi}(Y_{ij}, Y_{lm}) &= \begin{cases} \sigma_i^2 v_i(x_{ij}), & i = l, \\ 0, & i \neq l, \end{cases} \end{aligned}$$

donde  $\mu_i$ ,  $\sigma_i^2$  ( $\sigma_i^2 > 0 \forall i$ ) y  $\rho_i$  son cantidades desconocidas y  $v_i(\cdot)$  son funciones con imagen en los reales positivos conocidas ( $v_i : \mathbb{R} \rightarrow \mathbb{R}^+$ ) para  $i = 1, 2, \dots, K$ . Usualmente, se asume que las unidades  $\mathbf{Y}_i = (Y_{i1}, Y_{i2}, \dots, Y_{iN_i})^t$  condicionadas al vector de variables auxiliares  $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{iN_i})$ ,  $i = 1, \dots, K$  constituyen vectores aleatorios independientes según  $\xi$  (*modelo clásico simplificado de regresión independiente en dos etapas*), incorrelados según  $\xi$  (*modelo clásico simplificado de regresión incorrelado en dos etapas*) o intercambiables (veremos seguidamente el concepto) según  $\xi$  (*modelo simplificado clásico de regresión intercambiable en dos etapas*).

A los modelos anteriores se les suelen sumar hipótesis adicionales sobre la distribución  $\xi$  a fin de simplificar el proceso inferencial, estas hipótesis (algunas de ellas ya asumidas directamente al plantear los modelos en dos etapas) se añaden al nombre del modelo y habitualmente son las siguientes:

- **Intercambiabilidad.** Una secuencia finita o infinita numerable,  $Z_1, Z_2, \dots$ , de variables aleatorias es intercambiable si, para  $m \geq 1$  y para todos los índices  $1 \leq i_1 < i_2 < \dots < i_m$ , la distribución del vector  $(Z_{i_1}, Z_{i_2}, \dots, Z_{i_m})^t$  no depende de los subíndices actuales  $i_1, i_2, \dots, i_m$  sino solamente de  $m$ ; en particular, un modelo de superpoblación  $\xi$  es intercambiable si  $Y_1, Y_2, \dots, Y_N$  es una secuencia de variables aleatorias intercambiables. Este concepto introducido por de Finetti(37)[32] y desarrollado por Hewitt y Savage(55)[59] entre otros, aplicado al modelo de superpoblación  $\xi$ , constituye una restricción muy natural para la población en estudio y de ella se deriva que el conocimiento relevante comunicado al estadístico a través de  $s$  y el vector de características muestral  $\mathbf{y}_s$  asociado a  $s$  son los valores de las  $n$  características observadas independientemente de a qué elemento de la población pertenecen.
- **Incorrelación.** Un modelo de superpoblación  $\xi$  es incorrelado cuando las variables  $Y_1, Y_2, \dots, Y_N$  son incorreladas según  $\xi$ , esto es, si  $Cov_{\xi}(Y_i, Y_j) = 0$ , para  $i \neq j$ ,  $i, j = 1, 2, \dots, N$ . La hipótesis de incorrelación permite simplificaciones importantes en el cálculo de las varianzas de los posibles estimadores lineales de  $g(\mathbf{Y})$ .

- **Independencia.** Un modelo de superpoblación  $\xi$  es independiente cuando las variables  $Y_1, Y_2, \dots, Y_N$  son independientes entre sí. Es consecuencia de la independencia la incorrelación.
- **Normalidad.** Un modelo de superpoblación  $\xi$  es normal cuando  $\xi$  es Normal. Recordemos que bajo la hipótesis de normalidad los conceptos de incorrelación e independencia son equivalentes.

(La generalización de los conceptos anteriores para estructuras de población en varias etapas es inmediata señalando en qué nivel de la estructura se aplica la correspondiente hipótesis.)

### 1.3.1 Criterios para seleccionar el estimador

Dada una población finita con  $P = \{1, 2, \dots, N\}$  sobre la que asumimos un modelo de superpoblación  $\xi$  dado, según el cual el vector  $\mathbf{y} = (y_1, y_2, \dots, y_N)^t$  de características de la población es una realización del vector  $\mathbf{Y} = (Y_1, Y_2, \dots, Y_N)^t$  con distribución  $\xi$ , nos planteamos el problema de estimar  $g(\mathbf{Y})$  (esto es, una cantidad aleatoria) para lo cual seleccionamos una muestra  $s$  de la población de tamaño  $n$  y elaboramos un estimador que será función de  $s$  y del vector de características observado  $\mathbf{y}_s = (y_{i_1}, y_{i_2}, \dots, y_{i_n})^t$  asociado a  $s$ . Los criterios de selección del estimador óptimo (y en algunos casos el procedimiento de muestreo óptimo para seleccionar la muestra  $s$  para un modelo dado) se pueden clasificar en dos grandes grupos: criterios de población fija - superpoblación y criterios predictivos.

1.- **Criterios de población fija - superpoblación.** Según estos criterios debemos considerar dos fuentes de aleatoriedad en la modelización de la población finita: la inducida por el mecanismo de muestreo aplicado para seleccionar la muestra y la del modelo de superpoblación. Si denotamos por:  $E_p(\cdot)$  al operador esperanza respecto al procedimiento de muestreo y  $V_p(\cdot)$  al operador varianza según el procedimiento de muestreo, según la importancia que se otorga al modelo de superpoblación o al procedimiento de muestreo podemos distinguir las siguientes variantes:

- **Criterio 1.** Consiste en minimizar

$$E_{\xi}[E_p(g(\widehat{\mathbf{Y}}_s) - g(\mathbf{Y}))^2],$$

entre el conjunto de *estimadores  $g(\widehat{\mathbf{y}}_s)$  insesgados respecto al procedimiento de muestreo*, esto es, aquellos tales que:

$$E_p(g(\widehat{\mathbf{y}}_s)) = g(\mathbf{y}).$$

- **Criterio 2.** Consiste en minimizar

$$E_{\xi}[E_p(g(\widehat{\mathbf{Y}}_s) - g(\mathbf{Y}))^2],$$

entre el conjunto de *estimadores  $g(\widehat{\mathbf{y}}_s)$  insesgados respecto al modelo de superpoblación*, esto es, aquellos tales que:

$$E_{\xi}(g(\widehat{\mathbf{Y}}_s) - g(\mathbf{Y})) = 0.$$

- **Criterio 3.** Consiste en minimizar

$$E_{\xi}[E_p(g(\widehat{\mathbf{Y}}_s) - g(\mathbf{Y}))^2],$$

entre el conjunto de *estimadores*  $g(\widehat{\mathbf{y}}_s)$  *insesgados respecto al procedimiento de muestreo y el modelo de superpoblación*, esto es aquellos tales que:

$$E_p[E_{\xi}(g(\widehat{\mathbf{Y}}_s) - g(\mathbf{Y}))] = 0.$$

Notemos que en el criterio 1 se otorga mayor importancia al procedimiento de muestreo que al modelo de superpoblación y es defendido desde dos puntos de vista distintos: por autores defensores del punto de vista de población fija, como procedimiento para la obtención de un estimador óptimo para un procedimiento de muestreo dado en el que el modelo de superpoblación juega el papel de un mecanismo adicional de juicio (en este sentido destacar los trabajos de Godambe(55)[48], Cochran(77)[18], Cumberland y Royall(81)[19] y Padmawar(81)[79] ) y desde un punto de vista de superpoblación como mecanismo de protección ante una mala elección del modelo de superpoblación (una solución alternativa sería el utilizar en caso de dudas sobre el modelo una clase de modelos más amplia, dentro de la cual podamos considerar que está incluida nuestra población). El criterio 2 se sitúa, en cambio, en la posición contraria dando mayor "importancia" y credibilidad al modelo de superpoblación que al procedimiento de muestreo y el criterio 3 busca una postura intermedia entre los otros dos criterios intentando proteger la estimación ante malas elecciones del modelo pero manteniendo a la vez un sentimiento de credibilidad hacia el mismo.

**2.- Criterios Predictivos.** Son aquellos criterios que eligen a posteriori de la selección de la muestra cuál es el mejor estimador para un modelo de superpoblación dado. Dentro de ellos podemos distinguir dos tratamientos o criterios clásicos distintos:

- **Criterio 1.** Consiste en seleccionar el estimador que minimice:

$$E_{\xi}(g(\widehat{\mathbf{Y}}_s) - g(\mathbf{Y}))^2,$$

entre el conjunto de *estimadores*  $g(\widehat{\mathbf{y}}_s)$  *insesgados respecto al modelo de superpoblación*, esto es, aquellos tales que:

$$E_{\xi}(g(\widehat{\mathbf{Y}}_s) - g(\mathbf{Y})) = 0.$$

Este criterio en la línea de los estudiados hasta ahora proporciona el mismo estimador que el criterio 2 de población fija - superpoblación en el caso de procedimientos de muestreo no informativos ya que en este caso los operadores esperanza pueden permutarse, de modo que:

$$E_{\xi}[E_p(g(\widehat{\mathbf{Y}}_s) - g(\mathbf{Y}))^2] = E_p[E_{\xi}(g(\widehat{\mathbf{Y}}_s) - g(\mathbf{Y}))^2],$$

con lo cual al obtener un estimador  $\widehat{g}$  que minimice  $E_{\xi}(g(\widehat{\mathbf{Y}}_s) - g(\mathbf{Y}))^2$  para un muestreo dado entre los estimadores  $\xi$ -insesgados obtenemos un estimador que minimiza  $E_{\xi}[E_p(g(\widehat{\mathbf{Y}}_s) - g(\mathbf{Y}))^2]$  entre los  $\xi$ -insesgados.

- **Criterio 2.- Estimadores Predictivos Mximo Verosimiles.** Se basan en la obtenci3n de estimadores que maximicen una determinada funci3n de verosimilitud sobre el vector de caracteristicas de la poblaci3n no observado  $\mathbf{Y}_u = (Y_i : i \notin s)^t$  que represente nuestro actual conocimiento una vez obtenida  $s$  y observado el vector  $\mathbf{y}_s = (y_{i_1}, y_{i_2}, \dots, y_{i_n})^t$  de caracteristicas muestral asociado a  $s$ . Siguiendo el libro de Bolfarine y Zacks(92)[10][capitulo 4] y la tesis de Iglesias(93)[63][capitulo 4] podemos sealar 5 funciones de verosimilitud:

- *Funci3n de verosimilitud predictiva estimada.*
- *Funci3n de verosimilitud predictiva perfilada.*
- *Funci3n de verosimilitud predictiva de Lauritzen y Hinkley (Ver Hinkley(79)[61]).*
- *Funci3n de verosimilitud predictiva de Royall (Ver Royall(76)[94]).*
- *Funci3n de verosimilitud predictiva de Iglesias(93)[63].*

(Las definiciones de las cuatro primeras pueden consultarse en Bolfarine y Zacks(92)[10][pag. 99-105] y la de la quinta en Iglesias(93)[63][pag. 205-206].)

En la pr3xima subsecci3n centaremos fundamentalmente nuestro inter3s en los estimadores  $\widehat{\bar{Y}}$  de la media poblacional  $\bar{Y}$  basados en los criterios predictivos, ya que consideramos que son los que mejor reflejan el punto de vista de modelos de superpoblaci3n clsicos de esta secci3n y los ms comparables con los estimadores basados en modelos de superpoblaci3n tratados desde un punto de vista Bayesiano que analizaremos en la secci3n 1.4. y los que estudiaremos en los pr3ximos capitulo de esta tesis.

### 1.3.2 Estimadores de la media poblacional

Entre los trabajos publicados desde el punto de vista de modelos de superpoblaci3n clsicos con el objetivo de encontrar un estimador  $\widehat{\bar{Y}}$  de la media poblacional  $\bar{Y}$  destaquemos los siguientes:

**A.-** Entre los trabajos basados en los *criterios de poblaci3n fija - superpoblaci3n*, los resultados de Godambe y Joshi(65)[51] y Cassel, Srndal y Wretman(76,77) [16,17] respecto al criterio de estimaci3n 1 y de Cassel, Srndal y Wretman(76,77)[16,17] respecto al criterio de estimaci3n 3, que describiremos seguidamente.

1. **Resultado de Godambe y Joshi(65)[51].** Estos dos autores demostraron que los estimadores de la forma:

$$\widehat{Y} = \frac{n}{N}\bar{y}_s + \frac{N-n}{N}\widehat{Y}_u, \quad (1.8)$$

donde  $\widehat{Y}_u$  es un estimador de la media de los elementos no observados  $\bar{Y}_u = \frac{\sum_{i \in u} Y_i}{N-n}$  son insesgados respecto a un procedimiento de muestreo sin reemplazamiento con probabilidades de inclusión estrictamente positivas; cumpliéndose además si asumimos un modelo clásico simplificado de regresión independiente con  $v(x_i) = v_i, \forall i$  ( $v_i > 0$  constantes conocidas), la siguiente desigualdad:

$$E_{\xi}(E_p(\widehat{Y} - \bar{Y})^2) \geq \frac{\sigma^2}{N^2} \left[ \frac{1}{n} \left( \sum_{i=1}^N \sqrt{v_i} \right)^2 - \sum_{i=1}^N v_i \right].$$

A partir de estos dos resultados probaron que si asumimos el modelo clásico simplificado de regresión anterior, los procedimientos de muestreo sin reemplazamiento con probabilidades de inclusión  $\alpha_i = \frac{n\sqrt{v_i}}{\sum_{i=1}^N \sqrt{v_i}}$  para  $i = 1, 2, \dots, N$  y el estimador de Horwitz y Thomson (1.4) asociado son óptimos entre los procedimientos de muestreo sin reemplazamiento con probabilidades de inclusión estrictamente positivas y los estimadores de la forma (1.8).

2. **Resultados de Cassel, Särndal y Wretman(76,77)[16,17].** Estos autores demostraron que los muestreos sin reemplazamiento con probabilidades de inclusión  $\alpha_i = \frac{n}{N}$  (por ejemplo el muestreo aleatorio simple) y el estimador media muestral (1.1) son óptimos para un modelo básico, básico intercambiable o básico con variables  $Y_i$  independientes e igualmente distribuidas con  $\mathbf{V} = (1 - \rho)\sigma^2\mathbf{I}_N + \rho\sigma^2\mathbf{1}_N\mathbf{1}_N^t$  ( $\mathbf{I}_N$  la matriz identidad de dimensión  $N$  y  $\mathbf{1}_N$   $N$ -vector de unos,  $\sigma^2$  y  $\rho$  escalares desconocidos ( $\sigma^2 > 0$ )), entre los procedimientos de muestreo sin reemplazamiento con probabilidades de inclusión estrictamente positivas y los estimadores lineales insesgados respecto al procedimiento de muestreo. Y en el caso del modelo básico también lo son entre los procedimientos de muestreo anteriores y los estimadores insesgados respecto al procedimiento de muestreo y al modelo de superpoblación.

En esta relación de resultados se producen dos ausencias bastante significativas e intencionadas que deseamos justificar en este punto, por una parte no se incluyen los trabajos de Godambe(55)[48], Cochran(77)[18] Cumberland y Royall(81)[19] y Padmawar(81)[79], ya que consideramos que son trabajos de población fija porque a pesar de asumirse un modelo de superpoblación, éste sólo se usa como una información auxiliar para poder seleccionar un estimador óptimo como ya indicábamos en la subsección anterior, y por otra la ausencia de resultados según el criterio 2. Esta última omisión viene determinada por la circunstancia de que los resultados más importantes se realizan bajo la hipótesis de un procedimiento de muestreo no informativo con lo cual (siguiendo la reflexión que realizábamos en la subsección de criterios) el problema de obtener el estimador óptimo según el criterio 2 se reduce

a obtener el estimador óptimo según el criterio predictivo 1 y la diferencia única entre los dos criterios estriba en que los defensores del criterio 2 obtienen además del estimador, el procedimiento de muestreo óptimo (entre los no informativos) para el cual el valor esperado para el procedimiento de muestreo de la varianza del estimador respecto al modelo sea mínimo (señalemos los trabajos de Royall(70, 71)[91,92] y Murgui(82)[76] en la determinación de estos procedimientos de muestreo óptimos). En las próximas líneas presentaremos los estimadores predictivos según el criterio predictivo 1 más estudiados que pueden ser interpretados a la luz de la observación anterior como los estimadores de población fija - superpoblación según el criterio 2 omitidos.

**B.-** Entre los trabajos basados en *criterios predictivos* destaquemos los siguientes resultados correspondientes al criterio de encontrar un estimador que minimice el error cuadrático medio respecto al modelo entre los estimadores insesgados respecto al modelo (criterio predictivo 1).

1. **Estimador predictivo lineal para un modelo clásico básico.** Cassel, Särndal y Wretman(77)[17] demostraron que la media muestral  $\bar{y}_s$  (1.1) es el estimador óptimo según el criterio predictivo 1 para el modelo clásico básico con  $\mathbf{V} = (1 - \rho)\sigma^2\mathbf{I}_N + \rho\sigma^2\mathbf{1}_N\mathbf{1}_N^t$  ( $\mathbf{I}_N$  la matriz identidad de dimensión  $N$  y  $\mathbf{1}_N$   $N$ -vector de unos,  $\sigma^2$  y  $\rho$  escalares desconocidos ( $\sigma^2 > 0$ )), entre los estimadores *lineales* insesgados respecto al modelo.
2. **Estimador predictivo lineal para el modelo clásico simplificado de regresión independiente.** Brewer(63)[11] y Royall(70)[91] proponen el estimador:

$$\hat{Y}_{SC1} = \frac{N}{n}\bar{y}_s + (1 - \frac{n}{N})\hat{\beta}\bar{x}_u, \quad (1.9)$$

donde  $\bar{y}_s$  es la media muestral (1.1),  $\bar{x}_u = \frac{\sum_{i \notin s} x_i}{N-n}$  y  $\hat{\beta} = \frac{\sum_{i \in s} \frac{x_i y_i}{v(x_i)}}{\sum_{i \in s} \frac{1}{v(x_i)}}$ , demostrando que es el óptimo según el criterio predictivo 1 para el modelo clásico simplificado de regresión independiente entre los estimadores *lineales* insesgados respecto al modelo.

3. **Estimador Predictivo para el modelo Clásico de Regresión con matriz de varianzas-covarianzas conocida.** Royall(76)[93] propone el estimador (bajo la hipótesis de rango de  $\mathbf{X}_s$  igual a  $p$ ):

$$\hat{Y}_{SC2} = \frac{1}{N}[\mathbf{1}_n^t \mathbf{y}_s + \mathbf{1}_{N-n}^t [\mathbf{X}_u \hat{\beta} + \mathbf{V}_u \mathbf{V}_s^{-1} (\mathbf{y}_s - \mathbf{X}_s \hat{\beta})]], \quad (1.10)$$

donde  $\mathbf{y}_s = (y_{i_1}, y_{i_2}, \dots, y_{i_n})^t$ ,  $\mathbf{Y}_u = (Y_i : i \notin s)^t$ ,  $\mathbf{X}_s = (x_{i_1}^t, x_{i_2}^t, \dots, x_{i_n}^t)^t$ ,  $\mathbf{X}_u = (x_i^t : i \notin s)^t$ ,  $\mathbf{V}_s = \text{Var}_\xi(\mathbf{Y}_s)$ ,  $\mathbf{V}_u = \text{Var}_\xi(\mathbf{Y}_u)$ ,  $\mathbf{V}_{su} = \text{Cov}_\xi(\mathbf{Y}_s, \mathbf{Y}_u)$  y

$$\hat{\beta} = (\mathbf{X}_s^t \mathbf{V}_s^{-1} \mathbf{X}_s)^{-1} \mathbf{X}_s^t \mathbf{V}_s^{-1} \mathbf{y}_s,$$

es el estimador lineal insesgado óptimo de  $\beta$ , demostrando que es óptimo según el criterio predictivo 1 para el modelo clásico de regresión con  $\mathbf{V}$  conocido, entre los

estimadores *lineales* insesgados respecto al modelo. Una demostración alternativa de este resultado puede encontrarse de Rodrigues y Elian(89)[90] y en Bolfarine y Zacks (92)[capítulo 2][10].

A partir de este estimador se obtiene el **estimador lineal predictivo para el modelo clásico básico con  $\mathbf{V}$  conocido** para la media poblacional óptimo según el criterio predictivo 1 entre entre los estimadores lineales insesgados respecto al modelo, que vendrá dado (tomando  $p = 1$  y  $\mathbf{X} = \mathbf{1}_N$ ) por:

$$\widehat{Y}_{SC3} = \frac{1}{N}[n\bar{y}_s + (N - n)\widehat{\mu} + \mathbf{1}_{N-n}^t \mathbf{V}_{us} \mathbf{V}_s^{-1}(\mathbf{y}_s - \mathbf{1}_n \widehat{\mu})], \quad (1.11)$$

con  $\widehat{\mu} = (\mathbf{1}_n^t \mathbf{V}_s^{-1} \mathbf{1}_n)^{-1} \mathbf{1}_n^t \mathbf{V}_s^{-1} \mathbf{y}_s$ .

4. **Estimadores Predictivos Lineales para el modelo clásico básico incorrelado en dos etapas con medias para cada unidad iguales y matriz de varianzas-covarianzas conocida.** Royall(76)[93] demostró que el estimador

$$\begin{aligned} \widehat{Y}_{SC4} &= \frac{1}{N} \left\{ \sum_{(i,j) \in s} y_{ij} + \sum_{i \notin \{i_1, i_2, \dots, i_k\}} N_i \widehat{\mu} + \right. \\ &\quad \left. + \sum_{i \in \{i_1, i_2, \dots, i_k\}} (N_i - n_i) [\omega_i \bar{y}_{s_i} + (1 - \omega_i) \widehat{\mu}] \right\}, \quad (1.12) \end{aligned}$$

donde  $\bar{y}_{s_i} = \frac{\sum_{(i,j) \in s_i} y_{ij}}{n_i}$  es la media muestral en la unidad  $i$  para  $i \in \{i_1, i_2, \dots, i_k\}$  y

$$\begin{aligned} \omega_i &= \frac{\rho_i n_i}{1 - \rho_i + n_i \rho_i}, \quad i \in \{i_1, i_2, \dots, i_k\}, \\ u_i &= \frac{\frac{n_i}{\sigma_i^2 (1 - \rho_i + n_i \rho_i)}}{\sum_{i=1}^n \frac{n_i}{\sigma_i^2 (1 - \rho_i + n_i \rho_i)}}, \quad i \in \{i_1, i_2, \dots, i_k\}, \\ \widehat{\mu} &= \sum_{i \in \{i_1, i_2, \dots, i_k\}} u_i \bar{y}_{s_i}, \end{aligned}$$

es óptimo según el criterio predictivo 1 para el modelo clásico básico incorrelado en dos etapas con  $\mu_i = \mu$  para  $i = 1, 2, \dots, N$ , y  $\{\sigma_i\}_{i=1}^K$  y  $\{\rho_i\}_{i=1}^K$  conocidos, entre los estimadores *lineales* e insesgados respecto al modelo.

5. **Estimadores Predictivos Insesgados.** Todos los estimadores predictivos que hemos visto hasta ahora son estimadores óptimos según el criterio predictivo 1 para el modelo de superpoblación correspondiente entre los estimadores *lineales* e insesgados respecto al mismo, Bolfarine y Zacks(92)[10][capítulo 2] apoyándose en una extensión del teorema de Rao y Blackwell (ver Skiener(83)[109] y Rodrigues et al(85)[89]) demuestran que el estimador predictivo lineal para el modelo clásico de regresión (1.10) con varianza conocida es óptimo según el criterio predictivo 1 para el modelo básico de

regresión normal con matriz de varianzas-covarianzas conocida y  $\mathbf{V}_{su} = \mathbf{0}_{n \times (N-n)}$  entre los estimadores insesgados (no necesariamente lineales) respecto a dicho modelo. Consecuentemente, el estimador predictivo para el modelo clásico de regresión normal con matriz de varianzas-covarianzas conocida y  $\mathbf{V}_{su} = \mathbf{0}_{n \times (N-n)}$  viene dado por:

$$\widehat{Y}_{SC5} = \frac{n}{N} \bar{y}_s + \frac{1}{N} \mathbf{1}_{N-n}^t \mathbf{X}_u \widehat{\beta}, \quad (1.13)$$

donde  $\widehat{\beta} = (\mathbf{X}_s^t \mathbf{V}_s^{-1} \mathbf{X}_s)^{-1} \mathbf{X}_s^t \mathbf{V}_s^{-1} \mathbf{y}_s$ , y el estimador predictivo para el modelo clásico básico normal con  $\mathbf{V}_{su} = \mathbf{0}_{n \times (N-n)}$ ,

$$\widehat{Y}_{SC6} = \frac{n}{N} \bar{y}_s + (1 - \frac{n}{N}) (\mathbf{1}_n^t \mathbf{V}_s^{-1} \mathbf{1}_n)^{-1} \mathbf{1}_n^t \mathbf{V}_s^{-1} \mathbf{y}_s. \quad (1.14)$$

Respecto al criterio de estimadores predictivos máximo verosímiles, Bolfarine y Zacks(92) [10] [capítulo 4] obtienen para las cuatro primeras funciones de verosimilitud enumeradas en la subsección anterior, los estimadores máximo verosímiles predictivos para un modelo clásico de regresión normal con  $\mathbf{V}_s = \sigma^2 \mathbf{W}_s$ ,  $\mathbf{V}_u = \sigma^2 \mathbf{W}_u$  y  $\mathbf{V}_{su} = \mathbf{0}_{n \times (N-n)}$  con  $\mathbf{W}_s$  y  $\mathbf{W}_u$  matrices definidas positivas conocidas de dimensiones  $n \times n$  y  $(N-n) \times (N-n)$  respectivamente,  $\sigma^2$  un escalar positivo y  $\mathbf{0}_{n \times (N-n)}$  una matriz de ceros de dimensión  $n \times (N-n)$ . Para estas cuatro funciones de verosimilitud, se tiene que el estimador máximo verosímil predictivo coincide y es igual a  $\widehat{Y} = \frac{1}{N} [\mathbf{1}_n^t \mathbf{y}_s + \mathbf{1}_{N-n}^t \mathbf{X}_u \widehat{\beta}]$ , donde  $\widehat{\beta} = (\mathbf{X}_s^t \mathbf{W}_s^{-1} \mathbf{X}_s)^{-1} \mathbf{X}_s^t \mathbf{W}_s^{-1} \mathbf{y}_s$ , bajo las hipótesis:  $\sigma^2$  desconocida para la verosimilitud estimada,  $\sigma^2$  conocida para la verosimilitud perfilada,  $\sigma^2 = 1$  para la verosimilitud de L-H, y  $\sigma^2 = 1$  y condición L ( $\mathbf{W} \mathbf{1}_N = \mathbf{X} \delta$  para algún  $\delta$ ), para la verosimilitud de Royall. Notemos que los estimadores predictivos máximo verosímiles coinciden para las 4 primeras funciones de verosimilitud planteadas (bajo las correspondiente hipótesis adicionales), con el estimador predictivo insesgado óptimo para el modelo clásico de regresión normal con matriz de varianzas-covarianzas conocida y  $\mathbf{V}_{su} = \mathbf{0}_{n \times (N-n)}$  (ver (1.13)).

En relación a la función de verosimilitud quinta, Iglesias(93)[63] obtiene para el modelo clásico básico normal con matriz de varianzas covarianzas  $\mathbf{V} = \sigma^2 \mathbf{I}_N$ , con  $\sigma^2 > 0$  desconocido, que el estimador máximo verosímil para su función de verosimilitud predictiva para la media poblacional es la media muestral (1.1).

### 1.3.3 Comentarios de la sección

Tal y como hemos visto a lo largo de la sección, la aproximación al problema de estimación en poblaciones finitas basada en modelos de superpoblación sí obtiene estimadores únicos óptimos (la aproximación por población fija no alcanza resultados óptimos únicos -inexistencia de estimadores uniformes de mínima varianza-) según los criterios de población fija - superpoblación y predictivos. Del análisis de los resultados obtenidos mediante los diversos criterios de selección del estimador aplicados notemos que:

1.- En los resultados de estimación de la media poblacional según los criterios 1 y 3 de población fija - superpoblación se obtiene una solución de población fija apoyándose de manera accesoria en un modelo de superpoblación, esta solución confirma como estimador

óptimo de  $\bar{Y}$  a la media muestral  $\bar{y}$ , (1.1) (el estimador habitual en población fija en ausencia de información auxiliar) y al muestreo aleatorio simple sin reemplazamiento para modelos básicos, y al estimador de Horwitz y Thomson (1.4) y el muestreo proporcional al tamaño de la variable auxiliar para un modelo simplificado de regresión. Los criterios 1 y 3 de población fija - superpoblación son una solución “cuasi-población fija” al problema de encontrar un estimador óptimo mediante criterios de población fija “puros”.

2.- Los criterios basados en los criterios predictivos y el criterio 2 de población fija - superpoblación a diferencia de los anteriores consideran como elemento clave en la estimación el modelo de superpoblación (“creen” en el modelo), relegando a aspecto secundario el procedimiento de muestreo (en el proceso inferencial se considera la muestra como dada y en todo caso se busca un procedimiento que mejore el estimador predictivo, en el sentido de minimizar el valor esperado respecto a muestreos no informativos del error cuadrático medio respecto al modelo de superpoblación). Señalemos dos puntos de discusión para los resultados obtenidos a partir de estos criterios:

- Los modelos de superpoblación propuestos son muy generales y habitualmente no expresan “todo” lo que sabemos de la población finita en estudio. Así la hipótesis de intercambiabilidad finita se satisface para todos los modelos de superpoblación  $\xi$ , construibles a partir del siguiente procedimiento Bayesiano<sup>2</sup>:

“Si las variables  $Y_i$  son independientes e igualmente distribuidas para  $i = 1, 2, \dots, N$  dado un parámetro  $\theta$  (escalar o vector) con función de densidad  $f(Y|\theta)$  y asumimos que dicho parámetro  $\theta$  definido sobre  $\Theta$  se distribuye según función de densidad  $\pi(\theta)$ , entonces

$$f(Y_1, Y_2, \dots, Y_N) = \int_{\Theta} \prod_{i=1}^N f(Y_i|\theta)\pi(\theta)d\theta,$$

es una función de densidad correspondiente a un modelo de superpoblación intercambiable.”

en el cual la distribución asociada a  $\theta$  representa los distintos grados de conocimiento subjetivo del problema (una asignación clásica correspondería a tomar  $\pi(\theta) \propto 1$ ). Los modelos basados en las hipótesis de independencia o incorrelación son asimismo muy amplios.

- Los estimadores predictivos obtenidos son lineales y óptimos respecto al criterio de minimizar el error cuadrático medio entre los *estimadores lineales e insesgados* respecto al modelo de superpoblación y entre los *estimadores insesgados* en el caso modelos de superpoblación normales. Estos resultados se pueden obtener también a partir de los correspondientes modelos Bayesianos asumiendo una adecuada distribución mínimo informativa sobre los parámetros desconocidos, comprobándose además, desde este

---

<sup>2</sup>El procedimiento anterior supone una condición necesaria y suficiente de intercambiabilidad finita, pero no implica la existencia de una población infinita intercambiable de la que sea realización la población finita intercambiable así construida.

punto de vista, que la hipótesis de estimador lineal óptimo conduce a la normalidad del modelo<sup>3</sup>.

El análisis de la población finita a través de un modelo de superpoblación clásico proporciona resultados en el problema de estimación, constituye una aproximación estadística “natural” y supera las dos objeciones a la aproximación de población fija que señalábamos en la sección anterior. A pesar de ello, el análisis que permite la metodología de modelos de superpoblación clásica es incompleto desde nuestro punto de vista, los modelos son muy amplios y tanto ellos como los resultados para un modelo de superpoblación clásico dado pueden interpretarse de una forma más completa, tal y como muestran las dos observaciones anteriores, mediante una interpretación Bayesiana (que es siempre posible introduciendo las hipótesis adecuadas sobre los parámetros). Además, la aproximación Bayesiana tiene interés por sí misma, permite una modelización más natural y sencilla de las poblaciones, y permite además la obtención de la distribución predictiva de  $\mathbf{Y}_u|(s, \mathbf{y}_s)$  que reúne toda la información proporcionada por la muestra  $((s, \mathbf{y}_s))$  sobre el vector poblacional no observado (lo que supone una mejora conceptual frente a las funciones de verosimilitud predictivas máximo verosímiles comentadas en esta sección).

En la próxima sección y los capítulos siguientes aplicaremos técnicas Bayesianas sobre modelos de superpoblación.

Destaquemos como referencias fundamentales para el estudio de una población finita desde el punto de vista de superpoblación clásica (algunas de ellas ya referenciadas) por presentar un planteamiento global del tema, las siguientes: los artículos de Särndal(78)[104] y Godambe(92)[50] que presentan un amplio estudio teórico sobre los procedimientos y estimadores desde un punto de vista de método de muestreo y modelo de superpoblación y, desde un punto de vista predictivo los de Thomsen y Tesfu(88)[116] y Royall(88,92)[95,96], los libros de Cassel, Särndal y Wretman(77)[17][capítulos 4 y 5] y Bolfarine y Zacks(92)[10] y la tesis de Murgui(82)[76][capítulo 2].

## 1.4 Modelos de Superpoblación. Soluciones Bayesianas

Consideremos una población finita  $P = \{1, 2, \dots, N\}$ , un vector de características  $\mathbf{y} = (y_1, y_2, \dots, y_N)^t$  asociado a cada uno de los índices de la población, un modelo de superpoblación  $\xi$ , de modo que el vector  $\mathbf{y}$  es la realización de un vector aleatorio  $\mathbf{Y}$  que se distribuye según  $\xi$  y una muestra de la misma  $(s, \mathbf{y}_s)$  (a partir de ahora y en todo lo que resta de tesis cuando nos refiramos a la muestra nos referiremos a la pareja  $(s, \mathbf{y}_s)$ ), con  $s$  obtenido mediante un procedimiento de muestreo de tamaño fijo no informativo (habitualmente, la aproximación Bayes (y también la aproximación clásica predictiva) introduce toda la información sobre la población en el modelo y no hace uso de muestreos informativos, por lo que a partir de ahora y en lo que resta de tesis cuando nos refiramos a un procedimiento de muestreo éste será no informativo). Si considerada la muestra  $(s, \mathbf{y}_s)$ , denotamos por  $\mathbf{Y}_u = (Y_i : i \notin s)^t$ , al vector poblacional no observado (una parte de  $\mathbf{Y}$  desconocida una vez observada la muestra), la aproximación Bayesiana al problema para estimar  $g(\mathbf{Y})$

<sup>3</sup>Esta afirmación se apoya en los resultados de Goel y DeGroot(80)[52] y en el apartado de comentarios de la subsección 1.4.1

consiste en la asignación de una distribución sobre los parámetros desconocidos del modelo y la aplicación del teorema de Bayes para actualizar la distribución del vector  $\mathbf{Y}_u$  dada la muestra  $(s, \mathbf{y}_s)$ , esto es, la obtención de la distribución predictiva  $\mathbf{Y}_u|(s, \mathbf{y}_s)$ .

La distribución predictiva anterior, proporciona una información completa sobre el vector poblacional no observado condicionado a la muestra, que puede ser condensada (en cierta medida) para establecer inferencias acerca del mismo a partir de los primeros dos momentos de la distribución (en el caso de existir), que proporcionan un estimador del vector no observado y una medida sobre su variabilidad, la moda, la mediana, cuantiles de la distribución,...

El criterio más habitual consiste en la obtención del estimador Bayes de  $g(\mathbf{Y})$  que minimice el error esperado (respecto al modelo de superpoblación) asociado a una determinada función de pérdida. La función de pérdida que consideraremos en esta tesis (y la más aplicada en la práctica), es la función de pérdida cuadrática; para esta función de pérdida el estimador Bayes se obtendrá como aquel estimador  $g(\widehat{\mathbf{y}}_s)$  que minimice  $E_\xi(g(\widehat{\mathbf{Y}}_s) - g(\mathbf{Y}))^2$ . Dado que<sup>4</sup>:

$$E_\xi(g(\widehat{\mathbf{Y}}_s) - g(\mathbf{Y}))^2 = E_{\mathbf{Y}_s}[Var_{\mathbf{Y}_u|(s, \mathbf{y}_s)}(g(\mathbf{Y}))] + E_{\mathbf{Y}_s}[g(\widehat{\mathbf{Y}}_s) - E_{\mathbf{Y}_u|(s, \mathbf{y}_s)}(g(\mathbf{Y}))]^2,$$

el estimador Bayes  $g_B(\widehat{\mathbf{y}}_s)$  de  $g(\mathbf{Y})$  y su error cuadrático medio (respecto al modelo de superpoblación) vienen dados (cuando existen los correspondientes momentos) respectivamente por:

$$g_B(\widehat{\mathbf{y}}_s) = E_{\mathbf{Y}_u|(s, \mathbf{y}_s)}(g(\mathbf{Y})),$$

$$E_\xi(g_B(\widehat{\mathbf{y}}_s) - g(\mathbf{Y}))^2 = E_{\mathbf{Y}_s}[Var_{\mathbf{Y}_u|(s, \mathbf{y}_s)}(g(\mathbf{Y}))].$$

Notemos que el estimador Bayes de  $g(\mathbf{Y})$  viene dado por el valor esperado de  $g(\mathbf{Y})$  respecto a la distribución predictiva  $\mathbf{Y}_u|(s, \mathbf{y}_s)$  y su error cuadrático medio por la media respecto a la distribución de  $\mathbf{Y}_s$  de la varianza de  $g(\mathbf{Y})$  respecto a la distribución predictiva  $\mathbf{Y}_u|(s, \mathbf{y}_s)$ . Esto es, estimador Bayes y error cuadrático medio asociado se obtienen a partir de los dos primeros momentos (si existen) respecto a  $g(\mathbf{Y})$  de la distribución predictiva de  $\mathbf{Y}_u|(s, \mathbf{y}_s)$  (que además pueden interpretarse respectivamente como el estimador Bayes de  $g(\mathbf{Y})$  y su error cuadrático medio asociado bajo el criterio de minimizar la función de pérdida cuadrática esperada para la distribución predictiva).

Por otra parte, notemos que el estimador Bayes  $g_B(\widehat{\mathbf{y}}_s)$  es insesgado respecto al modelo de superpoblación  $\xi$ , puesto que:

$$\begin{aligned} E_\xi(g_B(\widehat{\mathbf{Y}}_s) - g(\mathbf{Y})) &= E_{\mathbf{Y}_s}[E_{\mathbf{Y}_u|(s, \mathbf{y}_s)}(g_B(\widehat{\mathbf{Y}}_s) - g(\mathbf{Y}))] = \\ &= E_{\mathbf{Y}_s}[g_B(\widehat{\mathbf{Y}}_s) - E_{\mathbf{Y}_u|(s, \mathbf{y}_s)}(g(\mathbf{Y}))] = 0, \end{aligned}$$

y que por lo tanto en el caso de modelos de superpoblación clásicos con parámetros conocidos el estimador Bayes coincide con el estimador predictivo (criterio predictivo 1).

<sup>4</sup>Al obtener  $s$  a partir de muestreos no informativos, notemos que  $s$  es subordinada (ancillarity) por lo que la inferencia sobre  $\mathbf{Y}$  se deriva únicamente de la distribución de  $\mathbf{Y}|s$

En particular, si nos centramos en la estimación de la media poblacional, tenemos que  $\bar{Y} = \frac{n}{N}\bar{y}_s + \frac{N-n}{N}\bar{Y}_u$ , donde  $\bar{y}_s$  es la media muestral (1.1) y  $\bar{Y}_u = \sum_{i \notin s} Y_i / (N - n)$  es la media de los elementos no observados, por lo que el estimador Bayes y su error cuadrático medio (respecto al modelo de superpoblación) asociado vienen dados por:

$$\begin{aligned}\hat{\bar{Y}}_B &= \frac{n}{N}\bar{y}_s + \frac{N-n}{N}\hat{\bar{Y}}_u^B, \\ E_{\xi}(\hat{\bar{Y}}_B - \bar{Y})^2 &= \frac{(N-n)^2}{N^2} E_{\mathbf{Y}_s} [Var_{\mathbf{Y}_u | (s, \mathbf{y}_s)}(\bar{Y}_u)],\end{aligned}$$

con  $\hat{\bar{Y}}_u^B = E(\bar{Y}_u | (s, \mathbf{y}_s))$ , el estimador Bayes con función de pérdida cuadrática de  $\bar{Y}_u$ .

El proceso inferencial anterior entraña bastantes dificultades (salvo en modelos conjugados), las distribuciones predictivas del vector poblacional no observado  $\mathbf{Y}_u | (s, \mathbf{y}_s)$ , no son fáciles de obtener, garantizar la existencia de la media y varianza de  $g(\mathbf{Y})$  respecto a la distribución predictiva anterior también presenta dificultades, que se ven acrecentadas considerablemente cuando la función  $g(\mathbf{Y})$  no es lineal y además en el caso de poder garantizar la existencia de los momentos anteriores, su cálculo requiere de técnicas de integración numérica. En los próximos capítulos se estudiarán algunos modelos en los que tendremos que hacer frente a estas dificultades.

En esta sección presentaremos los estimadores Bayes y sus errores para modelos Bayesianos sencillos y estimadores alternativos al estimador Bayes y sus errores surgidos por la dificultad en la obtención del estimador Bayes para un modelo dado (estimadores Bayes lineales y estimadores empíricos) o la exigencia de propiedades adicionales no satisfechas por el estimador Bayes (estimadores Bayes constreñidos). Dentro de la estimación de modelos de superpoblación desde un punto de vista Bayesiano se da una relación muy estrecha entre el modelo propuesto y la solución Bayes aportada por lo que distribuiremos el contenido de esta sección en cuatro subsecciones con títulos:

- 1.4.1 Modelos Bayesianos con distribuciones no totalmente especificadas. Estimadores Bayes lineales
- 1.4.2 Modelos Bayesianos para distribuciones especificadas. Estimadores Bayes y estimadores Bayes empíricos
- 1.4.3 Estimadores Bayes constreñidos
- 1.4.4 Modelos "rigurosamente" Bayesianos
- 1.4.5 Comentarios de la sección

en las cuatro primeras analizaremos los modelos habitualmente estudiados, el procedimiento de estimación empleado, las motivaciones que conducen a estos modelos y los resultados obtenidos por los distintos autores; en la subsección de comentarios se realizará un análisis general de la sección estableciendo la relación entre los estimadores propuestos y los modelos planteados.



### 1.4.1 Modelos Bayesianos con distribuciones no totalmente especificadas. Estimadores Bayes lineales

Un grupo significativo de autores se inclinan por una modelización Bayesiana de la población finita en estudio “amplia”, en el sentido de que la distribución  $\xi$  no esté completamente especificada, entre los modelos más estudiados podemos distinguir dos grupos:

1.- *Los modelos Bayesianos basados en especificaciones de primer y segundo momentos sobre el vector poblacional dados los parámetros y sobre los parámetros.* Entre ellos señalemos:

- **Modelo Bayes Básico.** Pertenecen a este modelo la clase de distribuciones de superpoblación  $\xi$  tales que:

$$Y_i = \mu + e_i, \quad i = 1, 2, \dots, N,$$

donde  $\mathbf{e} = (e_1, e_2, \dots, e_N)^t$  es un vector aleatorio de media  $\mathbf{0}_N$  y matriz de varianza-covarianzas  $\mathbf{V}$  y  $\mu$  una variable aleatoria de media  $b$  y varianza  $\sigma_b^2$ ,  $b$  es un escalar conocido,  $\mathbf{V}$  una matriz  $N \times N$  definida positiva conocida y  $\sigma_b^2$  un escalar ( $\sigma_b^2 > 0$ ) conocido.

- **Modelo Bayes de Regresión.** Pertenecen a este modelo la clase de distribuciones de superpoblación  $\xi$  para las cuales, asociada a cada variable  $Y_i$  tenemos una información auxiliar representada por un vector  $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})$  conocido cumpliéndose que:

$$Y_i = \mathbf{x}_i \boldsymbol{\beta} + e_i, \quad i = 1, 2, \dots, N,$$

donde  $\mathbf{X} = (\mathbf{x}_1^t, \mathbf{x}_2^t, \dots, \mathbf{x}_N^t)^t$ ,  $\mathbf{e} = (e_1, e_2, \dots, e_N)^t$  es vector aleatorio de medias  $\mathbf{0}_N$  y matriz de varianza-covarianzas  $\mathbf{V}$  y  $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_p)^t$  es vector aleatorio de media  $\mathbf{b}$  y matriz de varianza-covarianza  $\mathbf{B}$ ,  $\mathbf{b}$  un  $p$ -vector conocido y  $\mathbf{V}$  y  $\mathbf{B}$  son matrices definidas positivas conocidas de dimensiones  $N \times N$  y  $p \times p$  respectivamente. ( $\mathbf{0}_N$   $N$ -vector de ceros). Notemos que, en el caso  $p = 1$ ,  $\mathbf{X} = \mathbf{1}_N$  y  $\boldsymbol{\beta} = \mu$  ( $\mathbf{1}_N$   $N$  vector de unos) se obtiene el modelo Bayes Básico del párrafo anterior.

- **Modelos Bayesianos Básicos en dos etapas.** Constituyen la transformación del modelo clásico en dos etapas y modelo clásico simplificado de regresión en dos etapas a modelos Bayesianos lo que se realiza asumiendo para los *modelos Bayesianos Básicos en dos etapas con una jerarquía* que  $\boldsymbol{\mu} = (\mu_1, \mu_2, \dots, \mu_K)^t$  es un  $K$ -vector aleatorio con media  $\mu_\pi \mathbf{1}_K$  y matriz de varianzas-covarianzas  $\mathbf{B} = \sigma_\pi^2 \mathbf{I}_K$  donde todas las matrices de varianzas-covarianzas son definidas positivas y conocidas y  $\mu_\pi$  asimismo conocida y para los *modelos Bayesianos Básicos en dos etapas con dos jerarquías* que  $\boldsymbol{\mu} = (\mu_1, \mu_2, \dots, \mu_K)^t$  es un  $K$ -vector aleatorio con media  $\mu_\pi \mathbf{1}_K$  y matriz de varianzas-covarianzas  $\mathbf{B} = \sigma_\pi^2 \mathbf{I}_K$  con  $\mu_\pi$  variable aleatoria con media  $\mu_0$  y varianza  $\sigma_0^2$  donde todas las matrices de varianzas-covarianzas son definidas positivas y conocidas y  $\mu_0$  y  $\sigma_0^2$  escalares ( $\sigma_0^2 > 0$ ) conocidos.

Los modelos Bayesianos en dos etapas se aplican en poblaciones estructuradas en dos etapas. La aplicación del modelo con una o con dos jerarquías depende fundamentalmente de las características de la población y del nivel de información disponible, en general, el primer modelo se aplica en situaciones de muestreo estratificado, cuando los estratos son bastante heterogéneos entre sí y se dispone de bastante información acerca de los mismos (una generalización habitual es tomar  $\mu_\pi$  distinta en cada estrato), y el segundo cuando el nivel de información acerca de las unidades es más pobre o bien las características de la población hacen más adecuada la construcción de un modelo más flexible que permita establecer introduciendo una jerarquía más una relación de intercambiabilidad entre las unidades (esta modelización es adecuada en muchas situaciones de muestreo cluster, por rutas aleatorias e incluso por muestreos estratificados).

**2.- Los modelos Bayesianos intercambiables basados en especificaciones de primer y segundo momentos sobre el vector poblacional.** Los modelos que describiremos a continuación suelen enunciarse en términos del modelo de superpoblación colapsando los parámetros (que por otra parte no suelen ser objeto de estudio en poblaciones finitas). Entre ellos, señalemos:

- **Modelo Bayes Intercambiable Básico.** Pertenecen a este modelo la clase de distribuciones  $\xi$  tales que las variables  $Y_i$ ,  $i = 1, 2, \dots, N$  son intercambiables con

$$\begin{aligned} E(Y_i) &= m, \quad i = 1, 2, \dots, N, \\ \text{Cov}(Y_i, Y_j) &= \begin{cases} v, & i = j \quad i = 1, 2, \dots, N, \\ c, & i \neq j \quad i, j = 1, 2, \dots, N, \end{cases} \end{aligned}$$

donde  $m$ ,  $v$  y  $c$  son escalares conocidos ( $v > 0$ ).

Como casos particulares de este modelo tenemos todos los modelos de superpoblación construibles a partir del siguiente modelo:

- \*  $Y_i|\theta$ ,  $i = 1, 2, \dots, N$  independientes e igualmente distribuidas (i.i.d.) con media  $E(Y_i|\theta) = \theta$  y varianza  $\text{Var}(Y_i|\theta) = v - c$ ,  $i = 1, 2, \dots, N$ ,
- \*  $\theta$  variable aleatoria con media  $E(\theta) = m$  y varianza  $\text{Var}(\theta) = c$ ,

donde  $m$ ,  $c$  y  $v$  son escalares conocidos ( $v > 0$ ).

- **Modelos Bayesianos Intercambiables en dos etapas.** Podemos distinguir dos modelos: el *modelo intercambiable en dos etapas de una jerarquía* y el *modelo intercambiable en dos etapas de dos jerarquías*.

\* *El modelo intercambiable en dos etapas de una jerarquía.* Pertenecen a este modelo la clase de distribuciones  $\xi$  tales que:

- \* las unidades  $\mathbf{Y}_i = (Y_{i1}, Y_{i2}, \dots, Y_{iN_i})^t$ , son vectores aleatorios incorrelados para  $i = 1, 2, \dots, K$ ,
- \* para  $i = 1, 2, \dots, K$ , las variables  $Y_{i1}, Y_{i2}, \dots, Y_{iN_i}$  son intercambiables con  $E(Y_{ij}) = m_i$ ,  $j = 1, 2, \dots, N_i$ ,  $\text{Var}(Y_{ij}) = v_i$ ,  $j = 1, 2, \dots, N_i$  y  $\text{Cov}(Y_{ij}, Y_{jl}) = c_i$  para  $j \neq l$ ,  $j, l = 1, 2, \dots, N_i$ , con  $m_i$ ,  $v_i$  y  $c_i$  escalares conocidos para  $j, l = 1, 2, \dots, N_i$  ( $v_i > 0$ ,  $j, l = 1, 2, \dots, N_i$ ).

\* *El modelo intercambiable en dos etapas de dos jerarquías.* Pertenecen a este modelo la clase de distribuciones  $\xi$  tales que:

- \* las unidades  $\mathbf{Y}_i = (Y_{i1}, Y_{i2}, \dots, Y_{iN_i})^t$ , condicionadas a  $\bar{Y}_i$  y  $S_i = \frac{\sum_{j=1}^{N_i} (Y_{ij} - \bar{Y}_i)^2}{N_i}$  son intercambiables para  $i = 1, 2, \dots, K$ ,
- \* para  $i = 1, 2, \dots, K$ , las variables  $Y_{i1}, Y_{i2}, \dots, Y_{iN_i}$  son independientes entre sí,
- \* para  $i = 1, 2, \dots, K$ , los vectores  $(\bar{Y}_i, S_i)$  son intercambiables con  $E(\bar{Y}_i) = m$ ,  $Var(\bar{Y}_i) = v$ ,  $Cov(\bar{Y}_i, \bar{Y}_l) = c$  para  $i \neq l$  y  $E(S_i) = \phi$  donde  $m, v, c$  y  $\phi$  son escalares conocidos ( $|c| < v/(K - 1)$ ).

Estos modelos se aplican sobre poblaciones estructuradas en dos etapas, aplicando uno u otro modelo según las características de la población en estudio y nuestro nivel de información acerca de la misma.

## Estimación

Dada una muestra  $(s, \mathbf{y}_s)$ , la obtención en estos modelos del estimador Bayes  $g_B(\widehat{\mathbf{y}}_s)$  para función de pérdida cuadrática para estimar la función  $g(\mathbf{Y})$  no es posible, ya que la información disponible acerca de la distribución del vector de características  $\mathbf{Y}$  no es suficiente para el cálculo del estimador ni de su error cuadrático medio (respecto al modelo de superpoblación) asociado. Hartigan(69)[55] y Brunk(80)[12] introducen para modelos de similares características los *estimadores Bayes lineales*, que son aplicados al ámbito de estimación de poblaciones finitas por Ericson(69, 83, 88)[27,29,30], Smouse(82,84)[112,113] y Murgui(82)[76].

El estimador Bayes lineal  $g_{BL}(\widehat{\mathbf{y}}_s)$  (para función de pérdida cuadrática) de  $g(\mathbf{Y})$  se define como aquel estimador  $g(\widehat{\mathbf{y}}_s)$  de  $g(\mathbf{Y})$  que minimice

$$E_{\xi}(g(\widehat{\mathbf{Y}}_s) - g(\mathbf{Y}))^2 = E_{\mathbf{Y}_s}[Var_{\mathbf{Y}_s|(s, \mathbf{y}_s)}(g(\mathbf{Y}))] + E_{\mathbf{Y}_s}[g(\widehat{\mathbf{Y}}_s) - E_{\mathbf{Y}_s|(s, \mathbf{y}_s)}(g(\mathbf{Y}))]^2,$$

(equivalentemente  $E_{\mathbf{Y}_s}(g(\widehat{\mathbf{Y}}_s) - g_B(\widehat{\mathbf{Y}}_s))^2$  (ver Ericson(83, 88)[29,30] y Murgui(82)[76]) entre los estimadores de la forma

$$g(\widehat{\mathbf{y}}_s) = a_0 + \sum_{i \in s} a_i y_i,$$

con  $a_0$  y  $a_i, i = 1, 2, \dots, n$  escalares conocidos. El estimador Bayes lineal es pues, el estimador más próximo entre los lineales al estimador Bayes en error cuadrático medio respecto al modelo de superpoblación.

## Estimadores Bayes lineales de la media poblacional

Entre los estimadores Bayes lineales obtenidos para la media poblacional para los diversos modelos propuestos en esta subsección destaquemos en primer lugar, los correspondientes a los modelos Bayesianos basados en especificaciones de primer y segundo orden sobre el vector poblacional y los parámetros, y dentro de ellos el siguiente resultado para los modelos Bayesianos básico y de regresión. Rodrigues(88)[88] y Bolfarine y Zacks(92)[10][10] obtienen el *estimador Bayes lineal para el modelo Bayes de regresión* que se define por:

$$\widehat{Y}_{BL1} = \frac{n}{N}\bar{y}_s + \frac{1}{N}[1_{N-n}^t[\mathbf{X}_u\widehat{\beta}_{BL} + \mathbf{V}_u\mathbf{V}_s^{-1}(\mathbf{y}_s - \mathbf{X}_s\widehat{\beta}_{BL})]], \quad (1.15)$$

donde  $\bar{y}_s$  es la media muestral (1.1),  $\mathbf{y}_s = (y_{i_1}, y_{i_2}, \dots, y_{i_n})^t$ ,  $\mathbf{Y}_u = (Y_i : i \notin s)^t$ ,  $\mathbf{X}_s = (\mathbf{x}_{i_1}^t, \mathbf{x}_{i_2}^t, \dots, \mathbf{x}_{i_n}^t)^t$ ,  $\mathbf{X}_u = (\mathbf{x}_i^t : i \notin s)^t$ ,  $\mathbf{V}_s = \text{Var}_{\mathbf{Y}|\beta}(\mathbf{Y}_s)$ ,  $\mathbf{V}_u = \text{Var}_{\mathbf{Y}|\beta}(\mathbf{Y}_u)$ ,  $\mathbf{V}_{s,u} = \text{Cov}_{\mathbf{Y}|\beta}(\mathbf{Y}_s, \mathbf{Y}_u)$  y

$$\widehat{\beta}_{BL} = (\mathbf{X}_s^t\mathbf{V}_s^{-1}\mathbf{X}_s + \mathbf{B}^{-1})^{-1}(\mathbf{X}_s^t\mathbf{V}_s^{-1}\mathbf{y}_s + \mathbf{B}^{-1}\mathbf{b}),$$

que proporciona como caso particular para  $p = 1$ ,  $\mathbf{X} = \mathbf{1}_N$ ,  $\mathbf{b} = b$  y  $\mathbf{B} = \sigma_b^2$ , el *estimador Bayes lineal para el modelo Bayes Básico*, que viene dado por:

$$\widehat{Y}_{BL2} = \frac{n}{N}\bar{y}_s + \frac{N-n}{N}\widehat{\mu}_{BL} + \frac{1}{N}1_{N-n}^t\mathbf{V}_u\mathbf{V}_s^{-1}(\mathbf{y}_s - 1_n\widehat{\mu}_{BL}), \quad (1.16)$$

donde  $\widehat{\mu}_{BL} = (1_n^t\mathbf{V}_s^{-1}1_n + \sigma_b^{-2})^{-1}(1_n^t\mathbf{V}_s^{-1}\mathbf{y}_s + \sigma_b^{-2}b)$ . También calculan el error medio obtenido al emplear el estimador Bayes lineal para el modelo Bayes de Regresión (que podemos asimismo particularizar para el modelo Bayes básico) y los estimadores Bayes lineales y sus errores en los casos  $\mathbf{B} \rightarrow \mathbf{0}_{p \times p}$ ,  $\mathbf{V} \rightarrow \mathbf{0}_{N \times N}$ ,  $\mathbf{B}^{-1} \rightarrow \mathbf{0}_{p \times p}$  y  $\mathbf{V}^{-1} \rightarrow \mathbf{0}_{N \times N}$ . Otros casos particulares de este resultado son:

- para  $p = N$  y  $\mathbf{X} = \mathbf{I}_N$  se recupera el estimador Bayes lineal obtenido por Smouse(82)[112] para un modelo Bayesiano sin variables auxiliares y medias para cada elemento de la población  $Y_i$  dado el parámetro (que en este caso es un  $N$ -vector) distintas.
- en el caso  $\mathbf{B}^{-1} \rightarrow \mathbf{0}_{p \times p}$ , se recuperan los resultados clásicos; así el estimador (1.15) se transforma en el estimador predictivo para el modelo clásico de regresión (1.10) y el estimador (1.16) en el estimador predictivo para el modelo clásico básico (1.11), ambos modelos clásicos bajo la hipótesis de matriz de varianzas-covarianzas conocida.

Otros trabajos a destacar son: los resultados de Ericson(88)[30] para la estimación de  $\bar{Y}$  asumiendo los modelos intercambiables descritos en líneas anteriores (obtención de estimadores Bayes lineales y su error cuadrático medio asociado), así como su labor de estimación para algunos modelos de regresión y los estudios de Ericson(70)[28] y Ghosh

y Lahiri(87)[41] que obtienen los estimadores Bayes para un modelo Bayesiano básico intercambiable y para un modelo Bayesiano en dos etapas con una jerarquía con  $k = K$  respectivamente con hipótesis adicionales sobre la distribución de los parámetros.

### Comentarios

Los modelos Bayesianos basados en especificaciones de primer y segundo orden sobre el vector poblacional admiten como caso particular el modelo Bayes normal consistente en asumir la distribución normal sobre el vector poblacional condicionado al parámetro (la media o vector de medias) y sobre el escalar o vector de parámetros. Para los modelos Bayes normales se cumple que el estimador Bayes para la media poblacional (y en general para cualquier función  $g(\mathbf{Y})$  lineal) es lineal dado que la media de la distribución predictiva de  $\mathbf{Y}_u|(s, \mathbf{y}_s)$  es lineal como función de  $\mathbf{y}_s$  (ver en la próxima subsección la ecuación (1.19) junto con la definición de modelo  $m$ -jerárquico normal y los modelos que generaliza). La linealidad del estimador Bayes para los modelos Bayesianos normales implica que el estimador Bayes lineal y el estimador Bayes para  $g(\mathbf{Y})$  una función lineal del vector de características de la población coinciden en estos modelos y por lo tanto sus errores cuadráticos medios (respecto al modelo de superpoblación) asociados, lo que unido a la observación inicial de que los modelos Bayesianos normales son casos particulares de los modelos Bayesianos anteriores nos da como resultado que los estimadores Bayes lineales y sus errores cuadráticos medios asociados para estos modelos coinciden con el estimador Bayes y su error cuadrático medio para el modelo Bayesiano normal correspondiente pudiendo por tanto ser obtenidos a partir de éste.

Una crítica sobre los estimadores Bayes lineales para la media (y en general, para cualquier función  $g(\mathbf{Y})$  lineal del vector poblacional) estriba en que éstos sólo coinciden (bajo condiciones “suaves”) con los estimadores Bayes para modelos Bayesianos *normales*. Efectivamente, sigamos el siguiente razonamiento aplicado al modelo Bayes de regresión y al estimador media poblacional (generalizable de manera análoga al modelo Bayes básico y los modelos Bayesianos en dos etapas y a estimadores de funciones  $g(\mathbf{Y})$  lineales del vector poblacional).

“Supongamos el modelo Bayes de regresión definido al principio de la sección. Entonces, se tiene que:

$$E(\bar{Y}|(s, \mathbf{y}_s), \beta) = \frac{n}{N}\bar{y}_s + \frac{1}{N}[\mathbf{1}_{N-n}^t[\mathbf{X}_u\beta + \mathbf{V}_{us}\mathbf{V}_s^{-1}(\mathbf{y}_s - \mathbf{X}_s\beta)]],$$

y por lo tanto, el estimador Bayes del modelo, viene dado por:

$$\begin{aligned} \hat{Y}_B &= E_{\beta|(s, \mathbf{y}_s)}(E(\bar{Y}|(s, \mathbf{y}_s), \beta)) = \\ &= \frac{n}{N}\bar{y}_s + \frac{1}{N}[\mathbf{1}_{N-n}^t[\mathbf{X}_u E(\beta|(s, \mathbf{y}_s)) + \mathbf{V}_{us}\mathbf{V}_s^{-1}(\mathbf{y}_s - \mathbf{X}_s E(\beta|(s, \mathbf{y}_s)))]], \end{aligned}$$

El estimador Bayes  $\hat{Y}_B$  será pues lineal, cuando  $E(\beta|(s, \mathbf{y}_s))$  sea lineal. Y ello se produce (bajo hipótesis “suaves”) cuando el modelo es normal, ya que Goel y DeGroot(80) [52] demostraron que si:

$$\mathbf{Y}_s = \mathbf{X}_s \boldsymbol{\beta} + \mathbf{e}_s,$$

con  $\mathbf{e}_s$  y  $\boldsymbol{\beta}$  independientes, la hipótesis  $E(\boldsymbol{\beta} | (s, \mathbf{y}_s))$  lineal como función de  $\mathbf{y}_s$  implica bajo condiciones “suaves” la normalidad de la distribución de  $\mathbf{Y}_s | \boldsymbol{\beta}$  y del parámetro  $\boldsymbol{\beta}$ .”

#### 1.4.2 Modelos Bayesianos para distribuciones específicas. Estimadores Bayes y Estimadores Bayes empíricos

Con modelos Bayesianos para distribuciones específicas nos referimos a modelos de superpoblación en los que todas las distribuciones están totalmente especificadas (las del vector poblacional dado el parámetro o parámetros y las de los parámetros). Estudiaremos en esta subsección los siguientes modelos:

- **Modelo Bayes Básico Normal.** Dado el vector poblacional  $\mathbf{Y} = (Y_1, Y_2, \dots, Y_N)^t$  se define por:

$$\begin{aligned} \mathbf{Y} | \mu &\sim N_N(\mathbf{1}_N \mu, \mathbf{V}), \\ \mu &\sim N(b, \sigma_b^2), \end{aligned}$$

donde  $b$  es un escalar conocido,  $\mathbf{V}$  una matriz  $N \times N$  definida positiva conocida y  $\sigma_b^2$  un escalar ( $\sigma_b^2 > 0$ ) conocido.

- **Modelo Bayes de Regresión Normal.** Dado el vector poblacional  $\mathbf{Y} = (Y_1, Y_2, \dots, Y_N)^t$ , para el cual asociada a cada variable  $Y_i$  tenemos una información auxiliar representada por un vector  $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})$  conocido se define por:

$$\begin{aligned} \mathbf{Y} | \boldsymbol{\beta} &\sim N_N(\mathbf{X}\boldsymbol{\beta}, \mathbf{V}), \\ \boldsymbol{\beta} &\sim N_p(\mathbf{b}, \mathbf{B}), \end{aligned}$$

donde  $\mathbf{X} = (\mathbf{x}_1^t, \mathbf{x}_2^t, \dots, \mathbf{x}_N^t)^t$ ,  $\mathbf{b}$  un  $p$ -vector conocido y  $\mathbf{V}$  y  $\mathbf{B}$  son matrices definidas positivas conocidas de dimensiones  $N \times N$  y  $p \times p$  respectivamente. En el caso  $p = 1$ ,  $\mathbf{X} = \mathbf{1}_N$ ,  $\boldsymbol{\beta} = \mu$ ,  $\mathbf{b} = b$  y  $\mathbf{B} = \sigma_b^2 (\mathbf{1}_N \mathbf{1}_N^t)$  vector de unos se recupera el modelo Bayes Básico Normal.

- **Modelo Bayes Básico Normal-Gamma.** Dado el vector poblacional  $\mathbf{Y} = (Y_1, Y_2, \dots, Y_N)^t$  se define por:

$$\begin{aligned} \mathbf{Y} | \mu, \sigma^2 &\sim N_N(\mathbf{1}_N \mu, \sigma^2 \mathbf{W}), \\ \mu | \sigma^2 &\sim N(b, \sigma^2 c), \\ \sigma^2 &\sim Ga^{-1}(a_0, b_0), \end{aligned}$$

donde  $b$  es un escalar conocido,  $\mathbf{W}$  una matriz  $N \times N$  definida positiva conocida y  $a_0$  y  $b_0$  escalares positivos conocidos.

- **Modelo Bayes de Regresión Normal-Gamma.** Dado el vector poblacional  $\mathbf{Y} = (Y_1, Y_2, \dots, Y_N)^t$ , para el cual asociada a cada variable  $Y_i$  tenemos una información auxiliar representada por un vector  $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})$  conocido se define por:

$$\begin{aligned} \mathbf{Y}|\beta, \sigma^2 &\sim N_N(\mathbf{X}\beta, \sigma^2\mathbf{W}), \\ \beta|\sigma^2 &\sim N_p(\mathbf{b}, \sigma^2\mathbf{C}), \\ \sigma^2 &\sim Ga^{-1}(a_0, b_0), \end{aligned}$$

donde  $\mathbf{X} = (\mathbf{x}_1^t, \mathbf{x}_2^t, \dots, \mathbf{x}_N^t)^t$ ,  $\mathbf{b}$  un  $p$ -vector conocido,  $\mathbf{C}$  es una matriz  $p \times p$  definida positiva conocida y  $a_0$  y  $b_0$  escalares positivos conocidos. En el caso  $p = 1$ ,  $\mathbf{X} = \mathbf{1}_N$ ,  $\beta = \mu$ ,  $\mathbf{b} = b$  y  $\mathbf{C} = c$  ( $\mathbf{1}_N$   $N$  vector de unos) se recupera el modelo Bayes Básico Normal-Gamma.

- **Modelo Bayes en  $m$  jerarquías Normal.** Consiste en el siguiente modelo jerárquico:

$$\begin{aligned} \mathbf{Y}|\theta_1 &\sim N(\mathbf{A}_1\theta_1, \mathbf{V}_1), \\ \theta_1|\theta_2 &\sim N(\mathbf{A}_2\theta_2, \mathbf{V}_2), \\ &\vdots \\ \theta_{m-1}|\theta_m &\sim N(\mathbf{A}_m\theta_m, \mathbf{V}_m), \\ \theta_m &\sim N(\mathbf{A}_{m+1}, \mathbf{V}_{m+1}), \end{aligned}$$

donde  $\mathbf{Y}$  es el vector de características de la población (representando tanto una estructura monoetápica como bietápica, trietápica,  $\dots$ , o  $(m-1)$ -etápica de la población),  $\{\mathbf{A}_i\}_{i=1}^{m+1}$  son matrices conocidas y  $\{\mathbf{V}_i\}_{i=1}^{m+1}$  son matrices conocidas definidas positivas, todas ellas de dimensiones adecuadas. ( $\mathbf{A}_{m+1}$  y  $\mathbf{V}_{m+1}$  suelen ser de dimensión 1). Como casos particulares de este modelo señalemos:

- \* para  $m = 1$ ,  $\mathbf{Y} = (Y_1, Y_2, \dots, Y_N)^t$ ,  $\mathbf{A}_1 = \mathbf{1}_N$ ,  $\mathbf{A}_2 = b$ ,  $\mathbf{V}_1 = \mathbf{V}$  y  $\mathbf{V}_2 = \sigma_b^2$ , el modelo Bayes básico normal.
- \* para  $m = 1$ ,  $\mathbf{Y} = (Y_1, Y_2, \dots, Y_N)^t$ ,  $\mathbf{A}_1 = \mathbf{X}_N$ ,  $\mathbf{A}_2 = \mathbf{b}$ ,  $\mathbf{V}_1 = \mathbf{V}$  y  $\mathbf{V}_2 = \mathbf{B}$ , el modelo Bayes de regresión normal.
- \* para  $m = 1$  y  $\mathbf{Y} = (\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_K)^t$  con  $\mathbf{Y}_i = (Y_{i1}, Y_{i2}, \dots, Y_{iN_i})^t$  se tienen los *modelos Bayesianos en dos etapas con una jerarquía normales*.
- \* para  $m = 2$  y  $\mathbf{Y} = (\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_K)^t$  con  $\mathbf{Y}_i = (Y_{i1}, Y_{i2}, \dots, Y_{iN_i})^t$  se tienen los *modelos Bayesianos en dos etapas con dos jerarquías normales*.

- **Modelo Bayes en  $m$  jerarquías Normal-Gamma.** Consiste en el siguiente modelo jerárquico:

$$\begin{aligned} \mathbf{Y}|\theta_1, \sigma^2 &\sim N(\mathbf{A}_1\theta_1, \sigma^2\mathbf{W}_1), \\ \theta_1|\theta_2, \sigma^2 &\sim N(\mathbf{A}_2\theta_2, \sigma^2\mathbf{W}_2), \\ &\vdots \\ \theta_{m-1}|\theta_m, \sigma^2 &\sim N(\mathbf{A}_m\theta_m, \sigma^2\mathbf{W}_m), \\ \theta_m|\sigma^2 &\sim N(\mathbf{A}_{m+1}, \sigma^2\mathbf{W}_{m+1}), \\ \sigma^2 &\sim Ga^{-1}(a_0, b_0), \end{aligned}$$

donde  $\mathbf{Y}$  es el vector de características de la población (representando tanto una estructura monoetápica como bietápica, trietápica,  $\dots$ , o  $((m-1)$ -etápica),  $\{\mathbf{A}_i\}_{i=1}^{m+1}$  son matrices conocidas,  $\{\mathbf{W}_i\}_{i=1}^{m+1}$  son matrices conocidas definidas positivas, todas ellas de dimensiones adecuadas y  $a_0$  y  $b_0$  son escalares positivos conocidos. ( $\mathbf{A}_{m+1}$  y  $\mathbf{W}_{m+1}$  suelen ser de dimensión 1). Como casos particulares de este modelo señalemos:

- \* para  $m = 1$ ,  $\mathbf{Y} = (Y_1, Y_2, \dots, Y_N)^t$ ,  $\mathbf{A}_1 = \mathbf{1}_N$ ,  $\mathbf{A}_2 = b$ ,  $\mathbf{W}_1 = \mathbf{W}$  y  $\mathbf{W}_2 = c$ , el modelo Bayes básico normal-gamma.
- \* para  $m = 1$ ,  $\mathbf{Y} = (Y_1, Y_2, \dots, Y_N)^t$ ,  $\mathbf{A}_1 = \mathbf{X}_N$ ,  $\mathbf{A}_2 = b$ ,  $\mathbf{W}_1 = \mathbf{W}$  y  $\mathbf{W}_2 = \mathbf{C}$ , el modelo Bayes de regresión normal-gamma.
- \* para  $m = 1$  y  $\mathbf{Y} = (\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_K)^t$  con  $\mathbf{Y}_i = (Y_{i1}, Y_{i2}, \dots, Y_{iN_i})^t$  se tienen los *modelos Bayesianos en dos etapas con una jerarquía normales-gamma*.
- \* para  $m = 2$  y  $\mathbf{Y} = (\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_K)^t$  con  $\mathbf{Y}_i = (Y_{i1}, Y_{i2}, \dots, Y_{iN_i})^t$  se tienen los *modelos Bayesianos en dos etapas con dos jerarquías normales-gamma*.

Para estos modelos (que son generalizaciones de los modelos habitualmente aplicados, y serán comentados y analizados exhaustivamente en el capítulo 2 y 3) las distribuciones predictivas, los estimadores Bayes para una función  $g(\mathbf{Y})$  lineal y sus errores pueden calcularse con exactitud (puesto que se trata de modelos conjugados). Un primer inconveniente de estos modelos estriba en que las hipótesis sobre el conocimiento de las varianzas asumidas pueden ser muy fuertes para la modelización de una población finita dada. La hipótesis de varianzas conocidas de los modelos normales se debilita en los modelos normal-gamma al conocimiento de los cocientes entre varianzas (hipótesis extensamente aplicada en modelización Bayesiana y en ocasiones bastante realista) y la siguiente debilitación a cociente entre varianzas desconocidos conduce a modelos en los que la distribución predictiva, estimadores Bayes y sus errores requieren ser aproximados mediante procedimientos numéricos (en los capítulos 2 y 3 estudiaremos algunos modelos para modelizar estas situaciones).

En esta subsección nos restringiremos a citar los resultados obtenidos dentro de esta línea de trabajo para la estimación de la media poblacional para los modelos habitualmente estudiados por la bibliografía y a la descripción de algunos modelos presentados por la misma para la situación de cociente entre varianzas desconocido.

### Estimación

El criterio de estimación aplicado para estos modelos (con varianza o cociente entre varianzas conocido) es la obtención del correspondiente estimador Bayes. Recordemos que dada una muestra  $(s, \mathbf{y}_s)$ , el estimador Bayes  $g_{\mathbf{B}}(\widehat{\mathbf{y}}_s)$  de la función  $g(\mathbf{Y})$  y su error cuadrático medio respecto al modelo de superpoblación vienen dados respectivamente por:

$$g_{\mathbf{B}}(\widehat{\mathbf{y}}_s) = E_{\mathbf{Y}_{\cdot}|(s, \mathbf{y}_s)}(g),$$

$$E_{\xi}(g_{\mathbf{B}}(\widehat{\mathbf{y}}_s) - g(\mathbf{Y}))^2 = E_{\mathbf{Y}_{\cdot}}[Var_{\mathbf{Y}_{\cdot}|(s, \mathbf{y}_s)}(g(\mathbf{Y}))],$$

expresiones que en el caso de la estimación de la media poblacional se reducen a:

$$\begin{aligned}\widehat{Y}_B &= \frac{n}{N}\bar{y}_s + \frac{N-n}{N}\widehat{Y}_u^B, \\ E_\xi(\widehat{Y}_B - \bar{Y})^2 &= \frac{(N-n)^2}{N^2}E_{\mathbf{Y}_s}[Var_{\mathbf{Y}_u|(s, \mathbf{y}_s)}(\bar{Y}_u)],\end{aligned}$$

donde  $\widehat{Y}_u^B = E(\bar{Y}_u|(s, \mathbf{y}_s)) = (N-n)^{-1}\mathbf{1}_{N-n}^t E(\mathbf{Y}_u|(s, \mathbf{y}_s))$  y  $Var_{\mathbf{Y}_u|(s, \mathbf{y}_s)}(\bar{Y}_u) = (N-n)^{-2}\mathbf{1}_{N-n}^t Var(\mathbf{Y}_u|(s, \mathbf{y}_s))\mathbf{1}_{N-n}$ . Con lo que el cálculo de los estimadores Bayes de  $\bar{Y}$  y de su error cuadrático medio condicionado a la muestra (para obtener el error cuadrático respecto al modelo de superpoblación se tiene que promediar el error cuadrático medio condicionado a la muestra respecto a la distribución del vector de características observadas) se reduce a la obtención del primer y segundo momentos de la distribución predictiva de  $\mathbf{Y}_u|(s, \mathbf{y}_s)$ .

### Estimadores Bayes de la media poblacional

Los estimadores Bayes para la media poblacional y para los modelos descritos son:

1. **Estimador Bayes para el modelo Bayes básico normal.** El estimador Bayes para el modelo Bayes básico normal y su error cuadrático medio vienen dados respectivamente por:

$$\begin{aligned}\widehat{Y}_{B1} &= \frac{n}{N}\bar{y}_s + \frac{N-n}{N}\widehat{\mu}_B + \frac{1}{N}\mathbf{1}_{N-n}^t \mathbf{V}_{us} \mathbf{V}_s^{-1}(\mathbf{y}_s - \mathbf{1}_n \widehat{\mu}_B), \\ E_\xi(\widehat{Y}_{B1} - \bar{Y})^2 &= N^{-2}\{\mathbf{1}_{N-n}^t (\mathbf{V}_u + \mathbf{V}_{us} \mathbf{V}_s^{-1} \mathbf{V}_{us}^t) \mathbf{1}_{N-n} + \\ &\quad + (N-n - \mathbf{1}_{N-n}^t \mathbf{V}_{us} \mathbf{V}_s^{-1} \mathbf{1}_{N-n})^2 (\mathbf{1}_n^t \mathbf{V}_s^{-1} \mathbf{1}_n + \sigma_b^{-2})^{-1}\},\end{aligned}\quad (1.17)$$

donde  $\bar{y}_s$  es la media muestral (1.1),  $\mathbf{V}_s = Var_{\mathbf{Y}|\mu}(\mathbf{Y}_s)$ ,  $\mathbf{V}_u = Var_{\mathbf{Y}|\mu}(\mathbf{Y}_u)$ ,  $\mathbf{V}_{su} = Cov_{\mathbf{Y}|\mu}(\mathbf{Y}_s, \mathbf{Y}_u)$  y  $\widehat{\mu}_B = E(\mu|(s, \mathbf{y}_s)) = (\mathbf{1}_n^t \mathbf{V}_s^{-1} \mathbf{1}_n + \sigma_b^{-2})^{-1}(\mathbf{1}_n^t \mathbf{V}_s^{-1} \mathbf{y}_s + \sigma_b^{-2} b)$ , que coincide con el estimador lineal Bayes y su error para el modelo Bayesiano básico dado en (1.16) (ver subsección 1.4.1. apartado comentarios). El estimador y su error cuadrático medio asociado en la formulación arriba señalada proceden de los resultados de Bolfarine et al(87)[8] y Bolfarine y Zacks(91,92)[9, 10] para el modelo Bayesiano de regresión normal particularizado para  $p = 1$ ,  $\mathbf{X} = \mathbf{1}_N$ ,  $\mathbf{b} = b$  y  $\mathbf{B} = \sigma_b^2$ . Como casos particulares de (1.17), señalemos (además de (1.16)):

- para  $\mathbf{V} = \mathbf{I}_N \sigma^2$  se recuperan los resultados de Ericson(69)[26] y Murgui(82)[76].
- para  $\sigma_b^{-2} \rightarrow 0$  se recupera el estimador predictivo para el modelo clásico básico con matriz de varianzas-covarianzas conocida dado en (1.11).
- para  $\mathbf{V} = \mathbf{I}_N \sigma^2$  y  $\sigma_b^{-2} \rightarrow 0$  se recupera el estimador media muestral (1.1) (Este resultado lo podemos ver en Ericson(69)[26] y Murgui(82)[76]).

2. **Estimador Bayes para el modelo Bayes de regresión normal.** El estimador Bayes para el modelo Bayes de regresión normal y su error cuadrático medio vienen dados respectivamente por:

$$\begin{aligned}\widehat{Y}_{B2} &= \frac{n}{N}\bar{y}_s + \frac{1}{N}[\mathbf{1}_{N-n}^t[\mathbf{X}_u\widehat{\beta}_B + \mathbf{V}_{us}\mathbf{V}_s^{-1}(y_s - \mathbf{X}_s\widehat{\beta}_B)]], \\ E_{\xi}(\widehat{Y}_{B2} - \bar{Y})^2 &= N^{-2}\{\mathbf{1}_{N-n}^t(\mathbf{V}_u + \mathbf{V}_{us}\mathbf{V}_s^{-1}\mathbf{V}_{us}^t)\mathbf{1}_{N-n} + \quad (1.18) \\ &\quad + \mathbf{1}_{N-n}^t(\mathbf{X}_u - \mathbf{V}_{us}\mathbf{V}_s^{-1}\mathbf{X}_s)(\mathbf{X}_s^t\mathbf{V}_s^{-1}\mathbf{X}_s + \mathbf{B}^{-1})^{-1} \times \\ &\quad \times (\mathbf{X}_u - \mathbf{V}_{us}\mathbf{V}_s^{-1}\mathbf{X}_s)^t\mathbf{1}_{N-n}\},\end{aligned}$$

donde  $\bar{y}_s$  es la media muestral (1.1),  $\mathbf{y}_s = (y_{i_1}, y_{i_2}, \dots, y_{i_n})^t$ ,  $\mathbf{Y}_u = (Y_i : i \notin s)^t$ ,  $\mathbf{X}_s = (\mathbf{x}_{i_1}^t, \mathbf{x}_{i_2}^t, \dots, \mathbf{x}_{i_n}^t)^t$ ,  $\mathbf{X}_u = (\mathbf{x}_i^t : i \notin s)^t$ ,  $\mathbf{V}_s = \text{Var}_{\mathbf{Y}|s}\mathbf{Y}_s$ ,  $\mathbf{V}_u = \text{Var}_{\mathbf{Y}|s}\mathbf{Y}_u$ ,  $\mathbf{V}_{su} = \text{Cov}_{\mathbf{Y}|s}(\mathbf{Y}_s, \mathbf{Y}_u)$  y

$$\widehat{\beta}_B = E(\beta|(s, \mathbf{y}_s)) = (\mathbf{X}_s^t\mathbf{V}_s^{-1}\mathbf{X}_s + \mathbf{B}^{-1})^{-1}(\mathbf{X}_s^t\mathbf{V}_s^{-1}\mathbf{y}_s + \mathbf{B}^{-1}\mathbf{b}),$$

que coincide con el estimador y error para el modelo Bayesiano de regresión dado en (1.15) (ver subsección 1.4.1. apartado comentarios). El estimador y su error cuadrático medio asociado en la formulación arriba señalada proceden de los resultados de Bolfarine et al(87)[8] y Bolfarine y Zacks(91,92)[9,10] que obtienen además la distribución predictiva de  $\mathbf{Y}_u|(s, \mathbf{y}_s)$ . Como casos particulares de (1.18) señalemos (además de (1.15)):

- para  $p = 1$ ,  $\mathbf{X} = \mathbf{1}_N$ ,  $\mathbf{b} = b$  y  $\mathbf{B} = \sigma_b^2$  se recupera el estimador Bayes y error cuadrático medio asociado para el modelo Bayesiano básico normal (1.17).
- para  $\mathbf{B}^{-1} \rightarrow \mathbf{0}_{p \times p}$  se recupera el estimador predictivo para el modelo clásico de regresión con matriz de varianzas-covarianzas conocida dado en (1.10) (Resultados de Royall y Pfeffermann(82)[97]).
- para  $p = 1$ ,  $\mathbf{x}_i = x_i$ ,  $i = 1, 2, \dots, K$  y  $\mathbf{V} = \text{diag}_{i=1}^N\{x_i\}$  se recupera el estimador de razón (1.2).
- para  $p = 1$ ,  $\mathbf{x}_i = \alpha_i = \text{Pr}(i \in s)$ ,  $i = 1, 2, \dots, N$ ,  $\mathbf{V} = \text{diag}_{i=1}^N\{\frac{\alpha_i^2}{1-\alpha_i}\}$  y  $\mathbf{B}^{-1} \rightarrow 0$  se recupera el estimador de Horwitz y Thomson (1.4). (Resultado de Ghosh y Sinha(90)[43])

3. **Estimador Bayes para el modelo Bayes básico normal-gamma.** El artículo de Font(95)[33] describe la distribución predictiva de  $\mathbf{Y}_u|(s, \mathbf{y}_s)$ , el estimador Bayes de la media poblacional y su varianza condicionada a la muestra (que coincide al no depender de  $\mathbf{y}_s$  con el error cuadrático medio respecto al modelo de superpoblación), para el modelo Bayes de regresión normal-gamma, a partir de los cuales se pueden obtener los resultados para este modelo tomando  $p = 1$ ,  $\mathbf{x}_i = x_i = 1$ ,  $i = 1, 2, \dots, N$ . Los resultados completos y sus demostraciones se presentarán en la subsección 2.2.2. Como casos particulares de estos resultados señalemos:

- para  $\mathbf{W} = \mathbf{I}_N$  se recuperan los resultados de Ericson(69)[26] y Murgui(82)[76].
  - para  $\mathbf{W} = \mathbf{I}_N$  y  $\pi(\mu, \sigma^{-2}) \propto 1/\sigma^{-2}$  se recuperan los resultados de Ericson(69)[26] y Murgui(82)[76], que obtienen el estimador media muestral dado en (1.1).
4. **Estimador Bayes para el modelo Bayes de regresión normal-gamma.** Tal y como indicábamos en el párrafo anterior, las inferencias sobre este modelo aparecen descritas en Font(95)[33] y su estudio completo se realizará en la subsección 2.2.2 de esta tesis. Como casos particulares de estos resultados señalemos:
- para  $p = 1$ ,  $\mathbf{x}_i = x_i$ ,  $i = 1, 2, \dots, N$  y  $\mathbf{W} = \text{diag}_{i=1}^N \{v(x_i)\}$  con  $v : \mathfrak{R} \rightarrow \mathfrak{R}^+$  se recuperan los resultados de Ericson(69)[26] y Murgui(82)[76].
  - para  $p = 1$ ,  $\mathbf{x}_i = x_i$ ,  $i = 1, 2, \dots, N$ ,  $\mathbf{W} = \text{diag}_{i=1}^N \{v(x_i)\}$  con  $v : \mathfrak{R} \rightarrow \mathfrak{R}^+$  y  $\pi(\mu, \sigma^{-2}) \propto \frac{1}{\sigma^{-2}}$  se recuperan los resultados de Ericson(69)[26] y Murgui(82)[76].
  - para  $p = 2$   $\mathbf{x}_i = (1, x_i)$ ,  $i = 1, 2, \dots, N$  y  $\mathbf{W} = \text{diag}_{i=1}^N \{v(x_i)\}$  con  $v : \mathfrak{R} \rightarrow \mathfrak{R}^+$  se recuperan los resultados de Murgui(82)[76].
  - para  $\mathbf{W}_{us} = \mathbf{0}_{(N-n) \times n}$  se recuperan los resultados de Bolfarine y Zacks(91,92)[9,10]
  - para  $p = 1$ ,  $\mathbf{x}_i = x_i$ ,  $i = 1, 2, \dots, N$ ,  $\mathbf{W} = \text{diag}_{i=1}^N \{v(x_i)\}$  y  $\pi(\mu, \sigma^{-2}) \propto \frac{1}{\sigma^{-2}}$  se recupera el estimador predictivo para el modelo simplificado de regresión de Royall(70)[91] dado en (1.9).
5. **Estimador Bayes modelo Bayes en  $m$  jerarquías normal.** Malec y Sedransk (85)[72] estudiaron el modelo Bayesiano en  $m$  jerarquías normal (en su original el último nivel  $\theta_m$  es escalar) obteniendo que el vector  $\mathbf{Y}_u|(s, \mathbf{y}_s)$  se distribuye normal  $(N - n)$ -variante con media y varianza dadas respectivamente por:

$$\begin{aligned}
E(\mathbf{Y}_u|(s, \mathbf{y}_s)) &= \mathbf{A}_u \mathbf{D} \mathbf{d} + \mathbf{V}_{su}^t \mathbf{V}_s^{-1} (\mathbf{y}_s - \mathbf{A}_s \mathbf{D} \mathbf{d}), \\
\text{Var}(\mathbf{Y}_u|(s, \mathbf{y}_s)) &= \mathbf{V}_u - \mathbf{V}_{su}^t \mathbf{V}_s^{-1} \mathbf{V}_{su} + \\
&\quad + (\mathbf{A}_u - \mathbf{V}_{su}^t \mathbf{V}_s^{-1} \mathbf{A}_s) \mathbf{D} (\mathbf{A}_u - \mathbf{V}_{su}^t \mathbf{V}_s^{-1} \mathbf{A}_s)^t,
\end{aligned} \tag{1.19}$$

donde  $\mathbf{y}_s = (y_i : i \in s)^t$ ,  $\mathbf{Y}_u = (Y_i : i \notin s)^t$ ,  $\mathbf{A}_s$  es la matriz formada por las filas de la matriz  $\mathbf{A}_1$  correspondientes a índices  $i \in s$ ,  $\mathbf{A}_u$  es la matriz formada por las filas de la matriz  $\mathbf{A}_1$  correspondientes a índices  $i \notin s$ ,  $\mathbf{V}_s = \text{Var}_{\mathbf{Y}|\theta_1}(\mathbf{Y}_s)$ ,  $\mathbf{V}_u = \text{Var}_{\mathbf{Y}|\theta_1}(\mathbf{Y}_u)$ ,  $\mathbf{V}_{su} = \text{Cov}_{\mathbf{Y}|\theta_1}(\mathbf{Y}_s, \mathbf{Y}_u)$  y

$$\begin{aligned}
\mathbf{D}^{-1} &= \mathbf{A}_s^t \mathbf{V}_s^{-1} \mathbf{A}_s + (\mathbf{V} + \mathbf{A} \mathbf{V}_{m+1} \mathbf{A}^t)^{-1}, \\
\mathbf{d} &= \mathbf{A}_s^t \mathbf{V}_s^{-1} \mathbf{y}_s + (\mathbf{V} + \mathbf{A} \mathbf{V}_{m+1} \mathbf{A}^t)^{-1} \mathbf{A} \mathbf{A}_{m+1}, \\
\mathbf{A} &= \mathbf{A}_2 \mathbf{A}_3 \cdots \mathbf{A}_m, \\
\mathbf{V} &= \mathbf{V}_2 + \mathbf{A}_2 \mathbf{V}_3 \mathbf{A}_2^t + \dots + (\mathbf{A}_2 \mathbf{A}_3 \cdots \mathbf{A}_{m-1}) \mathbf{V}_m (\mathbf{A}_2 \mathbf{A}_3 \cdots \mathbf{A}_{m-1})^t.
\end{aligned}$$

De donde se obtiene que el estimador Bayes y el error cuadrático medio para la media poblacional para este modelo vendrá dados por<sup>5</sup>:

<sup>5</sup>Dado que la varianza de la distribución predictiva no depende del vector poblacional observado  $\mathbf{y}_s$ , y por lo tanto  $E_{\mathbf{Y}_s}[\text{Var}(\mathbf{Y}_u|(s, \mathbf{y}_s))] = \text{Var}(\mathbf{Y}_u|(s, \mathbf{y}_s))$

$$\begin{aligned}\widehat{Y}_{B3} &= \frac{n}{N}\bar{y}_s + \frac{1}{N}\mathbf{1}_{N-n}^t E(\mathbf{Y}_u|\mathbf{y}_s), \\ E_\xi(\widehat{Y}_{B3} - \bar{Y})^2 &= \frac{1}{N^2}\mathbf{1}_{N-n}^t \text{Var}(\mathbf{Y}_u|\mathbf{y}_s)\mathbf{1}_{N-n},\end{aligned}\quad (1.20)$$

con  $E(\mathbf{Y}_u|(s, \mathbf{y}_s))$  y  $\text{Var}(\mathbf{Y}_u|(s, \mathbf{y}_s))$  las definidas en (1.19).

Como casos particulares de (1.20), señalemos:

- para  $m = 1$ ,  $\mathbf{Y} = (Y_1, Y_2, \dots, Y_N)^t$ ,  $\mathbf{A}_1 = \mathbf{X}_N$ ,  $\mathbf{A}_2 = \mathbf{b}$ ,  $\mathbf{V}_1 = \mathbf{V}$  y  $\mathbf{V}_2 = \mathbf{B}$  se obtiene el estimador para el modelo Bayesiano de regresión normal (1.18).
- para  $m = 1$ ,  $\mathbf{Y} = (\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_K)^t$  con  $\mathbf{Y}_i = (Y_{i1}, Y_{i2}, \dots, Y_{iN_i})^t$ ,  $\mathbf{A}_1 = \mathbf{X}$  con  $\mathbf{X} = (\mathbf{X}_1^t, \mathbf{X}_2^t, \dots, \mathbf{X}_K^t)^t$ ,  $\mathbf{X}_i = (\mathbf{x}_{i1}^t, \mathbf{x}_{i2}^t, \dots, \mathbf{x}_{iN_i}^t)^t$ ,  $i = 1, 2, \dots, K$ ,  $\mathbf{x}_{ij} = (x_{ij1}, x_{ij2}, \dots, x_{ijp})$ ,  $j = 1, 2, \dots, N_i$ ,  $i = 1, 2, \dots, K$ ,  $\theta_1 = (\beta_1^t, \beta_2^t, \dots, \beta_K^t)^t$ ,  $\beta_i = (\beta_{i1}, \beta_{i2}, \dots, \beta_{ip})^t$ ,  $i = 1, 2, \dots, K$ ,  $\mathbf{A}_2 = \underbrace{(\mathbf{I}_p, \dots, \mathbf{I}_p)^t}_K \beta$ ,  $\mathbf{V}_1 = \text{diag}_{i=1}^K \{\mathbf{V}_i\}$  y  $\mathbf{V}_2 = \text{diag}_{i=1}^K \{\mathbf{B}_\beta\}$  (modelo Bayes de regresión en dos etapas con una jerarquía normal) se obtiene el *estimador Bayes de regresión en dos etapas de una jerarquía normal* y su error cuadrático medio asociado. Murgui(82)[76] estudia el caso particular con  $\mathbf{A}_1 = \text{diag}_{i=1}^K \{\mathbf{1}_{N_i}\}$ ,  $\mathbf{V}_1 = \text{diag}_{i=1}^K \{\mathbf{I}_{N_i} \sigma_i^2\}$ ,  $\mathbf{A}_2 = \mathbf{1}_K \mu$ ,  $\theta_1 = \mu = (\mu_1, \mu_2, \dots, \mu_K)^t$  y  $\mathbf{V}_2 = \mathbf{I}_K \sigma_0^2$ , para  $k = K$  en los casos  $\sigma_0^2$  escalar positivo conocido y  $\sigma_0^{-2} \rightarrow 0$ .
- para  $m = 2$ , las definiciones de ítem anterior para  $\mathbf{Y}$ ,  $\mathbf{A}_1$ ,  $\theta_1$ ,  $\mathbf{V}_1$  y  $\mathbf{V}_2$ , y  $\theta_2 = \beta$ ,  $\mathbf{A}_2 = \underbrace{(\mathbf{I}_p, \dots, \mathbf{I}_p)^t}_K$ ,  $\mathbf{A}_3 = \beta_0$  y  $\mathbf{V}_3 = \mathbf{B}_\beta$  (modelo Bayesiano de regresión en dos etapas con dos jerarquías normal) se obtiene el *estimador Bayes de regresión en dos etapas con dos jerarquías normal* y su error cuadrático medio asociado. Font(95)[33] describe la distribución predictiva del vector poblacional no observado dado la muestra, el estimador Bayes de la media poblacional y su varianza condicionada a la muestra para este modelo, y Murgui(82)[76] estudia el caso particular con  $\mathbf{A}_1 = \text{diag}_{i=1}^K \{\mathbf{1}_{N_i}\}$ ,  $\mathbf{V}_1 = \text{diag}_{i=1}^K \{\mathbf{I}_{N_i} \sigma_i^2\}$ ,  $\mathbf{A}_2 = \mathbf{1}_K$ ,  $\mathbf{A}_3 = \mu_0$ ,  $\theta_1 = \mu = (\mu_1, \mu_2, \dots, \mu_K)^t$ ,  $\theta_2 = \mu_\pi$ ,  $\mathbf{V}_2 = \mathbf{I}_K \sigma^2$  y  $\mathbf{V}_3 = \sigma_0^2$  en los casos  $\sigma_0^2$  escalar positivo y  $\sigma_0^{-2} \rightarrow 0$ . Scott y Smith(69)[106] en un trabajo anterior estudio el mismo modelo de Murgui para el caso  $\sigma_0^{-2} \rightarrow 0$ .
- Malec y Sedransk(85)[72] estudian un caso particular  $m = 3$  de un modelo Bayesiano para una población estructurada en tres etapas con dos jerarquías con distribución sobre el último parámetro no informativa.

El inconveniente del modelo y sobre todo de las expresiones (1.19) y (1.20) es su complejidad matricial que hace difícil la interpretación de los resultados obtenidos. En esta tesis nos centraremos en modelos Bayesianos para estructuras de población monoetápicas y bietápicas con una y dos jerarquías y estudiaremos en estos casos particulares expresiones más fáciles de interpretar (subsecciones 2.3.1, 2.3.4, 3.2.1 y 3.2.4).

**6. Estimador Bayes para el modelo Bayes en  $m$  jerarquías normal-gamma.**

No tengo referencias de la obtención del estimador Bayes para este modelo concretamente (la distribución predictiva, el estimador Bayes para la media poblacional y su varianza condicionada a la muestra se obtendrán en la subsección 3.4.2 y 3.4.3 en un caso mínimo informativo). Bernardo y Girón(91)[7] estudian un modelo  $m$  jerárquico en el que se modeliza la situación en la que para cada elemento queremos estudiar un vector de característica, que se puede particularizar al modelo estudiado en este párrafo en el caso  $Cov_{\mathbf{Y}|\theta_1, \sigma^2}(\mathbf{Y}_s, \mathbf{Y}_u) = \mathbf{0}$ . Font(95)[33] describe el caso  $m = 2$ ,  $\mathbf{Y} = (\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_K)^t$  con  $\mathbf{Y}_i = (Y_{i1}, Y_{i2}, \dots, Y_{iN_i})^t$ ,  $\mathbf{A}_1 = \mathbf{X}$ ,  $\theta_1 = (\beta_1^t, \beta_2^t, \dots, \beta_K^t)^t$ ,  $\theta_2 = \beta$ ,  $\mathbf{A}_2 = (\mathbf{I}_p, \dots, \mathbf{I}_p)^t$ ,

$\mathbf{A}_3 = \beta_0$ ,  $\mathbf{W}_1 = diag_{i=1}^K \{\mathbf{W}_i\}$ ,  $\mathbf{W}_2 = diag_{i=1}^K \{\mathbf{C}_\beta\}$  y  $\mathbf{W}_3 = \mathbf{C}_{\beta_0}$ , (modelo Bayesiano de regresión en dos etapas con dos jerarquías normal-gamma), que es estudiado en la subsección 3.2.2 y 3.2.4 en un caso mínimo informativo, y Murgui(82)[76] estudia el caso particular con:  $m = 2$ ,  $\mathbf{Y} = (\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_K)^t$  con  $\mathbf{Y}_i = (Y_{i1}, Y_{i2}, \dots, Y_{iN_i})^t$ ,  $\theta_1 = \mu$ ,  $\theta_2 = \mu_\pi$ ,  $\mathbf{A}_1 = diag_{i=1}^K \{\mathbf{1}_{N_i}\}$ ,  $\mathbf{A}_2 = \mathbf{1}_K$ ,  $\mathbf{A}_3 = \mu$ ,  $\mathbf{W}_1 = diag_{i=1}^K \{\mathbf{I}_{N_i} c_i\}$ ,  $\mathbf{W}_2 = \mathbf{I}_K$  y  $\mathbf{W}_3 = c_0$ . Otros casos particulares son los trabajos de Ericson(69)[26] y Murgui(82)[76] que hemos indicado en los items de modelos normales-gamma anteriores.

**Otros modelos con varianzas desconocidas**

Dedicaremos este apartado a presentar algunos trabajos para modelos Bayesianos con varianzas desconocidas y los resultados obtenidos. Estos modelos nos servirán de inspiración para proponer y estudiar en los capítulos 2 y 3 (subsecciones 2.2.3, 2.3.3 y 3.2.3), tres modelos para la situación de varianza desconocida y cociente entre varianzas asimismo desconocido. Citemos:

1. **El modelo de Scott y Smith(69)[106].** Es un modelo para estructuras de población en dos etapas que generaliza el modelo Bayes básico en dos etapas con una jerarquía normal-gamma y viene definido por:

$$\begin{aligned} Y_{ij} | \beta_i, \sigma^2 &\sim N(\beta_i, \sigma^2), \text{ i.i.d., } j = 1, 2, \dots, N_i, i = 1, 2, \dots, K, \\ \beta_i | b, \sigma_b^2 &\sim N(b, \sigma_b^2), \text{ i.i.d., } i = 1, 2, \dots, K, \\ \pi(b, \sigma^2, \sigma_b^2) &\propto \lambda \sigma^{-2} \sigma_b^{-2}, \quad \lambda = \frac{\sigma_b^2}{\sigma_b^2 + \frac{\sigma^2}{m}}, \end{aligned}$$

donde  $m$  es el número de elementos muestreados en cada unidad con  $n_i = m$ ,  $i \in \{i_1, i_2, \dots, i_k\}$ . Scott y Smith obtienen  $E(\mathbf{Y} | (s, \mathbf{y}_s))$  y por lo tanto el estimador Bayes para la media poblacional. Entre los puntos de crítica al modelo señalar la dependencia de la distribución inicial asignada de  $m$ , la longitud de los cálculos para obtener  $E(\mathbf{Y} | (s, \mathbf{y}_s))$  y la no aportación de una valoración del error cometido al aplicar el estimador Bayes.

2. **El modelo de Ghosh y Meeden(86)[42].** Es un modelo para estructuras de población monoetápicas o bietápicas y se define por:

$$\begin{aligned} \mathbf{Y}|\beta, \sigma^2 &\sim N(\mathbf{1}_N\beta, \mathbf{I}_N\sigma^2), \\ \beta|b, \sigma_b^2 &\sim N(b, \sigma_b^2), \end{aligned}$$

con  $b$  y  $R = \frac{\sigma^2}{\sigma_b^2}$  desconocidos ( $\sigma > 0$  y  $\sigma_b > 0$ ). Ghosh y Meeden obtienen una solución empírica Bayes para la estimación de la media poblacional.

3. **El modelo de Pérez y Pericchi(91)[80]**. Es un modelo para estructuras de población en dos etapas que generaliza el modelo Bayes básico en dos etapas y dos jerarquías normal-gamma, y se define por:

$$\begin{aligned} Y_{ij}|\beta_i, \sigma^2 &\sim N(\beta_i, \sigma^2), \quad \text{i.i.d.}, \quad j = 1, 2, \dots, N_i, \quad i = 1, 2, \dots, K, \\ \beta_i|b, \sigma_b^2 &\sim N(b, \sigma_b^2), \quad \text{i.i.d.}, \quad i = 1, 2, \dots, K, \\ \pi(b) &\propto 1, \\ \frac{\nu_w \lambda_w}{\sigma^2} &\sim \chi_{\nu_w}^2, \\ \frac{\nu_b \lambda_b}{\sigma_b^2} &\sim \chi_{\nu_b}^2, \end{aligned}$$

con  $\nu_w$ ,  $\lambda_w$ ,  $\nu_b$  y  $\lambda_b$  escalares positivos conocidos. Pérez y Pericchi obtienen la distribución de  $\beta|(s, \mathbf{y}_s)$  salvo constante de integración (producto de dos  $t$  multivariantes) y su estimador modal.

4. **Modelos con un parámetro varianza conocido**. Son modelos para estructuras monoetápicas (generalizables a estructuras bietápicas) consistentes en asumir:

$$\begin{aligned} Y_i|\beta_i, \sigma_i^2 &\sim N(\beta_i, \sigma_i^2), \quad \text{independientes}, \quad i = 1, 2, \dots, N, \\ \beta_i|b, \sigma_{b_i}^2 &\sim N(\mathbf{A}_i^t \mathbf{b}, \sigma_{b_i}^2), \quad \text{independientes}, \quad i = 1, 2, \dots, N, \end{aligned}$$

con  $\sigma_i$ ,  $i = 1, 2, \dots, N$  escalares positivos conocidos,  $\mathbf{A}_i$  un  $p$ -vector conocido y diversos niveles de conocimiento sobre  $\mathbf{b}$  y  $\sigma_{b_i}^2$ ,  $i = 1, 2, \dots, N$ . Los resultados se centran en la estimación de  $\beta_i$ ,  $i = 1, 2, \dots, N$  dada la muestra, citemos los trabajos de Fay y Herriot(79)[31] (que obtiene un estimador Bayes empírico de  $\beta_i$ ,  $i = 1, 2, \dots, N$  para el caso particular  $\mathbf{A}_i = \mathbf{1}$ ,  $i = 1, 2, \dots, N$  y  $b$  y  $\sigma_{b_i}^2 = \sigma_b^2$ ,  $i = 1, 2, \dots, N$  desconocidos), Datta y Lahiri(94)[23] (que obtienen el estimador Bayes de  $\beta_i$ ,  $i = 1, 2, \dots, N$  para  $\mathbf{b}$  conocido y  $\sigma_{b_i}^2$  distribuidas según gammas invertidas independientes para  $i = 1, 2, \dots, N$ ) y los resultados presentados por Ghosh(94) en la NSF-CBMS Regional Conference on Bayesian Methods in Finite Populations Sampling. Theory and Application, en los que se estudiaba una generalización del modelo anterior para estructuras bietápicas presentando estimadores Bayes empíricos y las distribuciones condicionales necesarias para aplicar Gibbs Sampler y obtener el estimador Bayes por métodos numéricos del  $N_i$ - vector  $\beta_i$ ,  $i = 1, 2, \dots, K$ .

5. **Modelo mixto de regresión de Datta y Ghosh(91)[22]**. Es un modelo para estructuras en una (tomar  $K = 1$ ) o dos etapas consistente en asumir:

$$\begin{aligned}
\mathbf{Y} &= \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{v} + \mathbf{e}, \\
\mathbf{v}|\{\lambda_i\}_{i=1}^t, \sigma^2 &\sim N_q(\mathbf{0}_q, \sigma^2 \text{diag}_{i=1}^t \{\lambda_i \mathbf{I}_{q_i}\}), \\
\mathbf{e}|\sigma^2 &\sim N_N(\mathbf{0}_N, \sigma^2 \mathbf{I}_N), \\
\pi(\boldsymbol{\beta}) &\propto 1, \\
\sigma^2 &\sim Ga^{-1}(a_0, b_0), \\
\lambda_i \sigma^2 &\sim Ga^{-1}(a_i, b_i), \text{ independientes, } i = 1, 2, \dots, t,
\end{aligned}$$

donde  $\mathbf{X}$  y  $\mathbf{Z}$  son matrices de covariables conocidas de dimensiones  $N \times p$  y  $N \times q$  respectivamente,  $\sum_{i=1}^t q_i = q$  y  $a_0, b_0, \{a_i\}_{i=1}^t, \{b_i\}_{i=1}^t$  son escalares conocidos ( $a_0 \geq 0, b_0 \geq 0, a_i \geq 0, i = 1, 2, \dots, t$  y  $b_i > 0, i = 1, 2, \dots, t$ ). (Nótese que  $Var(\mathbf{v}|\{\lambda_i\}_{i=1}^t, \sigma^2) = \text{diag}_{i=1}^t \{\sigma_i^2 \mathbf{I}_{q_i}\}$  con  $\sigma_i^2 = (\sigma_i^2/\sigma^2)\sigma^2 = \lambda_i \sigma^2, i = 1, 2, \dots, t$ .) Este modelo generaliza el modelo Bayesiano de regresión (en el caso mínimo informativo) y los modelos anidados y de efectos cruzados. Datta y Ghosh(91) estudia el modelo anterior con cociente entre varianzas conocido obteniendo la distribución predictiva exacta y el modelo con cociente entre varianzas desconocido obteniendo las distribuciones condicionales para aplicar el método Gibbs Sampler.

Al describir los resultados hemos hecho referencia a los estimadores Bayes empíricos, por lo que haremos seguidamente una descripción general de los mismos. En las aproximaciones Bayes empíricas se distinguen entre dos tipos de parámetros, unos parámetros Bayes sobre los que se asigna una distribución inicial y unos parámetros desconocidos sobre los que no se asigna distribución inicial. Entonces, el estimador para el modelo se construye obteniendo el estimador Bayes condicionado sobre los parámetros sobre los que no se asigna distribución inicial (trantándolos como cantidades conocidas) y sustituyendo dichos parámetros por unos estimadores clásicos previamente obtenidos de los mismos. La aproximación Bayes empírica evita de este modo la aplicación de procedimientos de integración numérica y al apoyarse en el estimador Bayes mantiene propiedades semejantes a éste siendo el error estimado para los estimadores Bayes empíricos (con un número no muy alto de parámetros desconocidos estimados por procedimientos clásicos) de un orden similar al del estimador Bayes. Una discursión más documentada sobre el comportamiento del estimador Bayes empírico frente al estimador Bayes podemos encontrarla en Ghosh(92)[40].

Los modelos y resultados aquí descritos muestran que la solución Bayes (estimador Bayes) a la estimación de un función  $g(\mathbf{Y})$  del vector de características de la población (y en particular la media poblacional) va a requerir del empleo de técnicas de integración numérica, cuando los cocientes entre varianzas sean desconocidos.

## Comentarios

A lo largo de esta subsección hemos visto el núcleo fundamental de la estimación en poblaciones finitas desde un punto de vista Bayesiano y cómo se pueden recuperar los estimadores predictivos y de población fija que hemos visto en las secciones anteriores. El gran número de modelos estudiados desde este punto de vista para modelos

conjugados nos ha conducido a introducir unos modelos algo más generales a los habitualmente trabajados a fin de poder describir a través de ellos las fronteras entre lo analizado y lo que falta por analizar (cuyo estudio será objeto de los capítulos 2 y 3). Otra cuestión que no está resuelta en esta línea de trabajo (de la que también nos ocuparemos en los capítulos 2 y 3) es la modelización de las poblaciones finitas cuando las varianzas del modelo son desconocidas y el cociente entre las mismas es también desconocido.

### 1.4.3 Estimadores Bayes constreñidos

Otra línea de estimación en poblaciones finitas desde el punto de vista Bayesiano es la basada en los estimadores Bayes constreñidos. Los estimadores Bayes constreñidos (constrained Bayes estimation) se aplican en estructuras de población en dos etapas en la estimación de  $g_i(\mathbf{Y}_i)$ ,  $i = 1, 2, \dots, K$  con  $g_i(\mathbf{Y}_i)$  la función a estimar en la unidad  $i$ -ésima de la población  $i = 1, 2, \dots, K$ .

Dada una muestra de la población  $(s, \mathbf{y}_s)$ , con  $k = K$ , el estimador Bayes constreñido de  $(g_1(\mathbf{Y}_1), g_2(\mathbf{Y}_2), \dots, g_K(\mathbf{Y}_K))$  se define como aquel estimador  $(g_1^{\text{BC}}(\mathbf{y}_{s_1}), g_2^{\text{BC}}(\mathbf{y}_{s_2}), \dots, g_K^{\text{BC}}(\mathbf{y}_{s_K}))$  que minimiza

$$E\left[\sum_{i=1}^K (g_i(\mathbf{Y}_i) - g_i(\widehat{\mathbf{Y}}_{s_i}))^2 \mid (s, \mathbf{y}_s)\right],$$

entre los estimadores  $(g_1(\widehat{\mathbf{y}}_{s_1}), g_2(\widehat{\mathbf{y}}_{s_2}), \dots, g_K(\widehat{\mathbf{y}}_{s_K}))$  de  $(g_1(\mathbf{Y}_1), g_2(\mathbf{Y}_2), \dots, g_K(\mathbf{Y}_K))$  tales que:

$$E\left[\frac{1}{K} \sum_{i=1}^K g_i(\mathbf{Y}_i) \mid (s, \mathbf{y}_s)\right] = \frac{1}{K} \sum_{i=1}^K g_i(\widehat{\mathbf{y}}_{s_i}), \quad (1.21)$$

$$E\left[\sum_{i=1}^K \left(g_i(\mathbf{Y}_i) - \frac{1}{K} \sum_{i=1}^K g_i(\mathbf{Y}_i)\right)^2 \mid (s, \mathbf{y}_s)\right] = \sum_{i=1}^K \left(g_i(\widehat{\mathbf{y}}_{s_i}) - \frac{1}{K} \sum_{i=1}^K g_i(\widehat{\mathbf{y}}_{s_i})\right)^2. \quad (1.22)$$

Siguiendo la prueba de Louis(84)[71], el estimador Bayes  $(g_1^{\text{B}}(\widehat{\mathbf{y}}_{s_1}), g_2^{\text{B}}(\widehat{\mathbf{y}}_{s_2}), \dots, g_K^{\text{B}}(\widehat{\mathbf{y}}_{s_K}))$  definido por  $g_i^{\text{B}}(\widehat{\mathbf{y}}_{s_i}) = E(g_i(\mathbf{Y}_i) \mid (s, \mathbf{y}_s))$ ,  $i = 1, 2, \dots, K$ , satisface la condición (1.21) pero no la (1.22).

El concepto de estimador Bayes constreñido fue introducido por Louis(84)[71] y aplicado por Lahiri(90)[67] a las poblaciones finitas para encontrar un estimador óptimo según este criterio para las medias poblacionales de cada unidad.

#### 1.4.4 Modelos “estrictamente” Bayesianos

La hipótesis de intercambiabilidad finita es una hipótesis “natural”<sup>6</sup> y por lo tanto usual dentro de la aproximación por modelos de superpoblación al problema de estimación en poblaciones finitas, la hemos estudiado en algunos modelos de superpoblación y a lo largo de esta sección la hemos visto como parte de la definición de algunos modelos y como propiedad para una serie de casos particulares de los modelos estudiados (siempre que asumimos a las variables  $Y_i$  de la población independientes e igualmente distribuidas dado un parámetro o vector de parámetros sobre el que asumimos una distribución de probabilidad). Sin embargo, tal y como ya señalábamos en la subsección de comentarios de la sección 1.3., la intercambiabilidad finita no implica la existencia de una distribución de probabilidad intercambiable sobre una población infinita de modo que nuestra población finita intercambiable sea realización de esa población infinita intercambiable. Iglesias(93)[63] y Iglesias, Pereira y Wechsler(94)[64] realizan un esfuerzo para la construcción de un modelo de superpoblación que sea intercambiable de manera finita y sea a la vez realización de un modelo de superpoblación intercambiable sobre una población infinita. Fruto de este esfuerzo para una característica continua tenemos el siguiente modelo “estrictamente” Bayesiano (Strict (predictivistic) Bayesian Model) y el siguiente resultado adaptado a la estimación de la media poblacional.

El modelo “estrictamente” Bayesiano para la estimación de la media poblacional se define como aquellos modelos de superpoblación  $\xi$  tales que el vector  $\mathbf{Y} = (Y_1, Y_2, \dots, Y_N)^t$  toma valores en  $\mathfrak{R}_N^+$  (el conjunto de los  $N$ -vectores reales positivos) y:

$$\begin{aligned} \mathbf{Y} | \bar{Y} = \bar{y} &\sim \text{Uniforme}(\chi_{\bar{y}}^N), \\ \bar{Y} &\sim \text{Pareto}(a, m_0), \end{aligned}$$

donde  $\chi_{\bar{y}}^N = \{(z_1, z_2, \dots, z_N)^t \in \mathfrak{R}_N^+ : \sum_{i=1}^N z_i/N = \bar{y}\}$  y  $a$  y  $m_0$  escalares positivos conocidos. Y el estimador Bayes para la media poblacional y su varianza condicionada a la muestra vienen dados (si  $\bar{y}_s \geq m_0$ ) por:

$$\begin{aligned} \hat{\bar{Y}}_{\text{BI}} &= \frac{n}{N} \frac{N+a-1}{n+a-1} \bar{y}_s, \\ \text{Var}(\bar{Y} | (s, \mathbf{y}_s)) &= \frac{n^2}{N^2} \frac{N+a-1}{(n+a-1)^2} \frac{N-n}{n+a-2} \bar{y}_s^2, \end{aligned} \quad (1.23)$$

estimador que coincide para  $a = 1$  con la media muestral (1.1).

#### 1.4.5 Comentarios de la sección

La aproximación mediante modelos de superpoblación Bayesianos al problema de estimación en poblaciones finitas permite una análisis completo e integrado del problema,

---

<sup>6</sup>no hay generalmente razones para distinguir en un estudio a los elementos de la población por sus índices que suelen ser simplemente procedimiento de notación

en el cual se recuperan los resultados de superpoblación clásicos que veíamos en la sección 1.3. ( que pueden interpretarse como resultados Bayesianos cuando no introducimos información sobre la distribución inicial del parámetro) y de población fija que veíamos en la sección 1.2. No obstante, la aproximación Bayesiana se enfrenta con dos problemas:

1.- La determinación de la distribución inicial sobre los parámetros. La aproximación Bayesiana se apoya en la asignación de una distribución inicial sobre el parámetro y esta asignación, que permite añadir a un modelo de superpoblación clásico nuestras creencias y conocimientos acerca de la “estructura probabilística” de la población (el punto a favor tal vez más importante de la aproximación Bayesiana), ha de ser “correcta” ya que el análisis Bayesiano puede ser bastante sensible a la distribución inicial asignada (ver el ejemplo de Bolfarine y Zacks(92)[10])[pag. 91].

2.- La dificultad para obtener la distribución predictiva del vector poblacional no observado, para asegurar la existencia de los primeros dos momentos de  $g(\mathbf{Y})$  según esta distribución (que pueden no existir) y en su caso calcular el estimador Bayes de  $g(\mathbf{Y})$  y su varianza condicionada a la muestra en los modelos no conjugados. Este problema está en relación directa con el primero, si nos centramos en el estudio de una característica continua de la población (tal y como estamos realizando en esta tesis), la hipótesis de normalidad para  $\mathbf{Y}$  es en general asumible y el problema de modelización se sitúa en la asignación de una distribución inicial sobre los parámetros media y varianza, si las distribuciones asignadas corresponden a los modelos Bayesianos normales o normales-gamma que hemos visto en la sección, el cálculo de la distribución exacta del vector  $\mathbf{Y}_u|(s, y_s)$  y de los estimadores Bayes y sus errores (en estos modelos es sencillo garantizar la existencia de los momentos necesarios para su obtención) es relativamente fácil, pero si no corresponden a estos modelos, el cálculo de la distribución de  $\mathbf{Y}_u|(s, y_s)$  y de sus primeros momentos (en el caso de que existan), no es fácil y requiere de la aplicación de métodos de integración numérica.

La primera cuestión se resuelve asignando distribuciones no informativas cuando no tenemos conocimiento “claro” acerca de la distribución del parámetro en cuestión o no queremos introducir información sobre los parámetros y la segunda mediante la aplicación de métodos computacionales basados en la teoría sobre cadenas de Markov como por ejemplo la aplicación del Algoritmo de Gibbs Sampling (Gelfand y Smith(90)[35], Gelfand et al(90)[34] y Casella y George(91)[15]), que obtiene una muestra de la distribución predictiva y puede implementarse fácilmente en el ordenador para un amplio grupo de modelos mediante el programa BUGS, otras soluciones alternativas cuando el número de parámetros es pequeño son las aproximaciones asintóticas a la distribución predictiva y algoritmos EM (estimación-maximización), métodos de integración numérica (metodos de Laplace y Naylor-Smith), métodos de Monte Carlo y los estimadores Bayes empíricos.

Destaquemos como referencias fundamentales para el estudio de una población finita desde el punto de vista de superpoblación Bayesiano (todas ellas ya referenciadas) por presentar un planteamiento global del tema, las siguientes: los artículos de Ericson(69,

88)[26,30] y Font(95)[33], el libro de Bolfarine y Zacks(92)[10] y la tesis de Murgui(82)[76] [capítulo 3, 4 y 5].

## 1.5 Comentarios finales

En este capítulo hemos intentado dar una visión amplia y crítica sobre distintas aproximaciones (población fija y modelos de superpoblación clásicos y Bayesianos) al problema de estimación de una característica continua de una población finita dada. En aras a esa visión amplia han quedado sin tratar algunas cuestiones específicas como: la estimación de otras funciones de características de la población así como: parámetros de la misma (en la aproximación por modelos de superpoblación clásicos y Bayesianos), la estimación en poblaciones finitas para características de la población discretas, las estimación en presencia de no respuesta, modelos dinámicos,...,etc.

Hemos visto también, cómo la aproximación basada en modelos de superpoblación Bayesianos supone un mecanismo de estudio eficaz<sup>7</sup> para el problema de estimación en poblaciones finitas que supera los inconvenientes de las otras aproximaciones y permite una interpretación de las mismas (recuperando para distribuciones iniciales específicas los estimadores basados en las aproximaciones de población fija y modelos de superpoblación clásicos).

Nuestra tesis se centra en la línea de trabajo Bayesiana y tiene por objetivos: el presentar una gama de modelos suficientemente amplia para poder inferir a través de ellos sobre una característica continua de una población y a la vez estudiar como estos modelos y generalizaciones de los mismos permiten plantear y resolver cuestiones relacionadas con el procedimiento de selección de la muestra aplicado. El primer objetivo, tiene su punto de partida en los estudios realizados por otros autores en esta área, referenciados en la sección 1.4, y será tratado en los capítulos 2 y 3. El segundo objetivo, se apoyará en los modelos establecidos en estos tres primeros capítulos y se desarrollará en los capítulos 4 y 5, comparando en el primero de ellos, la calidad de las inferencias obtenidas a través de muestreos aleatorios, estratificados y cluster, y en el segundo introduciendo algunos modelos sencillos para estudiar los resultados obtenidos a partir de un muestreo por rutas.

La notación aplicada para la realización de todo el estudio se apoyará en la notación que hemos introducido en este capítulo, reunida para facilitar su rápida consulta en el apéndice A, y en las notaciones que excepcionalmente se introduzcan al principio del capítulo. Por otra parte, el estudio de los modelos, requerirá de la aplicación de algunos resultados sobre funciones, distribuciones, . . . , cuyo enunciado en forma de lemas y su demostración se estudiará en el apéndice B, para evitar así su reiterada enumeración y facilitar la comprensión de los resultados de poblaciones finitas estudiados en cada capítulo. También tendremos que hacer uso en el estudio de los modelos no conjugados de métodos numéricos, en concreto, en esta tesis se aplicarán las técnicas basadas en

---

<sup>7</sup>los problemas de cálculo de distribuciones y estimadores Bayes para modelos no conjugados pueden ser resueltos en la actualidad mediante técnicas de aproximación de Monte Carlo y simulación Gibbs Sampler fácilmente implementables en ordenadores cada vez más rápidos y potentes

el muestreo por Gibbs, que en los modelos estudiados podrán ser implementadas en el ordenador mediante el programa BUGS. El apéndice C recogerá una descripción del muestreo por Gibbs Sampling (sección C.1), del programa BUGS (sección C.2), y un análisis general sobre cómo se han aplicado estas técnicas en la implementación de los modelos no conjugados estudiados en esta tesis (sección C.3).

## Capítulo 2

# Modelos de Superpoblación Bayesianos I: Modelos en una jerarquía

### 2.1 Introducción

Tal y como avanzábamos en la última sección del capítulo anterior, este capítulo y el próximo se centrarán en un objetivo de modelización. En ellos, se definirán de una forma más formal los modelos descritos en la subsección 1.4.2, y algunas generalizaciones de los mismos destinadas a modelizar la situación de poblaciones con varianzas desconocidas y cociente entre las mismas asimismo desconocido. La inferencia sobre las poblaciones, a través de estos modelos, se apoyará (como es natural en la aproximación Bayes) en la obtención de las distribuciones predictivas del vector poblacional no observado dada la muestra  $((s, \mathbf{y}_s)$  con  $s$  un procedimiento de tamaño fijo  $n$  no informativo), la obtención de la distribución predictiva de la media poblacional no observada, y en el caso de poblaciones estructuradas en dos etapas, la media no observada de cada unidad y a partir de las mismas del estimador Bayes (para función de pérdida cuadrática) de la media poblacional y en su caso media de cada unidad, y su varianza condicionada a la muestra<sup>1</sup> (siempre que existan los dos primeros momentos de las distribuciones predictivas anteriores). El análisis realizado en estos capítulos es siempre a posteriori de la elección de la muestra y no entra en consideraciones sobre el mecanismo de muestreo aplicado para su obtención salvo la consideración de que los

---

<sup>1</sup>La inferencia realizada se apoya en los comentarios realizados en la introducción de la sección 1.4. El estimador Bayes de  $g(\mathbf{Y})$  y su varianza son respectivamente la media y varianza de  $g(\mathbf{Y})$  según la distribución  $\mathbf{Y}_u|(s, \mathbf{y}_s)$  y pueden interpretarse como el estimador que minimiza el error cuadrático medio dada la muestra (o si se prefiere dado el modelo de superpoblación) y como su error cuadrático medio asociado dada la muestra (y si tomamos esperanzas respecto a la distribución del vector poblacional observado, respecto al modelo de superpoblación)

muestreos ha de ser de tamaño  $n$  fijos, no informativos y los índices de los elementos muestreados han de ser distintos, y en el caso de poblaciones estructuradas en dos etapas que el muestreo aplicado sea en dos etapas con tamaños  $k$  y  $n_i$ ,  $i \in \{i_1, i_2, \dots, i_k\}$  fijos no informativos y que las unidades seleccionadas e índices de los elementos sean distintos. La notación aplicada para realizar todo el estudio se encuentra reunida en el apéndice A.

Volviendo a los objetivos de modelización de este capítulo y el capítulo 3, el trabajo realizado se ha apoyado en modelos jerárquicos Bayes y en la descripción de las relaciones entre el vector de características poblacional y las características auxiliares a través de un modelo de regresión. Este modelo de regresión se ha introducido en el primer nivel del modelo, atendiendo a la idea de que estas características auxiliares dan información acerca de la característica de la población en estudio directamente. Esta idea, aplicada en sus modelos por: Ericson(69)[26], Murgui(82)[76] y Bolfarine y Zacks(92)[10], entre otros, no es universal y convive en el análisis de poblaciones estructuradas en dos etapas con la introducción del modelo de regresión en el segundo nivel del modelo, asumiendo que las características auxiliares proporcionan información acerca de las medias de cada unidad (ver por ejemplo, dentro de la aproximación Bayes y en el análisis de una característica continua, Datta y Lahiri(94)[23]).

Obviamente, las dos posturas son igualmente correctas, y la elección entre una de ellas o la adopción de ambas de una manera conjunta, por ejemplo Malec y Sedransk(85)[72] en su modelo  $m$ -jerárquico normal, dependerá de la población en estudio y las características auxiliares conocidas. En esta tesis nos ha parecido más adecuado decantarnos por la primera opción, pero queremos señalar que el mecanismo de análisis aplicado para la obtención de la distribución predictiva del vector poblacional no observado, de la media no observada y la estimación de la media poblacional y su varianza condicionada, es aplicable de manera directa para obtener estos resultados cuando el modelo de regresión se introduce en el segundo nivel o cuando se introduce en el primer y segundo nivel del modelo.

La división de los contenidos de modelización de esta tesis en dos capítulos obedece al número de modelos analizados y se basa en la división que proporcionan los modelos jerárquicos, así pues, el capítulo 2 se dedica a los modelos Bayesianos sobre poblaciones finitas basados en modelos jerárquicos con una jerarquía y el capítulo 3 a los modelos Bayesianos sobre poblaciones finitas en dos o más jerarquías. Los contenidos del capítulo 2 se organizan en cuatro secciones de las cuales la sección 2.1 es la presente introducción. La sección 2.2 analiza los modelos Bayesianos en una etapa, que se aplican en poblaciones no estructuradas (o monoetápicas) y contiene el estudio de tres modelos, correspondientes a la situación de varianza conocida, varianza desconocida y cociente entre varianzas conocido y varianza desconocida y cociente entre varianzas desconocido, e incluye el análisis de dos modelos mínimos informativos para la situación varianza conocida y desconocida. La sección 2.3 analiza los modelos Bayesianos en dos etapas con una jerarquía, para poblaciones estructuradas en dos etapas, que se aplican habitualmente en poblaciones estructuradas en dos etapas como resultado de un muestreo estratificado, que permite dividir la población en unidades bastante heterogeneas entre sí (y homogéneas en sí), sobre las que poseemos un nivel

de información suficiente para poder asignar los parámetros de las distribuciones iniciales. Los resultados obtenidos se apoyan en la hipótesis de un muestreo de todas las unidades ( $k = K$ ); la división de la sección es similar a la de la sección 2.2. Para finalizar el estudio de los modelos de superpoblación en una jerarquía, en la sección 2.4 se presentará un ejemplo de aplicación de los modelos estudiados y en la sección 2.5 se realizará un comentario breve sobre el capítulo y en especial sobre las motivaciones que conducen al análisis de modelos con más jerarquías.

## 2.2 Modelos Bayesianos en una etapa

Esta sección se dedicará al análisis de modelos Bayesianos adecuados para describir una característica continua de una población no estructurada (monoetápica), en el sentido descrito en la sección 1.1. Se estudiarán en concreto cinco modelos que hemos clasificado en cuatro subsecciones atendiendo a las hipótesis introducidas para la modelización: el modelo Bayes de Regresión Normal (varianzas conocidas), el modelo Bayes de Regresión Normal-Gamma (varianzas desconocidas y cociente entre varianzas conocido), el modelo Bayes de Regresión para cociente entre varianzas desconocido (varianza desconocida y cociente entre varianzas desconocida) y los modelos Bayes de Regresión mínimo informativos (aplicables cuando no disponemos o no queremos introducir información sobre los parámetros iniciales), que analiza los casos de varianza conocida y desconocida.

Los modelos Bayes en una etapa han sido extensamente estudiados en la bibliografía de poblaciones finitas, tal y como vimos en la subsección 1.4.2. En concreto, los modelos Bayes de Regresión Normal y Normal-Gamma, (el primero de ellos estudiado en Bolfarine y Zacks(92)[10]) y los modelos mínimo informativos son generalizaciones de los modelos Bayesianos tradicionalmente aplicados en Poblaciones Finitas por Ericson(69)[26], Murgui(82)[76],... y proporcionan como casos particulares estimadores de población fija (estimador media muestral, estimador de Horwitz-Thompson,...) y estimadores clásicos (estimador media muestral, estimador de regresión clásico,...). El objetivo de esta sección es el de completar y analizar los resultados obtenidos por otros autores en los casos varianza conocida y varianza desconocida y cociente entre varianzas conocido e introducir un modelo para el caso más general de cociente entre varianzas desconocido. La notación aplicada en esta sección corresponde a la introducida para poblaciones estructuradas en una etapa a lo largo del capítulo 1, que para facilitar su consulta rápida ha sido reunida en la sección estructura monoetápica del Apéndice A. Los resultados inferenciales que se obtendrán en esta sección: distribuciones predictivas dada la muestra y estimadores Bayes y varianza condicionada a la muestra se apoyarán en una muestra obtenida mediante un procedimiento de muestreo dado de tamaño fijo y no informativo, en el que los índices de los elementos seleccionados sean distintos entre sí.

Las demostraciones de los teoremas se apoyan en algunos resultados previos formulados y demostrados en forma de lemas en el apéndice B.

## 2.2.1 El modelo Bayes de Regresión Normal

Dedicaremos esta subsección al análisis del modelo Bayes de Regresión Normal, que se aplica para describir una característica continua en poblaciones finitas no estructuradas (monoetápicas) cuando la varianza de la población es conocida. Su definición, a efectos de su estudio en esta tesis, es la siguiente:

**Definición 2.1** Dado el vector poblacional  $\mathbf{Y} = (Y_1, Y_2, \dots, Y_N)^t$ , para el cual asociada a cada variable  $Y_i$  tenemos una información auxiliar representada por un vector  $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})$  conocido llamaremos modelo Bayes de Regresión Normal al definido por:

$$\begin{aligned} \mathbf{Y}|\boldsymbol{\beta} &\sim N_N(\mathbf{X}\boldsymbol{\beta}, \mathbf{V}), \\ \boldsymbol{\beta} &\sim N_p(\mathbf{b}, \mathbf{B}), \end{aligned} \quad (2.1)$$

donde  $\mathbf{X} = (\mathbf{x}_1^t, \mathbf{x}_2^t, \dots, \mathbf{x}_N^t)^t$  es una matriz de rango  $p$ ,  $\mathbf{b}$  un  $p$ -vector conocido y  $\mathbf{V}$  y  $\mathbf{B}$  son matrices simétricas, definidas positivas conocidas de dimensiones  $N \times N$  y  $p \times p$  respectivamente.

Este modelo tiene como caso particular para  $p = 1$ ,  $\mathbf{X} = \mathbf{1}_N$ , el modelo Bayes Básico Normal descrito en la subsección 1.4.3 y generaliza los modelos Bayesianos tradicionalmente aplicados.

El siguiente teorema proporciona las distribuciones predictivas del vector poblacional no observado y la media no observada, así como el estimador Bayes para la media poblacional y su varianza condicionada a la muestra.

**Teorema 2.1** Consideremos el modelo Bayes de Regresión Normal dado en (2.1), entonces se tiene que:

(a)  $\mathbf{Y}_u|(s, \mathbf{y}_s)$  se distribuye normal  $(N - n)$ -variante con media y varianza dadas respectivamente por:

$$\begin{aligned} E(\mathbf{Y}_u|(s, \mathbf{y}_s)) &= \mathbf{X}_u \hat{\boldsymbol{\beta}}_B + \mathbf{V}_{su}^t \mathbf{V}_s^{-1} (\mathbf{y}_s - \mathbf{X}_s \hat{\boldsymbol{\beta}}_B), \\ \text{Var}(\mathbf{Y}_u|(s, \mathbf{y}_s)) &= \mathbf{V}_u - \mathbf{V}_{su}^t \mathbf{V}_s^{-1} \mathbf{V}_{su} + \\ &\quad + (\mathbf{X}_u - \mathbf{V}_{su}^t \mathbf{V}_s^{-1} \mathbf{X}_s) \boldsymbol{\alpha} (\mathbf{X}_s^t \mathbf{V}_s^{-1} \mathbf{X}_s)^{-1} (\mathbf{X}_u - \mathbf{V}_{su}^t \mathbf{V}_s^{-1} \mathbf{X}_s)^t, \end{aligned} \quad (2.2)$$

(b)  $\bar{Y}_u|(s, \mathbf{y}_s)$  se distribuye normal con media y varianza dadas por:

$$\begin{aligned} E(\bar{Y}_u|(s, \mathbf{y}_s)) &= \frac{1}{N - n} \mathbf{1}_{N-n}^t [\mathbf{X}_u \hat{\boldsymbol{\beta}}_B + \mathbf{V}_{su}^t \mathbf{V}_s^{-1} (\mathbf{y}_s - \mathbf{X}_s \hat{\boldsymbol{\beta}}_B)], \\ \text{Var}(\bar{Y}_u|(s, \mathbf{y}_s)) &= \frac{1}{(N - n)^2} \{ \mathbf{1}_{N-n}^t (\mathbf{V}_u - \mathbf{V}_{su}^t \mathbf{V}_s^{-1} \mathbf{V}_{su}) \mathbf{1}_{N-n} + \\ &\quad + \mathbf{1}_{N-n}^t (\mathbf{X}_u - \mathbf{V}_{su}^t \mathbf{V}_s^{-1} \mathbf{X}_s) \boldsymbol{\alpha} (\mathbf{X}_s^t \mathbf{V}_s^{-1} \mathbf{X}_s)^{-1} (\mathbf{X}_u - \mathbf{V}_{su}^t \mathbf{V}_s^{-1} \mathbf{X}_s)^t \mathbf{1}_{N-n} \}, \end{aligned} \quad (2.3)$$

y (c) el estimador Bayes para la media poblacional y su varianza condicionada a la muestra vienen dados por:

$$\begin{aligned}\widehat{Y} &= \frac{n}{N}\bar{y}_s + \frac{1}{N}\mathbf{1}_{N-n}^t[\mathbf{X}_u\widehat{\beta}_B + \mathbf{V}_{su}^t\mathbf{V}_s^{-1}(y_s - \mathbf{X}_s\widehat{\beta}_B)], \\ \text{Var}(\widehat{Y}|(s, y_s)) &= \frac{1}{N^2} \{ \mathbf{1}_{N-n}^t(\mathbf{V}_u - \mathbf{V}_{su}^t\mathbf{V}_s^{-1}\mathbf{V}_{su})\mathbf{1}_{N-n} + \\ &\quad + \mathbf{1}_{N-n}^t(\mathbf{X}_u - \mathbf{V}_{su}^t\mathbf{V}_s^{-1}\mathbf{X}_s)\alpha(\mathbf{X}_s^t\mathbf{V}_s^{-1}\mathbf{X}_s)^{-1}(\mathbf{X}_u - \mathbf{V}_{su}^t\mathbf{V}_s^{-1}\mathbf{X}_s)^t\mathbf{1}_{N-n} \},\end{aligned}\quad (2.4)$$

donde  $\mathbf{V}_s = \text{Var}_{\mathbf{Y}|\beta}(\mathbf{Y}_s)$ ,  $\mathbf{V}_u = \text{Var}_{\mathbf{Y}|\beta}(\mathbf{Y}_u)$ ,  $\mathbf{V}_{su} = \text{Cov}_{\mathbf{Y}|\beta}(\mathbf{Y}_s, \mathbf{Y}_u)$ ,  $\bar{y}_s$  es la media muestral,  $\widehat{\beta} = (\mathbf{X}_s^t\mathbf{V}_s^{-1}\mathbf{X}_s)^{-1}\mathbf{X}_s^t\mathbf{V}_s^{-1}\mathbf{y}_s$  es el estimador de regresión clásico y,

$$\begin{aligned}\alpha &= (\mathbf{X}_s^t\mathbf{V}_s^{-1}\mathbf{X}_s + \mathbf{B}^{-1})^{-1}\mathbf{X}_s^t\mathbf{V}_s^{-1}\mathbf{X}_s, \\ \widehat{\beta}_B &= E(\beta|(s, y_s)) = \alpha\widehat{\beta} + (\mathbf{I}_p - \alpha)\mathbf{b},\end{aligned}\quad (2.5)$$

**Demostración.** La demostración de este teorema es inmediata a partir de los resultados del teorema 3.5 (capítulo 3), tomando:  $m = 1$ ,  $\mathbf{Y} = (Y_1, Y_2, \dots, Y_N)^t$ ,  $\theta_1 = \beta$ ,  $\mathbf{A}_1 = \mathbf{X}$ ,  $\mathbf{A}_2 = \mathbf{I}_p$ ,  $\mathbf{V}_1 = \mathbf{V}$  y  $\mathbf{V}_2 = \mathbf{B}$  y denotando  $\widehat{\theta}_1^{\mathbf{B}} = \widehat{\beta}_B$ .

C.S.Q.D.

Una demostración alternativa de estos resultados puede encontrarse en Bolfarine et al(87)[8] y Bolfarine y Zacks(91,92)[9,10].

Notemos a partir de las expresiones (2.2), (2.3) y (2.4), que la contribución de la muestra a las distribuciones predictivas y a la estimación de la media poblacional se articula a través de  $\mathbf{y}_s$ ,  $\bar{y}_s$  (la media muestral),  $\widehat{\beta}$  (el estimador clásico de regresión) y  $\alpha$  (e indirectamente a través de  $\mathbf{V}_s$ ,  $\mathbf{V}_{su}$  y  $\mathbf{X}_s$ ). El factor  $\alpha$ , es un factor de encogimiento (shrinking factor), dado que por definición,  $\alpha = (\mathbf{I}_p - \alpha)\mathbf{B}^{-1}\mathbf{X}_s^t\mathbf{V}_s^{-1}\mathbf{X}_s$ , que mide la proporción de varianza respecto a  $\beta$  no explicada por la muestra, ya que:

$$\alpha = [\text{Var}(\beta) - \text{Var}(\beta|(s, y_s))]\text{Var}^{-1}(\beta).\quad (2.6)$$

La expresión (2.6) se obtiene a partir de la definición de  $\alpha$ , notando que  $\text{Var}(\beta|(s, y_s)) = (\mathbf{X}_s^t\mathbf{V}_s^{-1}\mathbf{X}_s + \mathbf{B}^{-1})^{-1}$  (resultado previo en la demostración anterior).

A partir de esta reflexión sobre  $\alpha$  podemos interpretar el comportamiento de las distribuciones predictivas dadas en el teorema 2.1. En primer lugar notemos que el estimador Bayes de la pendiente de regresión,  $\widehat{\beta}_B$ , es una “combinación lineal” del estimador de regresión clásico y la pendiente a priori de la población de modo que si  $\alpha \rightarrow \mathbf{I}_p$  (situación que se produce cuando la precisión de la distribución inicial tiende a cero), entonces el estimador Bayes de la pendiente de regresión tiende al estimador de regresión clásico y si  $\alpha \rightarrow \mathbf{0}_p$  (que se produce cuando la varianza de la distribución

inicial tiende a cero), entonces el estimador Bayes de la pendiente de regresión tiende a la pendiente a priori asignada, esto es, a medida que  $\alpha$  (la proporción de varianza no explicada por la muestra) disminuye, la pendiente a priori asignada va tomando mayor importancia en la estimación Bayes de la pendiente de regresión y por lo tanto en la estimación de la media poblacional. Esta interpretación es compatible con la que podemos extraer fijándonos en las varianzas de las distribuciones predictivas, ya que a medida que  $\alpha$  (la proporción de varianza no explicada por la muestra) disminuye, las varianzas de las distribuciones predictivas y la varianza del estimador de la media poblacional condicionada a la muestra disminuyen. (Estas dos observación señalan un comportamiento lógico del modelo ya que cuando disminuye  $\alpha$  lo que esta sucediendo es que estamos introduciendo una mayor información al modelo a través de una distribución inicial más precisa).

Del teorema anterior, se recuperan los resultados Bayesianos tradicionales, (ver apartado estimadores Bayes de la media poblacional de la subsección 1.4.2), tomando la forma adecuada de  $\mathbf{V}$  y de  $\mathbf{X}$ .

## 2.2.2 El modelo Bayes de Regresión Normal-Gamma

El modelo Bayes de Regresión Normal-Gamma que definiremos a continuación es una generalización de los modelos normales-gamma tradicionales que se aplican para modelizar una característica continua de un población no estructurada (monoetápica), cuando la varianza es desconocida y el cociente entre varianzas (en nuestro caso  $\mathbf{W}\mathbf{C}^{-1}$ ) conocido.

**Definición 2.2** *Dado el vector poblacional  $\mathbf{Y} = (Y_1, Y_2, \dots, Y_N)^t$ , para el cual asociada a cada variable  $Y_i$  tenemos una información auxiliar representada por un vector  $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})$  conocido llamaremos modelo Bayes de Regresión Normal-Gamma al definido por:*

$$\begin{aligned} \mathbf{Y}|\beta, \sigma^2 &\sim N_N(\mathbf{X}\beta, \sigma^2\mathbf{W}), \\ \beta|\sigma^2 &\sim N_p(\mathbf{b}, \sigma^2\mathbf{C}), \\ \sigma^2 &\sim Ga^{-1}(a_0, b_0), \end{aligned} \tag{2.7}$$

donde  $\mathbf{X} = (\mathbf{x}_1^t, \mathbf{x}_2^t, \dots, \mathbf{x}_N^t)^t$  es una matriz de rango  $p$ ,  $\mathbf{b}$  es un  $p$ -vector conocido,  $\mathbf{W}$  y  $\mathbf{C}$  son matrices simétricas y definidas positivas conocidas de dimensiones  $N \times N$  y  $p \times p$  respectivamente, y  $a_0$  y  $b_0$  son escalares positivos conocidos.

Este modelo, presenta como caso particular para  $p = 1$  y  $\mathbf{X} = \mathbf{1}_N$ , el modelo Bayes Básico Normal-Gamma y generaliza los modelos Bayesianos tradicionalmente aplicados para modelizar este tipo de poblaciones (ver subsección 1.4.2).

El siguiente teorema proporciona las distribuciones predictivas para el vector poblacional no observado, la media poblacional no observada y el estimador para la media

poblacional y su varianza condicionada a la muestra. Estos resultados pueden interpretarse de manera análoga a los correspondientes resultados para el modelo Bayes de Regresión Normal; y proporcionan los correspondientes resultados para los modelos tradicionales normal-gamma, a los que nos referíamos en el párrafo anterior mediante la adecuada descripción de  $\mathbf{W}$ ,  $\mathbf{C}$ ,  $a_0$  y  $b_0$  (ver apartado estimadores Bayes de la media poblacional de la subsección 1.4.2).

**Teorema 2.2** Consideremos el modelo Bayes de Regresión Normal-Gamma dado en (2.7), entonces se tiene que:

(a)  $\mathbf{Y}_u|(s, \mathbf{y}_s)$  se distribuye  $t$ ,  $(N - n)$ -variante con  $2a_1$  grados de libertad y media y varianza dadas, para  $a_1 > 1$ , respectivamente por:

$$\begin{aligned} E(\mathbf{Y}_u|(s, \mathbf{y}_s)) &= \mathbf{X}_u \hat{\beta}_B + \mathbf{W}_{su}^t \mathbf{W}_s^{-1} (\mathbf{y}_s - \mathbf{X}_s \hat{\beta}_B), \\ \text{Var}(\mathbf{Y}_u|(s, \mathbf{y}_s)) &= \frac{b_1}{a_1 - 1} \{ \mathbf{W}_u - \mathbf{W}_{su}^t \mathbf{W}_s^{-1} \mathbf{W}_{su} + \\ &\quad + (\mathbf{X}_u - \mathbf{W}_{su}^t \mathbf{W}_s^{-1} \mathbf{X}_s) \alpha (\mathbf{X}_s^t \mathbf{W}_s^{-1} \mathbf{X}_s)^{-1} (\mathbf{X}_u - \mathbf{W}_{su}^t \mathbf{W}_s^{-1} \mathbf{X}_s)^t \}, \end{aligned} \quad (2.8)$$

(b)  $\bar{\mathbf{Y}}_u|(s, \mathbf{y}_s)$  se distribuye  $t$  con  $2a_1$  grados de libertad y media y varianza dadas, para  $a_1 > 1$ , por:

$$\begin{aligned} E(\bar{\mathbf{Y}}_u|(s, \mathbf{y}_s)) &= \frac{1}{N - n} \mathbf{1}_{N-n}^t [\mathbf{X}_u \hat{\beta}_B + \mathbf{W}_{su}^t \mathbf{W}_s^{-1} (\mathbf{y}_s - \mathbf{X}_s \hat{\beta}_B)] \\ \text{Var}(\bar{\mathbf{Y}}_u|(s, \mathbf{y}_s)) &= \frac{1}{(N - n)^2} \frac{b_1}{a_1 - 1} \{ \mathbf{1}_{N-n}^t [\mathbf{W}_u - \mathbf{W}_{su}^t \mathbf{W}_s^{-1} \mathbf{W}_{su}] \mathbf{1}_{N-n} + \\ &\quad + \mathbf{1}_{N-n}^t (\mathbf{X}_u - \mathbf{W}_{su}^t \mathbf{W}_s^{-1} \mathbf{X}_s) \alpha (\mathbf{X}_s^t \mathbf{W}_s^{-1} \mathbf{X}_s)^{-1} (\mathbf{X}_u - \mathbf{W}_{su}^t \mathbf{W}_s^{-1} \mathbf{X}_s)^t \mathbf{1}_{N-n} \}, \end{aligned} \quad (2.9)$$

y (c) el estimador Bayes para la media poblacional y su error cuadrático medio vienen dados, para  $a_1 > 1$ , por:

$$\begin{aligned} \hat{\bar{\mathbf{Y}}} &= \frac{n}{N} \bar{\mathbf{y}}_s + \frac{1}{N} \mathbf{1}_{N-n}^t [\mathbf{X}_u \hat{\beta}_B + \mathbf{W}_{su}^t \mathbf{W}_s^{-1} (\mathbf{y}_s - \mathbf{X}_s \hat{\beta}_B)], \\ \text{Var}(\bar{\mathbf{Y}}|(s, \mathbf{y}_s)) &= \frac{1}{N^2} \frac{b_1}{a_1 - 1} \{ \mathbf{1}_{N-n}^t [\mathbf{W}_u - \mathbf{W}_{su}^t \mathbf{W}_s^{-1} \mathbf{W}_{su}] \mathbf{1}_{N-n} + \\ &\quad + \mathbf{1}_{N-n}^t (\mathbf{X}_u - \mathbf{W}_{su}^t \mathbf{W}_s^{-1} \mathbf{X}_s) \alpha (\mathbf{X}_s^t \mathbf{W}_s^{-1} \mathbf{X}_s)^{-1} (\mathbf{X}_u - \mathbf{W}_{su}^t \mathbf{W}_s^{-1} \mathbf{X}_s)^t \mathbf{1}_{N-n} \}, \end{aligned} \quad (2.10)$$

donde  $\mathbf{W}_s = \sigma^{-2} \text{Var}_{\mathbf{Y}|\beta}(\mathbf{Y}_s)$ ,  $\mathbf{W}_u = \sigma^{-2} \text{Var}_{\mathbf{Y}|\beta}(\mathbf{Y}_u)$ ,  $\mathbf{W}_{su} = \sigma^{-2} \text{Cov}_{\mathbf{Y}|\beta}(\mathbf{Y}_s, \mathbf{Y}_u)$ ,  $\bar{\mathbf{y}}_s$  es la media muestral,  $\hat{\beta} = (\mathbf{X}_s^t \mathbf{W}_s^{-1} \mathbf{X}_s)^{-1} \mathbf{X}_s^t \mathbf{W}_s^{-1} \mathbf{y}_s$  es el estimador de regresión clásico y,

$$\begin{aligned}
\alpha &= (\mathbf{X}_s^t \mathbf{W}_s^{-1} \mathbf{X}_s + \mathbf{C}^{-1})^{-1} \mathbf{X}_s^t \mathbf{W}_s^{-1} \mathbf{X}_s, \\
\hat{\beta}_B &= E(\beta | (s, \mathbf{y}_s)) = \alpha \hat{\beta} + (\mathbf{I}_p - \alpha) \mathbf{b}, \\
a_1 &= \frac{n}{2} + a_0 \\
b_1 &= b_0 + \frac{1}{2} (\mathbf{y}_s - \mathbf{X}_s \mathbf{b})^t (\mathbf{W}_s + \mathbf{X}_s \mathbf{C} \mathbf{X}_s^t)^{-1} (\mathbf{y}_s - \mathbf{X}_s \mathbf{b}).
\end{aligned} \tag{2.11}$$

**Demostración.** La demostración de este teorema es inmediata a partir de los resultados del teorema 3.6 (capítulo 3), tomando:  $m = 1$ ,  $\mathbf{Y} = (Y_1, Y_2, \dots, Y_N)^t$ ,  $\theta_1 = \beta$ ,  $\mathbf{A}_1 = \mathbf{X}$ ,  $\mathbf{A}_2 = \mathbf{I}_p$ ,  $\mathbf{W}_1 = \mathbf{W}$  y  $\mathbf{W}_2 = \mathbf{C}$  y denotando  $\hat{\theta}_1^B = \hat{\beta}_B$ .

C.S.Q.D.

### 2.2.3 El modelo Bayes de Regresión con cociente entre varianzas desconocido

El modelo Bayes de Regresión Normal-Gamma definido y estudiado en la subsección anterior, proporciona una solución con distribuciones predictivas cerradas y estimador Bayes de la media poblacional y varianza condicionada a la muestra fácilmente calculable mediante simples operaciones matriciales en la modelización de una característica continua en poblaciones finitas con varianza desconocida. Sin embargo, esta solución no es totalmente satisfactoria por las dificultades que entraña la elección de las matrices  $\mathbf{W}$  y  $\mathbf{C}$  de manera adecuada; y en la circunstancia de que las varianzas de la población y de la inicial han de ser proporcionales a un mismo escalar  $\sigma^2$ . Estas restricciones tienen su origen en las limitaciones que supone elaborar un modelo conjugado.

En la actualidad, la implementación en ordenadores de algoritmos como el Gibbs Sampling (ver C.1), que permiten la obtención de muestras de las distribuciones finales de un modelo cuando éste no es conjugado, hacen factible la elaboración y estudio de modelos no conjugados y de esta manera permiten una mayor libertad en la modelización. El modelo que definiremos a continuación es un modelo no conjugado que puede resultar adecuado en la modelización de una característica continua en poblaciones no estructuradas (monoetápicas), cuando la varianza es desconocida y el cociente entre las varianzas asimismo desconocido. Este modelo se inspira en los modelos de Ghosh y Meeden(86)[42], Pérez y Pericchi(91)[80] y fundamentalmente Datta y Ghosh(91)[22], que describíamos en la subsección 1.4.2.

**Definición 2.3** *Dado el vector poblacional  $\mathbf{Y} = (Y_1, Y_2, \dots, Y_N)^t$ , para el cual asociada a cada variable  $Y_i$  tenemos una información auxiliar representada por un vector  $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})$  conocido llamaremos modelo Bayes de Regresión con cociente entre varianzas desconocido al definido por:*

$$\begin{aligned}
\mathbf{Y}|\beta, \sigma^2 &\sim N_N(\mathbf{X}\beta, \sigma^2 \mathbf{I}_N), \\
\sigma^2 &\sim Ga^{-1}(a_0, b_0), \\
\beta|\sigma_\beta^2 &\sim N_p(\mathbf{b}, \sigma_\beta^2 \mathbf{C}), \\
\sigma_\beta^2 &\sim Ga^{-1}(a_\beta, b_\beta),
\end{aligned} \tag{2.12}$$

donde  $\mathbf{X} = (\mathbf{x}_1^t, \mathbf{x}_2^t, \dots, \mathbf{x}_N^t)^t$  es una matriz de rango  $p$ ,  $\mathbf{b}$  un  $p$ -vector conocido,  $\mathbf{C}$  es una matriz  $p \times p$  simétrica definida positiva conocida,  $\mathbf{I}_p$  es la matriz identidad de dimensión  $p$ , y  $a_0, a_\beta, b_0$  y  $b_\beta$  escalares positivos conocidos.

El siguiente teorema establece las distribuciones predictiva para el vector poblacional no observado y la media no observada condicionadas a  $\lambda = \frac{\sigma^2}{\sigma_\beta^2}$ , la distribución final de  $\lambda|(s, \mathbf{y}_s)$  y el estimador Bayes y varianza condicionada a la muestra en función de los operadores  $E^{\lambda|(s, \mathbf{y}_s)}[\cdot]$  y  $Var^{\lambda|(s, \mathbf{y}_s)}[\cdot]$ .

**Teorema 2.3** Consideremos el modelo Bayes de Regresión con cociente entre varianzas desconocido dado en (2.12); si definimos por  $\lambda = \frac{\sigma^2}{\sigma_\beta^2}$ , entonces se tiene que:

(a)  $\mathbf{Y}_u|(s, \mathbf{y}_s), \lambda$  se distribuye  $t$ ,  $(N - n)$ -variante con  $2a_1$  grados de libertad y media y varianza dadas, respectivamente, para  $a_1 > 1$ , por:

$$\begin{aligned}
E(\mathbf{Y}_u|(s, \mathbf{y}_s), \lambda) &= \mathbf{X}_u \hat{\beta}_B, \\
Var(\mathbf{Y}_u|(s, \mathbf{y}_s), \lambda) &= \frac{b_1}{a_1 - 1} \{ \mathbf{I}_{N-n} \lambda + \mathbf{X}_u (\mathbf{I}_p - \alpha) \mathbf{C} \mathbf{X}_u^t \}
\end{aligned} \tag{2.13}$$

(b)  $\bar{Y}_u|(s, \mathbf{y}_s), \lambda$  se distribuye  $t$  con  $2a_1$  grados de libertad y media y varianza, para  $a_1 > 1$ , dadas por:

$$\begin{aligned}
E(\bar{Y}_u|(s, \mathbf{y}_s), \lambda) &= \frac{1}{N - n} \mathbf{1}_{N-n}^t \mathbf{X}_u \hat{\beta}_B, \\
Var(\bar{Y}_u|(s, \mathbf{y}_s), \lambda) &= \frac{1}{(N - n)^2} \frac{b_1}{a_1 - 1} \{ (N - n) \lambda + \mathbf{1}_{N-n}^t \mathbf{X}_u (\mathbf{I}_p - \alpha) \mathbf{C} \mathbf{X}_u^t \mathbf{1}_{N-n} \}
\end{aligned} \tag{2.14}$$

(c)  $\lambda|(s, \mathbf{y}_s)$  se distribuye con función de densidad dada por:

$$\pi(\lambda|(s, \mathbf{y}_s)) \propto |\mathbf{I}_n \lambda + \mathbf{X}_s \mathbf{C} \mathbf{X}_s^t|^{-\frac{1}{2}} \lambda^{-(a_0+1)} b_1^{-a_1}. \tag{2.15}$$

y (d) el estimador Bayes para la media poblacional (para  $n \geq 2$ ) y su varianza condicionada por la muestra (para  $n \geq 2$  y  $a_0 > 1$ ) vienen dados por:

$$\begin{aligned}\widehat{Y} &= \frac{n}{N}\bar{y}_s + \frac{1}{N}\mathbf{1}_{N-n}^t \mathbf{X}_u E^{\lambda(s, \mathbf{y}_s)}[\widehat{\beta}_B], \\ \text{Var}(\bar{Y}|(s, \mathbf{y}_s)) &= \frac{1}{N^2} \left\{ \mathbf{1}_{N-n}^t \mathbf{X}_u \text{Var}^{\lambda(s, \mathbf{y}_s)}[\widehat{\beta}_B] \mathbf{X}_u^t \mathbf{1}_{N-n} + \right. \\ &\quad \left. + \frac{1}{a_1 - 1} E^{\lambda(s, \mathbf{y}_s)} \{ b_1 [(N-n)\lambda + \mathbf{1}_{N-n}^t \mathbf{X}_u (\mathbf{I}_p - \alpha) \mathbf{C} \mathbf{X}_u^t \mathbf{1}_{N-n}] \} \right\},\end{aligned}\quad (2.16)$$

donde  $E^{\lambda(s, \mathbf{y}_s)}[\cdot]$  y  $\text{Var}^{\lambda(s, \mathbf{y}_s)}[\cdot]$ , son respectivamente, los operadores esperanza y varianza respecto a la distribución de  $\lambda|(s, \mathbf{y}_s)$  dada en (2.15),  $\bar{y}_s$  es la media muestral,  $\widehat{\beta} = (\mathbf{X}_s^t \mathbf{X}_s)^{-1} \mathbf{X}_s^t \mathbf{y}_s$  es el estimador de regresión clásico y,

$$\begin{aligned}\alpha &= (\mathbf{X}_s^t \mathbf{X}_s + \mathbf{C}^{-1} \lambda)^{-1} \mathbf{X}_s^t \mathbf{X}_s, \\ \widehat{\beta}_B &= E(\beta|(s, \mathbf{y}_s), \lambda) = \alpha \widehat{\beta} + (\mathbf{I}_p - \alpha) \mathbf{b}, \\ a_1 &= \frac{n}{2} + a_0 + a_\beta \\ b_1 &= b_\beta + \frac{b_0}{\lambda} + \frac{1}{2} (\mathbf{y}_s - \mathbf{X}_s \mathbf{b})^t (\mathbf{I}_n \lambda + \mathbf{X}_s \mathbf{C} \mathbf{X}_s^t)^{-1} (\mathbf{y}_s - \mathbf{X}_s \mathbf{b}).\end{aligned}\quad (2.17)$$

**Demostración.** La demostración de este teorema se apoya en el teorema 2.1 y los lemas B.3, B.8 y B.9; para su mejor desarrollo la dividiremos en los siguientes pasos:

1.- A partir de los resultados (a) y (b) del teorema 2.1, tomando  $\mathbf{V} = \lambda \sigma_\beta^2 \mathbf{I}_N$  y  $\mathbf{B} = \sigma_\beta^2 \mathbf{C}$ , se obtienen las distribuciones de  $\mathbf{Y}_u|(s, \mathbf{y}_s), \sigma_\beta^2, \lambda$  y  $\bar{Y}_u|(s, \mathbf{y}_s), \sigma_\beta^2, \lambda$ , que son:

$$\begin{aligned}\mathbf{Y}_u|(s, \mathbf{y}_s), \sigma_\beta^2, \lambda &\sim N_{N-n}(\mathbf{X}_u \widehat{\beta}_B^\lambda, \sigma_\beta^2 (\mathbf{I}_{N-n} \lambda + \mathbf{X}_u (\mathbf{I}_p - \alpha) \mathbf{C} \mathbf{X}_u^t)), \\ \bar{Y}_u|(s, \mathbf{y}_s), \sigma_\beta^2, \lambda &\sim N\left(\frac{1}{N-n} \mathbf{1}_{N-n}^t \mathbf{X}_u \widehat{\beta}_B^\lambda, \frac{\sigma_\beta^2 (a_1 - 1)}{b_1} \text{Var}(\bar{Y}_u|(s, \mathbf{y}_s), \lambda)\right).\end{aligned}\quad (2.18)$$

$$(2.19)$$

2.- A partir de las distribuciones iniciales de  $\mathbf{Y}|\beta, \sigma^2$  y de  $\beta|\sigma_\beta^2$  se obtiene aplicando, sobre la primera las propiedades de las distribuciones marginales de una normal y después, sobre el resultado y la segunda, el apartado (a) del lema B.8, la distribución de  $\mathbf{Y}_s|\sigma_\beta^2, \lambda$ , que es:

$$\mathbf{Y}_s|\sigma_\beta^2, \lambda \sim N_n(\mathbf{X}_s \mathbf{b}, \sigma_\beta^2 [\mathbf{I}_n \lambda + \mathbf{X}_s \mathbf{C} \mathbf{X}_s^t]).\quad (2.20)$$

3.- A partir de las distribuciones iniciales de  $\sigma^2$  y  $\sigma_\beta^2$  se obtiene, aplicando el método de obtención de una f.d.p por el método directo (para una transformación 1 a 1) la distribución inicial conjunta de  $(\sigma_\beta^2, \lambda)$ , que es:

$$\pi(\sigma_{\beta}^2, \lambda) \propto [\sigma_{\beta}^2]^{-(a_0+a_{\beta}+1)} \exp\left\{-\frac{1}{\sigma_{\beta}^2}\left(b_{\beta} + \frac{b_0}{\lambda}\right)\right\} \lambda^{-(a_0+1)} \quad (2.21)$$

4.- A partir de (2.20) y (2.21), aplicando el teorema de Bayes e identificando los términos se obtienen las distribuciones de  $\sigma_{\beta}^2|\lambda, (s, \mathbf{y}_s)$ , que viene dada por:

$$\sigma_{\beta}^2|\lambda, (s, \mathbf{y}_s) \sim Ga^{-1}(a_1, b_1) \quad (2.22)$$

y de  $\lambda|(s, \mathbf{y}_s)$  dada en (2.15).

5.- A partir de (2.18) y (2.22) aplicando el apartado (a) del lema B.9 se obtiene la distribución de  $\mathbf{Y}_u|(s, \mathbf{y}_s), \lambda$  dada en (2.13).

6.- A partir de (2.19) y (2.22) aplicando el apartado (a) del lema B.9 se obtiene la distribución de  $\bar{Y}_u|(s, \mathbf{y}_s), \lambda$  dada en (2.14).

7.- A partir de (2.14) y de (2.15) se obtiene, aplicando la definición de estimador Bayes, el lema B.3, y que:

$$\begin{aligned} E(\bar{Y}_u|(s, \mathbf{y}_s)) &= E^{\lambda|(s, \mathbf{y}_s)} E(\bar{Y}_u|(s, \mathbf{y}_s), \lambda), \\ Var(\bar{Y}_u|(s, \mathbf{y}_s)) &= E^{\lambda|(s, \mathbf{y}_s)} Var(\bar{Y}_u|(s, \mathbf{y}_s), \lambda) + Var^{\lambda|(s, \mathbf{y}_s)} E(\bar{Y}_u|(s, \mathbf{y}_s), \lambda), \end{aligned}$$

se obtiene el resultado (d) de este teorema dado en (2.16).

**C.S.Q.D.**

El resultado anterior proporciona las distribuciones necesarias para obtener las distribuciones predictivas del vector poblacional no observado y la media no observada, que son propias por serlo las correspondientes distribuciones condicionadas a  $\lambda$  así como la distribución de  $\lambda|(s, \mathbf{y}_s)$ , aplicando integración numérica. Y establece la existencia del estimador Bayes de la media poblacional y su varianza condicionada a la muestra, ya que el teorema 2.3 apoyado en el lema B.3 garantiza la existencia de  $E^{\lambda|(s, \mathbf{y}_s)}[\hat{\beta}_B]$ ,  $Var^{\lambda|(s, \mathbf{y}_s)}[\hat{\beta}_B]$  y  $E^{\lambda|(s, \mathbf{y}_s)}\{b_1[(N-n)\lambda + \mathbf{1}_{N-n}^t \mathbf{X}_u (\mathbf{I}_p - \alpha) \mathbf{C} \mathbf{X}_u^t \mathbf{1}_{N-n}]\}$ ; estos momentos pueden calcularse nuevamente por integración numérica.

El teorema 2.3 permite, por tanto, la obtención de todos los resultados de interés para el modelo propuesto con el único inconveniente de la realización de algunas integrales numéricas. Un método alternativo, que evita el estudio una por una de estas integrales numéricas y es más fácil de implementar en el ordenador (por ejemplo con el programa BUGS (Bayesian Inference using Gibbs Sampler)), es el método basado en el procedimiento de muestreo Gibbs (ver C1 y C2 para un descripción del método Gibbs y del programa BUGS). Las condicionales necesarias para la aplicación del algoritmo Gibbs Sampler vienen dadas en el siguiente teorema.

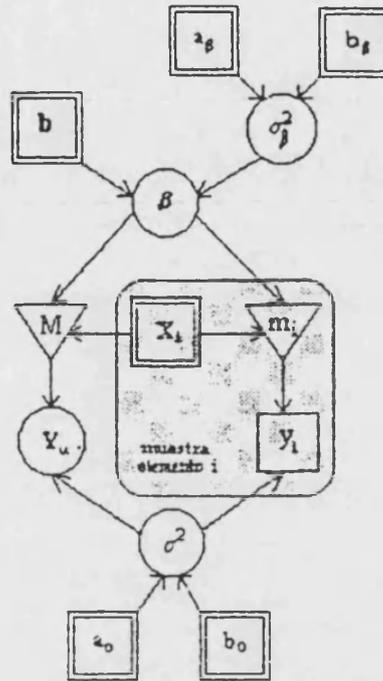


Figura 2.1: Representación gráfica del modelo Bayes de Regresión con cociente entre varianzas desconocido

**Teorema 2.4** *El modelo Bayes de Regresión con cociente entre varianzas desconocido dado en (2.12), es un modelo DAG<sup>2</sup> (figura 2.1), con distribuciones condicionales dadas por:*

$$f(y_u | y_s, \beta, \sigma^2, \sigma_\beta^2) = f(y_u | \beta, \sigma^2) = N_{N-n}(y_u | X_u \beta, \sigma^2 I_{N-n}), \quad (2.23)$$

$$\pi(\beta | y, \sigma^2, \sigma_\beta^2) = N_p(\beta | \alpha^* \hat{\beta}^* + (I_p - \alpha^*) b, (I_p - \alpha^*) C \sigma_\beta^2), \quad (2.24)$$

$$\pi(\sigma_\beta^2 | y, \beta, \sigma^2) = \pi(\sigma_\beta^2 | \beta) = Ga^{-1}(\sigma_\beta^2 | \frac{p}{2} + a_\beta, \frac{1}{2}(\beta - b)^t C^{-1}(\beta - b) + b_\beta), \quad (2.25)$$

$$\pi(\sigma^2 | y, \beta, \sigma_\beta^2) = \pi(\sigma^2 | y, \beta) = Ga^{-1}(\sigma^2 | \frac{N}{2} + a_0, \frac{1}{2}(y - X\beta)^t (y - X\beta) + b_0), \quad (2.26)$$

donde:

$$\hat{\beta}^* = (X^t X)^{-1} X^t y, \quad (2.27)$$

<sup>2</sup>Las siglas DAG significan Directed Acyclic Graphical model y se aplican para referirse a modelos jerárquicos en los que la dependencia probabilística entre las variables es directa de modo que la distribución conjunta de todo el modelo es producto de las distribuciones conjuntas de cada una de sus jerarquías. Estos modelos fueron estudiados por Whittaker(90)[118]. Las subsecciones C.2.1 y C.2.2 proporcionan una definición más completa de estos modelos y de sus propiedades

$$\alpha^* = (\mathbf{X}^t \mathbf{X} + \mathbf{C}^{-1} \frac{\sigma^2}{\sigma_\beta^2})^{-1} \mathbf{X}^t \mathbf{X}.$$

**Demostración.** De la definición 2.3 del modelo en estudio se deduce directamente la representación gráfica de la figura 2.1, que corresponde a un modelo DAG; según la definición introducida en la subsección C.2.1. Y los resultados de las distribuciones condicionales se obtienen a partir de la distribución conjunta de  $(\mathbf{Y}_u^t, \mathbf{Y}_s^t, \beta^t, \sigma^2, \sigma_\beta^2)^t$ , que viene dada por:

$$\begin{aligned} f((\mathbf{y}_u^t, \mathbf{y}_s^t, \beta^t, \sigma^2, \sigma_\beta^2)^t) = \\ N(\mathbf{y}_u | \mathbf{X}_u \beta, \sigma^2 \mathbf{I}_{N-n}) \times N(\mathbf{y}_s | \mathbf{X}_s \beta, \sigma^2 \mathbf{I}_n) \times N(\beta | \mathbf{b}, \sigma_\beta^2 \mathbf{C}) \times \\ Ga^{-1}(\sigma^2 | a_0, b_0) \times Ga^{-1}(\sigma_\beta^2 | a_\beta, b_\beta), \end{aligned}$$

separando los términos correspondientes a cada condicional (que contienen la variable cuya condicional estamos hallando), de modo que:

- $f(\mathbf{y}_u | \mathbf{y}_s, \beta, \sigma^2, \sigma_\beta^2) = N(\mathbf{y}_u | \mathbf{X}_u \beta, \sigma^2 \mathbf{I}_{N-n})$ , con lo que se obtiene (2.23).
- $f(\beta | \mathbf{y}, \sigma^2, \sigma_\beta^2) \propto N(\mathbf{y}_u | \mathbf{X}_u \beta, \sigma^2 \mathbf{I}_{N-n}) \times N(\mathbf{y}_s | \mathbf{X}_s \beta, \sigma^2 \mathbf{I}_n) \times N(\beta | \mathbf{b}, \sigma_\beta^2 \mathbf{C})$ , de donde se obtiene, completando cuadrados respecto a  $\beta$ , la distribución dada en (2.24).
- $\pi(\sigma_\beta^2 | \mathbf{y}, \beta, \sigma^2) \propto N(\beta | \mathbf{b}, \sigma_\beta^2 \mathbf{C}) \times Ga^{-1}(\sigma_\beta^2 | a_\beta, b_\beta)$ , de donde se obtiene, juntando los términos correspondientes a potencias y funciones exponenciales de  $\sigma_\beta^2$ , la distribución dada en (2.25).
- $\pi(\sigma^2 | \mathbf{y}, \beta, \sigma_\beta^2) \propto N(\mathbf{y}_u | \mathbf{X}_u \beta, \sigma^2 \mathbf{I}_{N-n}) \times N(\mathbf{y}_s | \mathbf{X}_s \beta, \sigma^2 \mathbf{I}_n) \times Ga^{-1}(\sigma^2 | a_0, b_0)$ , de donde se obtiene, juntando los términos correspondientes a potencias y funciones exponenciales de  $\sigma^2$ , la distribución dada en (2.26).

**C.S.Q.D.**

Notemos que todas las condicionales necesarias para la aplicación del algoritmo de muestreo Gibbs son distribuciones estándar. Esta circunstancia facilita la aplicación del algoritmo Gibbs, que en este caso puede ser implementado a través del programa BUGS (el modelo a implementar es DAG, las distribuciones iniciales del modelo son independientes entre sí y las iniciales y condicionales tienen forma estándar), tal y como veremos en la sección 2.4.

## 2.2.4 Modelos Bayes en una etapa mínimo informativos

Para concluir el estudio de los modelos Bayesianos de Regresión en una etapa, incluimos a continuación unas versiones mínimo informativas del modelo Bayes de Regresión Normal y el modelo Bayes de Regresión Normal-Gamma. El modelo Bayes de Regresión Normal mínimo informativo y el modelo Bayes de Regresión Normal-Gamma mínimo

informativos, se aplican para modelizar una característica continua de una población no estructurada (monoetápica), cuando no se desea introducir información inicial sobre  $\beta$  y sobre  $\beta$  y  $\sigma^2$ , en los casos varianza de la población conocida y desconocida respectivamente.

### Modelo Bayes de Regresión Normal mínimo informativo

**Definición 2.4** Llamaremos modelo Bayes de Regresión Normal mínimo informativo a un modelo Bayes de Regresión Normal en el que la distribución que asumimos sobre  $\beta$  es una mínimo informativa de la forma  $\pi(\beta) \propto 1$ .

**Corolario 2.1** Consideremos el modelo Bayes de Regresión Normal mínimo informativo dado en la definición 2.4, entonces se tiene que las distribuciones predictivas para el vector no observado y la media poblacional no observada y el estimador de la media poblacional y su varianza condicionada a la muestra son los dados en (2.2), (2.3) y (2.4), tomando  $\alpha = \mathbf{I}_p$ .

**Demostración.** La demostración del corolario es inmediata a partir del corolario 3.3 (capítulo 3), tomando:  $m = 1$ ,  $\mathbf{Y} = (Y_1, Y_2, \dots, Y_N)^t$ ,  $\theta_1 = \beta$ ,  $\mathbf{A}_1 = \mathbf{X}$ ,  $\mathbf{V}_1 = \mathbf{V}$  y denotando  $\hat{\theta}_1^B = \hat{\beta}_B$ .

C.S.Q.D.

### Modelo Bayes de Regresión Normal-Gamma mínimo informativo

**Definición 2.5** Llamaremos modelo Bayes de Regresión Normal-Gamma mínimo informativo a un modelo Bayesiano de regresión normal-gamma en el que las distribuciones iniciales para  $\beta$  y  $\sigma^2$  son sustituidas por una distribución inicial conjunta para  $(\beta, \sigma^2)$  mínimo informativa<sup>3</sup> de la forma  $\pi(\beta, \sigma^2) \propto \frac{1}{\sigma^2}$ .

**Corolario 2.2** Consideremos el modelo Bayes de Regresión Normal-Gamma mínimo informativo dado en la definición 2.5, entonces se tiene que las distribuciones predictivas para el vector no observado y la media poblacional no observada y el estimador de la media poblacional y su varianza condicionada a la muestra son las dadas en (2.8), (2.9) y (2.10) tomando:

$$\begin{aligned} \alpha &= \mathbf{I}_p, \\ a_1 &= \frac{n-p}{2}, \\ b_1 &= \frac{1}{2} \mathbf{y}_s^t [\mathbf{W}_s^{-1} + \mathbf{W}_s^{-1} \mathbf{X}_s (\mathbf{X}_s^t \mathbf{W}_s^{-1} \mathbf{X}_s)^{-1} \mathbf{X}_s^t \mathbf{W}_s^{-1}] \mathbf{y}_s. \end{aligned} \tag{2.28}$$

<sup>3</sup>ver nota 7 a pie de página de la subsección 3.4.3

**Demostración.** La demostración de este corolario es inmediata a partir de los resultados del corolario 3.4 (capítulo 3), tomando:  $m = 1$ ,  $\mathbf{Y} = (Y_1, Y_2, \dots, Y_N)^t$ ,  $\theta_1 = \beta$ ,  $\mathbf{A}_1 = \mathbf{X}$ ,  $\mathbf{W}_1 = \mathbf{W}$  y denotando  $\hat{\theta}_1^B = \hat{\beta}_B$ .

C.S.Q.D.

Los modelos Bayes mínimo informativos tradicionales, comentados en la subsección 1.4.2 son casos particulares de estos modelos para una elección adecuada de  $\mathbf{V}$  y  $\mathbf{X}$ , obteniéndose las distribuciones predictivas del vector poblacional no observado, la media poblacional no observada y el estimador Bayes para la media poblacional y su varianza condicionada a la muestra aplicando los corolarios anteriores para esas definiciones de  $\mathbf{V}$  y  $\mathbf{X}$  (ver el apartado estimadores de la media poblacional de la subsección 1.4.2). Señalemos, en este punto, que en los modelos normal-gamma mínimo informativos los resultados obtenidos por algunos autores pueden diferir en el número de grados de libertad de las distribuciones predictivas por una elección diferente de la distribución mínimo informativa.

Por último, recordemos que a partir de los estimadores para la media poblacional y varianza condicionada a la muestra obtenidos para el modelo Bayes de Regresión Normal mínimo informativo, se recuperan los estimadores para los correspondientes modelos de superpoblación clásicos con matriz de varianzas-covarianzas conocida, y para el modelo Bayes de Regresión Normal-Gamma mínimo informativo, los estimadores para el correspondiente modelo clásico con matriz de varianzas-covarianzas desconocida. También se recuperan a partir de estos dos modelos los estimadores más habituales de población fija para una elección adecuada de  $p$ ,  $\mathbf{V}$  (o  $\mathbf{W}$ ) y  $\mathbf{X}$ .

## 2.3 Modelos Bayesianos en dos etapas con una jerarquía

En esta sección nos dedicaremos a la definición y estudio de los modelos Bayes en dos etapas con una jerarquía para poblaciones con estructura bietápica (en las que los elementos de la población se encuentran divididos en unidades). Estos modelos se aplican, habitualmente, cuando la población en estudio ha sido estructurada como resultado de un muestreo estratificado en unidades, cuando estas unidades son heterogéneas entre sí o simplemente cuando disponemos de una buena información para la asignación de las distribuciones iniciales sobre las pendientes de cada unidad que hacen innecesario invocar una jerarquía más.

La notación aplicada en toda la sección es la que fue introducida en el capítulo 1 para poblaciones estructuradas en dos etapas, y se encuentra reunida en el apartado estructura bietápica del apéndice A. Los resultados inferenciales obtenidos en esta sección, distribuciones predictivas del vector poblacional no observado y de la media no observada dada la muestra y los estimadores de las medias poblacional y media de cada unidad y sus varianzas condicionadas a la muestra, se apoyarán en una muestra

obtenida mediante un muestreo en dos etapas dado de tamaños fijos y no informativo, en el que se muestrean todas las unidades, esto es,  $k = K$  y los índices de los elementos muestreados en cada unidad son distintos entre sí.

### 2.3.1 Modelo Bayes de Regresión en dos etapas con una jerarquía Normal

En primer lugar definiremos y estudiaremos el modelo Bayes de Regresión en dos etapas con una jerarquía Normal, este modelo se aplica para inferir sobre una característica continua en una población bietápica de las características señaladas al principio de la sección, cuando la varianza de la población es conocida.

**Definición 2.6** Dado el vector poblacional  $\mathbf{Y} = (\mathbf{Y}_1^t, \mathbf{Y}_2^t, \dots, \mathbf{Y}_K^t)^t$  con  $\mathbf{Y}_i = (Y_{i1}, Y_{i2}, \dots, Y_{iN_i})^t$ ,  $i = 1, 2, \dots, K$  para el cual asociada a cada variable  $Y_{ij}$  tenemos una información auxiliar representada por un vector  $\mathbf{x}_{ij} = (x_{ij1}, \dots, x_{ijp})$  conocido,  $j = 1, 2, \dots, N_i$ ,  $i = 1, 2, \dots, K$ , llamaremos modelo Bayes de Regresión en dos etapas con una jerarquía Normal al definido por:

$$\begin{aligned} \mathbf{Y}_i | \mathbf{X}_i, \boldsymbol{\beta}_i &\sim N(\mathbf{X}_i \boldsymbol{\beta}_i, \mathbf{V}_i), \text{ independientes, } i = 1, 2, \dots, K, \\ \boldsymbol{\beta}_i &\sim N_p(\boldsymbol{\beta}, \mathbf{B}_\beta), \text{ i.i.d., } i = 1, 2, \dots, K, \end{aligned} \quad (2.29)$$

donde  $\mathbf{X}_i = (\mathbf{x}_{i1}^t, \mathbf{x}_{i2}^t, \dots, \mathbf{x}_{iN_i}^t)^t$ , es una matriz de rango  $p$ ,  $i = 1, 2, \dots, K$ ,  $\boldsymbol{\beta}$  es un  $p$ -vector conocido y  $\mathbf{V}_i$ ,  $i = 1, 2, \dots, K$  y  $\mathbf{B}_\beta$  son matrices simétricas definidas positivas conocidas de dimensiones  $N_i \times N_i$ ,  $i = 1, 2, \dots, K$  y  $p \times p$  respectivamente.

Este modelo presenta como caso particular para  $p = 1$ ,  $\mathbf{X}_i = \mathbf{1}_{N_i}$ ,  $i = 1, 2, \dots, K$ , el modelo Bayes Básico en dos etapas con una jerarquía Normal y como particularización de éste el modelo estudiado por Murgui(82)[76], referenciado en el apartado estimadores de la media poblacional de la subsección 1.4.2.

Los resultados inferenciales asociados a este modelo dada una muestra con las características señaladas al principio de la sección, aparecen reflejados en el siguiente teorema:

**Teorema 2.5** Consideremos el modelo Bayes de Regresión en dos etapas con una jerarquía Normal dado en (2.29), entonces se tiene que:

(a)  $\mathbf{Y}_{u_i} | (s, \mathbf{y}_s)$  son independientes para  $i = 1, 2, \dots, K$ , con distribuciones predictivas normales con media y varianza dadas por:

$$\begin{aligned} E(\mathbf{Y}_{u_i} | (s, \mathbf{y}_s)) &= \mathbf{X}_{u_i} \hat{\boldsymbol{\beta}}_i^B + \mathbf{V}_{su_i}^t \mathbf{V}_{s_i}^{-1} (\mathbf{y}_{s_i} - \mathbf{X}_{s_i} \hat{\boldsymbol{\beta}}_i^B), \\ \text{Var}(\mathbf{Y}_{u_i} | (s, \mathbf{y}_s)) &= \mathbf{V}_{u_i} - \mathbf{V}_{su_i}^t \mathbf{V}_{s_i}^{-1} \mathbf{V}_{su_i} + \\ &\quad + (\mathbf{X}_{u_i} - \mathbf{V}_{su_i}^t \mathbf{V}_{s_i}^{-1} \mathbf{X}_{s_i}) \boldsymbol{\alpha}_i (\mathbf{X}_{s_i}^t \mathbf{V}_{s_i}^{-1} \mathbf{X}_{s_i})^{-1} (\mathbf{X}_{u_i} - \mathbf{V}_{su_i}^t \mathbf{V}_{s_i}^{-1} \mathbf{X}_{s_i})^t, \end{aligned} \quad (2.30)$$

(b)  $\bar{Y}_{u_i}|(s, \mathbf{y}_s)$  se distribuye normal con media y varianza dadas por:

$$\begin{aligned} E(\bar{Y}_{u_i}|(s, \mathbf{y}_s)) &= \frac{1}{N_i - n_i} \mathbf{1}_{N_i - n_i}^t [\mathbf{X}_{u_i} \hat{\beta}_i^B + \mathbf{V}_{su_i}^t \mathbf{V}_{s_i}^{-1} (\mathbf{y}_{s_i} - \mathbf{X}_{s_i} \hat{\beta}_i^B)], \\ \text{Var}(\bar{Y}_{u_i}|(s, \mathbf{y}_s)) &= \frac{1}{(N_i - n_i)^2} [\mathbf{1}_{N_i - n_i}^t (\mathbf{V}_{u_i} - \mathbf{V}_{su_i}^t \mathbf{V}_{s_i}^{-1} \mathbf{V}_{su_i}) \mathbf{1}_{N_i - n_i} + \\ &\quad + \mathbf{1}_{N_i - n_i}^t (\mathbf{X}_{u_i} - \mathbf{V}_{su_i}^t \mathbf{V}_{s_i}^{-1} \mathbf{X}_{s_i}) \alpha_i (\mathbf{X}_{s_i}^t \mathbf{V}_{s_i}^{-1} \mathbf{X}_{s_i})^{-1} (\mathbf{X}_{u_i} - \mathbf{V}_{su_i}^t \mathbf{V}_{s_i}^{-1} \mathbf{X}_{s_i})^t \mathbf{1}_{N_i - n_i}], \end{aligned} \quad (2.31)$$

(c) el estimador Bayes para la media de la unidad  $i$  y su varianza condicionada a la muestra vienen dados por:

$$\begin{aligned} \hat{Y}_i &= \frac{n_i}{N_i} \bar{y}_{s_i} + \frac{1}{N_i} \mathbf{1}_{N_i - n_i}^t [\mathbf{X}_{u_i} \hat{\beta}_i^B + \mathbf{V}_{su_i}^t \mathbf{V}_{s_i}^{-1} (\mathbf{y}_{s_i} - \mathbf{X}_{s_i} \hat{\beta}_i^B)], \\ \text{Var}(\bar{Y}_i|(s, \mathbf{y}_s)) &= \frac{1}{N_i^2} [\mathbf{1}_{N_i - n_i}^t (\mathbf{V}_{u_i} - \mathbf{V}_{su_i}^t \mathbf{V}_{s_i}^{-1} \mathbf{V}_{su_i}) \mathbf{1}_{N_i - n_i} + \\ &\quad + \mathbf{1}_{N_i - n_i}^t (\mathbf{X}_{u_i} - \mathbf{V}_{su_i}^t \mathbf{V}_{s_i}^{-1} \mathbf{X}_{s_i}) \alpha_i (\mathbf{X}_{s_i}^t \mathbf{V}_{s_i}^{-1} \mathbf{X}_{s_i})^{-1} (\mathbf{X}_{u_i} - \mathbf{V}_{su_i}^t \mathbf{V}_{s_i}^{-1} \mathbf{X}_{s_i})^t \mathbf{1}_{N_i - n_i}], \end{aligned} \quad (2.32)$$

(d)  $\bar{Y}_u|(s, \mathbf{y}_s)$  se distribuye normal con media y varianza dadas por:

$$\begin{aligned} E(\bar{Y}_u|(s, \mathbf{y}_s)) &= \frac{1}{N - n} \sum_{i=1}^K \mathbf{1}_{N_i - n_i}^t [\mathbf{X}_{u_i} \hat{\beta}_i^B + \mathbf{V}_{su_i}^t \mathbf{V}_{s_i}^{-1} (\mathbf{y}_{s_i} - \mathbf{X}_{s_i} \hat{\beta}_i^B)], \\ \text{Var}(\bar{Y}_u|(s, \mathbf{y}_s)) &= \frac{1}{(N - n)^2} \left\{ \sum_{i=1}^K [\mathbf{1}_{N_i - n_i}^t (\mathbf{V}_{u_i} - \mathbf{V}_{su_i}^t \mathbf{V}_{s_i}^{-1} \mathbf{V}_{su_i}) \mathbf{1}_{N_i - n_i} + \right. \\ &\quad \left. + \mathbf{1}_{N_i - n_i}^t (\mathbf{X}_{u_i} - \mathbf{V}_{su_i}^t \mathbf{V}_{s_i}^{-1} \mathbf{X}_{s_i}) \alpha_i (\mathbf{X}_{s_i}^t \mathbf{V}_{s_i}^{-1} \mathbf{X}_{s_i})^{-1} (\mathbf{X}_{u_i} - \mathbf{V}_{su_i}^t \mathbf{V}_{s_i}^{-1} \mathbf{X}_{s_i})^t \mathbf{1}_{N_i - n_i}] \right\}, \end{aligned} \quad (2.33)$$

y (e) el estimador Bayes para la media poblacional y su varianza condicionada a la muestra vienen dados por:

$$\begin{aligned} \hat{Y} &= \frac{n}{N} \bar{y}_s + \frac{1}{N} \sum_{i=1}^K \mathbf{1}_{N_i - n_i}^t [\mathbf{X}_{u_i} \hat{\beta}_i^B + \mathbf{V}_{su_i}^t \mathbf{V}_{s_i}^{-1} (\mathbf{y}_{s_i} - \mathbf{X}_{s_i} \hat{\beta}_i^B)], \\ \text{Var}(\bar{Y}|(s, \mathbf{y}_s)) &= \frac{1}{N^2} \left\{ \sum_{i=1}^K [\mathbf{1}_{N_i - n_i}^t (\mathbf{V}_{u_i} - \mathbf{V}_{su_i}^t \mathbf{V}_{s_i}^{-1} \mathbf{V}_{su_i}) \mathbf{1}_{N_i - n_i} + \right. \\ &\quad \left. + \mathbf{1}_{N_i - n_i}^t (\mathbf{X}_{u_i} - \mathbf{V}_{su_i}^t \mathbf{V}_{s_i}^{-1} \mathbf{X}_{s_i}) \alpha_i (\mathbf{X}_{s_i}^t \mathbf{V}_{s_i}^{-1} \mathbf{X}_{s_i})^{-1} (\mathbf{X}_{u_i} - \mathbf{V}_{su_i}^t \mathbf{V}_{s_i}^{-1} \mathbf{X}_{s_i})^t \mathbf{1}_{N_i - n_i}] \right\}, \end{aligned} \quad (2.34)$$

donde  $\mathbf{V}_{s_i} = \text{Var}_{\mathbf{Y}|\beta}(\mathbf{Y}_{s_i})$ ,  $\mathbf{V}_{u_i} = \text{Var}_{\mathbf{Y}|\beta}(\mathbf{Y}_{u_i})$ ,  $\mathbf{V}_{su_i} = \text{Cov}_{\mathbf{Y}|\beta}(\mathbf{Y}_{s_i}, \mathbf{Y}_{u_i})$ ,  $i = 1, 2, \dots, K$ ,  $\bar{y}_s = \frac{\sum_{(i,j) \in s} y_{ij}}{n}$  es la media muestral de la población,  $\bar{y}_{s_i}$  y  $\hat{\beta}_i^B = (\mathbf{X}_{s_i}^t \mathbf{V}_{s_i}^{-1}$

$\mathbf{X}_{s_i})^{-1}\mathbf{X}_{s_i}^t \mathbf{V}_{s_i}^{-1} \mathbf{y}_{s_i}$  son, respectivamente la media muestra y el estimador de regresión clásico para la unidad  $i$  para  $i = 1, 2, \dots, K$ , y donde para  $i = 1, 2, \dots, K$ :

$$\begin{aligned}\alpha_i &= (\mathbf{X}_{s_i}^t \mathbf{V}_{s_i}^{-1} \mathbf{X}_{s_i} + \mathbf{B}_\beta^{-1})^{-1} \mathbf{X}_{s_i}^t \mathbf{V}_{s_i}^{-1} \mathbf{X}_{s_i}, \\ \hat{\beta}_i^B &= E(\beta_i | (s, \mathbf{y}_s)) = \alpha_i \hat{\beta}_i + (\mathbf{I}_p - \alpha_i) \beta.\end{aligned}\quad (2.35)$$

**Demostración.** La demostración de los apartados (a), (d) y (e) son inmediatas a partir del teorema 3.5 (capítulo 3), tomando:  $m = 1$ ,  $\mathbf{Y} = (\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_K)^t$ ,  $\mathbf{Y}_i = (Y_{i1}, Y_{i2}, \dots, Y_{iN_i})^t$ ,  $i = 1, 2, \dots, K$ ,  $\theta_1 = (\beta_1^t, \dots, \beta_K^t)^t$ ,  $\mathbf{A}_1 = \text{diag}_{i=1}^K \{\mathbf{X}_i\}$ ,  $\mathbf{X}_i = (\mathbf{x}_{i1}^t, \mathbf{x}_{i2}^t, \dots, \mathbf{x}_{iN_i}^t)^t$ ,  $i = 1, 2, \dots, K$ ,  $\mathbf{x}_{ij} = (x_{ij1}, x_{ij2}, \dots, x_{ijp})$ ,  $j = 1, 2, \dots, N_i$ ,  $i = 1, 2, \dots, K$ ,  $\mathbf{A}_2 = \underbrace{(\mathbf{I}_p, \dots, \mathbf{I}_p)}_K^t \beta$ ,  $\mathbf{V}_1 = \text{diag}_{i=1}^K \{\mathbf{V}_i\}$  y  $\mathbf{V}_2 = \text{diag}_{i=1}^K \{\mathbf{B}_\beta\}$ , denotando

por  $\hat{\beta}_i^B$  el  $p$ -vector de  $\hat{\theta}_1^B$  correspondiente a la unidad  $i$ . Los restantes resultados son inmediatos a partir de la distribución predictiva anterior y que la marginal de una normal y la combinación lineal de normales es normal, y la definición de estimador Bayes adaptada a la estimación de la media de la unidad  $i$  con  $i = 1, 2, \dots, K$ .

C.S.Q.D.

Notemos que la independencia a priori entre las unidades en (2.29) se traduce en la independencia de las distribuciones predictivas de los vectores poblacionales no observados y la media no observada de cada unidad, y en la obtención para cada unidad de las mismas distribuciones que se obtenían en el modelo Bayes de Regresión Normal.

Análogamente al modelo Bayes de Regresión Normal, pero ahora referido a cada unidad, se pueden traducir las observaciones que se realizaban en la sección anterior. Así, notemos a partir de las expresiones (2.30) a (2.34), que la contribución de la muestra a las distribuciones predictivas del vector poblacional no observado de la unidad  $i$  y a la estimación de la media poblacional no observada de la unidad  $i$ ,  $i = 1, 2, \dots, K$ , se articula a través de  $\mathbf{y}_{s_i}$ ,  $\bar{y}_{s_i}$  (la media muestral de la unidad  $i$ ),  $\hat{\beta}_i$  (el estimador clásico de regresión de la unidad  $i$ ) y  $\alpha_i$ ,  $i = 1, 2, \dots, K$ , (e indirectamente a través de  $\mathbf{V}_{s_i}$ ,  $\mathbf{V}_{su_i}$  y  $\mathbf{X}_{s_i}$ ,  $i = 1, 2, \dots, K$ ). Con  $\alpha_i$ ,  $i = 1, 2, \dots, K$ , un factor de encogimiento ( $\alpha_i = (\mathbf{I}_p - \alpha_i) \mathbf{B}_\beta^{-1} \mathbf{X}_{s_i}^t \mathbf{V}_{s_i}^{-1} \mathbf{X}_{s_i}$ ,  $i = 1, 2, \dots, K$ ), que mide la proporción de varianza respecto a  $\beta_i$  no explicada por la muestra, ya que, para  $i = 1, 2, \dots, K$ :

$$\alpha_i = [\text{Var}(\beta_i) - \text{Var}(\beta_i | (s, \mathbf{y}_s))] \text{Var}^{-1}(\beta_i). \quad (2.36)$$

La expresión (2.36) se obtiene a partir de la definición de  $\alpha_i$ ,  $i = 1, 2, \dots, K$ , notando que  $\text{Var}(\beta_i | (s, \mathbf{y}_s)) = (\mathbf{X}_{s_i}^t \mathbf{V}_{s_i}^{-1} \mathbf{X}_{s_i} + \mathbf{B}_\beta^{-1})^{-1}$  (resultado previo en la demostración anterior).

A partir de esta expresión se tiene que, el estimador Bayes de la pendiente de Regresión de cada unidad, es una "combinación lineal" del estimador de regresión clásico y

la pendiente asignada a priori a la unidad, que refleja en qué medida predomina la información introducida en la distribución inicial sobre la muestra. Por otra parte, notemos que la varianza de las distribuciones predictivas de cada unidad y varianza de los estimadores de cada unidad, disminuye en la medida que el factor  $\alpha_i$  de la unidad  $i$  disminuye, esto es, en la medida que nuestros conocimientos iniciales acerca de la pendiente de la unidad aumentan (para  $i = 1, 2, \dots, K$ ).

Este modelo puede generalizarse, para describir situaciones en las que las unidades son independientes y muy distintas entre sí, asumiendo que las variables pendiente de cada unidad son independientes y normales pero no idénticamente distribuidas, esto es, sustituir en el modelo (2.29), las distribuciones iniciales sobre la pendiente de cada unidad por:  $\beta_i \sim N_p(\mathbf{b}_i, \mathbf{B}_i)$ , independientes para  $i = 1, 2, \dots, K$ , con  $\mathbf{b}_i$  un  $p$ -vector conocido, para  $i = 1, 2, \dots, K$ , y  $\mathbf{B}_i$  una matriz simétrica y definida positiva de dimensión  $p \times p$  conocida, para  $i = 1, 2, \dots, K$ . Para este nuevo modelo se obtienen, aplicando el teorema 3.5 adecuadamente, los resultados (2.30) a (2.34), sustituyendo  $\mathbf{B}_\beta$  por  $\mathbf{B}_i$  y  $\beta$  por  $\mathbf{b}_i$ , para  $i = 1, 2, \dots, K$ . El inconveniente de esta generalización reside en el número de parámetros iniciales que han de ser asignados.

### 2.3.2 Modelo Bayes de Regresión en dos etapas con una jerarquía Normal-Gamma

El modelo Bayes de Regresión en dos etapas con una jerarquía Normal-Gamma, que definiremos a continuación, se aplica cuando deseamos inferir sobre una característica continua de una población bietápica de las características señaladas al principio de la sección cuando la varianza no es conocida, aunque sí el cociente entre la varianza de la población y de la distribución inicial asignada por unidad (esto es,  $\mathbf{W}_i \mathbf{C}_\beta^{-1}$ ,  $i = 1, 2, \dots, K$ ).

**Definición 2.7** Dado el vector poblacional  $\mathbf{Y} = (\mathbf{Y}_1^t, \mathbf{Y}_2^t, \dots, \mathbf{Y}_K^t)^t$  con  $\mathbf{Y}_i = (Y_{i1}, Y_{i2}, \dots, Y_{iN_i})^t$ ,  $i = 1, 2, \dots, K$  para el cual asociada a cada variable  $Y_{ij}$  tenemos una información auxiliar representada por un vector  $\mathbf{x}_{ij} = (x_{ij1}, \dots, x_{ijp})$  conocido,  $j = 1, 2, \dots, N_i$ ,  $i = 1, 2, \dots, K$ , llamaremos modelo Bayes de Regresión en dos etapas con una jerarquía Normal-Gamma al definido por:

$$\begin{aligned} \mathbf{Y}_i | \mathbf{X}_i, \beta_i, \sigma^2 &\sim N(\mathbf{X}_i \beta_i, \sigma^2 \mathbf{W}_i), \text{ independientes, } i = 1, 2, \dots, K, \\ \beta_i | \sigma^2 &\sim N_p(\beta, \sigma^2 \mathbf{C}_\beta), \text{ i.i.d., } i = 1, 2, \dots, K, \\ \sigma^2 &\sim Ga^{-1}(a_0, b_0) \end{aligned} \quad (2.37)$$

donde  $\mathbf{X}_i = (\mathbf{x}_{i1}^t, \mathbf{x}_{i2}^t, \dots, \mathbf{x}_{iN_i}^t)^t$ , es una matriz de rango  $p$ , para  $i = 1, 2, \dots, K$ ,  $\beta$  es un  $p$ -vector conocido,  $\mathbf{W}_i$ ,  $i = 1, 2, \dots, K$  y  $\mathbf{C}_\beta$  son matrices simétricas, definidas positivas conocidas de dimensiones  $N_i \times N_i$ ,  $i = 1, 2, \dots, K$  y  $p \times p$  respectivamente,  $a_0$  y  $b_0$  escalares positivos conocidos.

Este modelo presenta como caso particular para  $p = 1$ ,  $\mathbf{X}_i = \mathbf{1}_{N_i}$ ,  $i = 1, 2, \dots, K$ , el modelo Bayes Básico en dos etapas con una jerarquía Normal-Gamma (descrito en la sección 1.4.2).

Notemos de la definición anterior que la distribución sobre  $\sigma^2$  vincula al vector  $\mathbf{Y}$  y al pendiente  $\beta_i$ ,  $i = 1, 2, \dots, K$ . Esta circunstancia se refleja, en la no independencia (aunque si incorrelación) de las distribuciones predictivas del vector no observado de cada unidad, tal y como veremos en el próximo resultado, que recoge los resultados inferenciales de interés para este modelo dada una muestra. Estos resultados pueden interpretarse en términos de  $\mathbf{y}_{s_i}$ ,  $\bar{y}_{s_i}$  (la media muestral de la unidad  $i$ ),  $\hat{\beta}_i$  (el estimador clásico de regresión de la unidad  $i$ ) y  $\alpha_i$ ,  $i = 1, 2, \dots, K$ , (e indirectamente a través de  $\mathbf{W}_{s_i}$ ,  $\mathbf{W}_{s_{u_i}}$  y  $\mathbf{X}_{s_i}$ ,  $i = 1, 2, \dots, K$ ), de forma totalmente análoga al modelo Bayes de Regresión en dos etapas con una jerarquía Normal.

**Teorema 2.6** Consideremos el modelo Bayes de Regresión en dos etapas con una jerarquía Normal-Gamma dado en (2.37), entonces se tiene que:

(a)  $\mathbf{Y}_{u_i} | (s, \mathbf{y}_s)$  se distribuye  $t$  ( $N - n$ )-variante con  $2a_1$  grados de libertad con, para  $a_1 > 0$ , media:

$$E(\mathbf{Y}_{u_i} | (s, \mathbf{y}_s)) = \mathbf{X}_{u_i} \hat{\beta}_i^B + \mathbf{W}_{s_{u_i}}^t \mathbf{W}_{s_i}^{-1} (\mathbf{y}_{s_i} - \mathbf{X}_{s_i} \hat{\beta}_i^B), \quad (2.38)$$

y matriz de varianzas y covarianzas,

\* para  $i = 1, 2, \dots, K$ ,

$$\begin{aligned} \text{Var}(\mathbf{Y}_{u_i} | (s, \mathbf{y}_s)) &= \frac{b_1}{a_1 - 1} \{ \mathbf{W}_{u_i} - \mathbf{W}_{s_{u_i}}^t \mathbf{W}_{s_i}^{-1} \mathbf{W}_{s_{u_i}} + \\ &+ (\mathbf{X}_{u_i} - \mathbf{W}_{s_{u_i}}^t \mathbf{W}_{s_i}^{-1} \mathbf{X}_{s_i}) \alpha_i (\mathbf{X}_{s_i}^t \mathbf{W}_{s_i}^{-1} \mathbf{X}_{s_i})^{-1} (\mathbf{X}_{u_i} - \mathbf{W}_{s_{u_i}}^t \mathbf{W}_{s_i}^{-1} \mathbf{X}_{s_i})^t \}, \end{aligned} \quad (2.39)$$

\* para  $i \neq j$ ,  $i, j = 1, 2, \dots, K$ ,

$$\text{Cov}(\mathbf{Y}_{u_i}, \mathbf{Y}_{u_j} | (s, \mathbf{y}_s)) = \mathbf{0}_{N_i - n_i \times N_j - n_j}. \quad (2.40)$$

(b)  $\bar{Y}_{u_i} | (s, \mathbf{y}_s)$  se distribuye  $t$  con  $2a_1$  grados de libertad y media y varianza, para  $a_1 > 0$ , dadas por:

$$\begin{aligned} E(\bar{Y}_{u_i} | (s, \mathbf{y}_s)) &= \frac{1}{N_i - n_i} \mathbf{1}_{N_i - n_i}^t [\mathbf{X}_{u_i} \hat{\beta}_i^B + \mathbf{W}_{s_{u_i}}^t \mathbf{W}_{s_i}^{-1} (\mathbf{y}_{s_i} - \mathbf{X}_{s_i} \hat{\beta}_i^B)], \quad (2.41) \\ \text{Var}(\bar{Y}_{u_i} | (s, \mathbf{y}_s)) &= \frac{1}{(N_i - n_i)^2} \frac{b_1}{a_1 - 1} [\mathbf{1}_{N_i - n_i}^t (\mathbf{W}_{u_i} - \mathbf{W}_{s_{u_i}}^t \mathbf{W}_{s_i}^{-1} \mathbf{W}_{s_{u_i}}) \mathbf{1}_{N_i - n_i} + \\ &+ \mathbf{1}_{N_i - n_i}^t (\mathbf{X}_{u_i} - \mathbf{W}_{s_{u_i}}^t \mathbf{W}_{s_i}^{-1} \mathbf{X}_{s_i}) \alpha_i (\mathbf{X}_{s_i}^t \mathbf{W}_{s_i}^{-1} \mathbf{X}_{s_i})^{-1} (\mathbf{X}_{u_i} - \mathbf{W}_{s_{u_i}}^t \mathbf{W}_{s_i}^{-1} \mathbf{X}_{s_i})^t \mathbf{1}_{N_i - n_i}], \end{aligned}$$

(c) el estimador Bayes para la media de la unidad  $i$ ,  $i = 1, 2, \dots, K$  y su varianza condicionada a la muestra, para  $a_1 > 0$ , vienen dados por:

$$\begin{aligned}\widehat{Y}_i &= \frac{n_i}{N_i} \bar{y}_{s_i} + \frac{1}{N_i} \mathbf{1}_{N_i-n_i}^t [\mathbf{X}_{u_i} \widehat{\beta}_i^B + \mathbf{W}_{s_{u_i}}^t \mathbf{W}_{s_i}^{-1} (\mathbf{y}_{s_i} - \mathbf{X}_{s_i} \widehat{\beta}_i^B)], \\ \text{Var}(\bar{Y}_i | (s, \mathbf{y}_s)) &= \frac{1}{N_i^2} \frac{b_1}{a_1 - 1} [\mathbf{1}_{N_i-n_i}^t (\mathbf{W}_{u_i} - \mathbf{W}_{s_{u_i}}^t \mathbf{W}_{s_i}^{-1} \mathbf{W}_{s_{u_i}}) \mathbf{1}_{N_i-n_i} + \\ &+ \mathbf{1}_{N_i-n_i}^t (\mathbf{X}_{u_i} - \mathbf{W}_{s_{u_i}}^t \mathbf{W}_{s_i}^{-1} \mathbf{X}_{s_i}) \alpha_i (\mathbf{X}_{s_i}^t \mathbf{W}_{s_i}^{-1} \mathbf{X}_{s_i})^{-1} (\mathbf{X}_{u_i} - \mathbf{W}_{s_{u_i}}^t \mathbf{W}_{s_i}^{-1} \mathbf{X}_{s_i})^t \mathbf{1}_{N_i-n_i}],\end{aligned}\quad (2.42)$$

(d)  $\bar{Y}_u | (s, \mathbf{y}_s)$  se distribuye  $t$  con  $2a_1$  grados de libertad y media y varianza, para  $a_1 > 0$ , dadas por:

$$\begin{aligned}E(\bar{Y}_u | (s, \mathbf{y}_s)) &= \frac{1}{N-n} \sum_{i=1}^K \mathbf{1}_{N_i-n_i}^t [\mathbf{X}_{u_i} \widehat{\beta}_i^B + \mathbf{W}_{s_{u_i}}^t \mathbf{W}_{s_i}^{-1} (\mathbf{y}_{s_i} - \mathbf{X}_{s_i} \widehat{\beta}_i^B)], \quad (2.43) \\ \text{Var}(\bar{Y}_u | (s, \mathbf{y}_s)) &= \frac{1}{(N-n)^2} \frac{b_1}{a_1 - 1} \left\{ \sum_{i=1}^K [\mathbf{1}_{N_i-n_i}^t (\mathbf{W}_{u_i} - \mathbf{W}_{s_{u_i}}^t \mathbf{W}_{s_i}^{-1} \mathbf{W}_{s_{u_i}}) \mathbf{1}_{N_i-n_i} + \right. \\ &+ \left. \mathbf{1}_{N_i-n_i}^t (\mathbf{X}_{u_i} - \mathbf{W}_{s_{u_i}}^t \mathbf{W}_{s_i}^{-1} \mathbf{X}_{s_i}) \alpha_i (\mathbf{X}_{s_i}^t \mathbf{W}_{s_i}^{-1} \mathbf{X}_{s_i})^{-1} (\mathbf{X}_{u_i} - \mathbf{W}_{s_{u_i}}^t \mathbf{W}_{s_i}^{-1} \mathbf{X}_{s_i})^t \mathbf{1}_{N_i-n_i} \right\},\end{aligned}$$

y (e) el estimador Bayes para la media poblacional y su varianza condicionada a la muestra vienen dados por:

$$\begin{aligned}\widehat{Y} &= \frac{n}{N} \bar{y}_s + \frac{1}{N} \sum_{i=1}^K \mathbf{1}_{N_i-n_i}^t [\mathbf{X}_{u_i} \widehat{\beta}_i^B + \mathbf{W}_{s_{u_i}}^t \mathbf{W}_{s_i}^{-1} (\mathbf{y}_{s_i} - \mathbf{X}_{s_i} \widehat{\beta}_i^B)], \\ \text{Var}(\bar{Y} | (s, \mathbf{y}_s)) &= \frac{1}{N^2} \frac{b_1}{a_1 - 1} \left\{ \sum_{i=1}^K [\mathbf{1}_{N_i-n_i}^t (\mathbf{W}_{u_i} - \mathbf{W}_{s_{u_i}}^t \mathbf{W}_{s_i}^{-1} \mathbf{W}_{s_{u_i}}) \mathbf{1}_{N_i-n_i} + \right. \\ &+ \left. \mathbf{1}_{N_i-n_i}^t (\mathbf{X}_{u_i} - \mathbf{W}_{s_{u_i}}^t \mathbf{W}_{s_i}^{-1} \mathbf{X}_{s_i}) \alpha_i (\mathbf{X}_{s_i}^t \mathbf{W}_{s_i}^{-1} \mathbf{X}_{s_i})^{-1} (\mathbf{X}_{u_i} - \mathbf{W}_{s_{u_i}}^t \mathbf{W}_{s_i}^{-1} \mathbf{X}_{s_i})^t \mathbf{1}_{N_i-n_i} \right\},\end{aligned}\quad (2.44)$$

donde  $\mathbf{W}_{s_i} = \sigma^{-2} \text{Var}_{\mathbf{Y} | \beta}(\mathbf{Y}_{s_i})$ ,  $\mathbf{W}_{u_i} = \sigma^2 \text{Var}_{\mathbf{Y} | \beta}(\mathbf{Y}_{u_i})$ ,  $\mathbf{W}_{s_{u_i}} = \sigma^{-2} \text{Cov}_{\mathbf{Y} | \beta}(\mathbf{Y}_{s_i}, \mathbf{Y}_{u_i})$ ,  $\bar{y}_s = \frac{\sum_{(i,j) \in s} y_{ij}}{n}$  es la media muestral de la población,  $\bar{y}_s$ , y  $\widehat{\beta}_i = (\mathbf{X}_{s_i}^t \mathbf{W}_{s_i}^{-1} \mathbf{X}_{s_i})^{-1} \mathbf{X}_{s_i}^t \mathbf{W}_{s_i}^{-1} \mathbf{y}_{s_i}$  son, respectivamente la media muestra y el estimador de regresión clásico para la unidad  $i$  para  $i = 1, 2, \dots, K$ , y donde para  $i = 1, 2, \dots, K$ :

$$\begin{aligned}\alpha_i &= (\mathbf{X}_{s_i}^t \mathbf{W}_{s_i}^{-1} \mathbf{X}_{s_i} + \mathbf{C}_\beta^{-1})^{-1} \mathbf{X}_{s_i}^t \mathbf{W}_{s_i}^{-1} \mathbf{X}_{s_i}, \\ \widehat{\beta}_i^B &= E(\beta_i | (s, \mathbf{y}_s)) = \alpha_i \widehat{\beta}_i + (\mathbf{I}_p - \alpha_i) \beta,\end{aligned}$$

$$a_1 = a_0 + \frac{n}{2}, \quad (2.45)$$

$$b_1 = b_0 + \frac{1}{2} \sum_{i=1}^K (y_{s_i} - \mathbf{X}_{s_i} \boldsymbol{\beta})^t (\mathbf{W}_{s_i} + \mathbf{X}_{s_i} \mathbf{C}_\beta \mathbf{X}_{s_i}^t)^{-1} (y_{s_i} - \mathbf{X}_{s_i} \boldsymbol{\beta}).$$

**Demostración.** La demostración de los apartados (a), (d) y (e) son inmediatas a partir del teorema 3.6 (capítulo 3), tomando:  $m = 1$ ,  $\mathbf{Y} = (\mathbf{Y}_1^t, \mathbf{Y}_2^t, \dots, \mathbf{Y}_K^t)^t$ ,  $\mathbf{Y}_i = (Y_{i1}, Y_{i2}, \dots, Y_{iN_i})^t$ ,  $i = 1, 2, \dots, K$ ,  $\boldsymbol{\theta}_1 = (\boldsymbol{\beta}_1^t, \dots, \boldsymbol{\beta}_K^t)^t$ ,  $\mathbf{A}_1 = \text{diag}_{i=1}^K \{\mathbf{X}_i\}$ ,  $\mathbf{X}_i = (\mathbf{x}_{i1}^t, \mathbf{x}_{i2}^t, \dots, \mathbf{x}_{iN_i}^t)^t$ ,  $i = 1, 2, \dots, K$ ,  $\mathbf{x}_{ij} = (x_{ij1}, x_{ij2}, \dots, x_{ijp})$ ,  $j = 1, 2, \dots, N_i$ ,  $i = 1, 2, \dots, K$ ,  $\mathbf{A}_2 = \underbrace{(\mathbf{I}_p, \dots, \mathbf{I}_p)}_K^t \boldsymbol{\beta}$ ,  $\mathbf{W}_1 = \text{diag}_{i=1}^K \{\mathbf{W}_i\}$  y  $\mathbf{W}_2 = \text{diag}_{i=1}^K \{\mathbf{C}_\beta\}$ ,

y denotando por  $\hat{\boldsymbol{\beta}}_i^B$  el  $p$ -vector de  $\hat{\boldsymbol{\theta}}_1^B$  correspondiente a la unidad  $i$ . El resultado (b) se comprueba<sup>4</sup> aplicando el resultado (a) del lema B.9 sobre la distribución de  $\bar{Y}_{u_i} | (s, \mathbf{y}_s), \sigma^2$  que se obtiene directamente de (2.31) tomando  $\mathbf{V}_i = \sigma^2 \mathbf{W}_i$ ,  $i = 1, 2, \dots, K$  y  $\mathbf{B}_\beta = \sigma^2 \mathbf{C}_\beta$  y sobre la distribución (3.79) de la demostración del teorema 3.6 (adaptada a este modelo a través de los cambios señalados en el primer párrafo de esta demostración), y el resultado (c) se obtiene a partir de (b) aplicando la definición de estimador Bayes adaptada a la estimación de la media de la unidad  $i$  con  $i = 1, 2, \dots, K$ .

C.S.Q.D.

Una generalización de este modelo para establecer inferencias sobre una característica continua cuando las unidades son independientes y muy distintas entre sí y la varianza es desconocida consiste en asumir que las variables pendiente de cada unidad son independientes y normales con medias y varianzas distintas, esto es, sustituir en (2.37), las distribuciones iniciales sobre las pendientes para cada unidad por:  $\boldsymbol{\beta}_i \sim N_p(\mathbf{c}_i, \sigma^2 \mathbf{C}_i)$ , independientes para  $i = 1, 2, \dots, K$ , con  $\mathbf{c}_i$  un  $p$ -vector conocido, para  $i = 1, 2, \dots, K$ , y  $\mathbf{C}_i$  una matriz simétrica y definida positiva de dimensión  $p \times p$  conocida, para  $i = 1, 2, \dots, K$ . Este modelo para el que se obtienen, aplicando el teorema 3.6 adecuadamente, los resultados (2.38) a (2.44), sustituyendo  $\mathbf{C}_\beta$  por  $\mathbf{C}_i$  y  $\boldsymbol{\beta}$  por  $\mathbf{c}_i$ , para  $i = 1, 2, \dots, K$ ; tiene el inconveniente del número a parámetros a determinar para asignar las distribuciones iniciales y que las matrices varianzas-covarianzas de  $\mathbf{Y}$  y de  $(\boldsymbol{\beta}_1^t, \boldsymbol{\beta}_2^t, \dots, \boldsymbol{\beta}_K^t)^t$  han de ser proporcionales a una misma variable aleatoria  $\sigma^2$ , tal y como ya sucedía en el modelo Bayes de Regresión en dos etapas con una jerarquía Normal-Gamma.

### 2.3.3 Modelo Bayes de Regresión en dos etapas con una jerarquía con cociente entre varianzas desconocido

El modelo Bayes de Regresión en dos etapas con una jerarquía Normal-Gamma y la generalización del mismo propuesta en la subsección anterior tienen el inconveniente

<sup>4</sup>Una manera alternativa de comprobar el resultado consiste en aplicar sobre la distribución predictiva del vector no observado dada la muestra que la marginal de una  $t$  y la combinación lineal de  $t$  es  $t$

de que la matriz de varianzas-covarianzas de  $\mathbf{Y}$  y de  $(\beta_1^t, \beta_2^t, \dots, \beta_K^t)^t$  han de ser proporcionales a una misma variable aleatoria  $\sigma^2$ . Esta dificultad puede evitarse si estudiamos modelos no conjugados, un ejemplo de estos modelos es el modelo Bayes de Regresión en dos etapas con una jerarquía con cociente entre varianzas desconocido, que definiremos a continuación, que se aplica para inferir sobre una característica continua en poblaciones bietápicas con las características señaladas al principio de la sección y varianzas y cociente entre varianzas asimismo desconocidos.

**Definición 2.8** *Dado el vector poblacional  $\mathbf{Y} = (\mathbf{Y}_1^t, \mathbf{Y}_2^t, \dots, \mathbf{Y}_K^t)^t$  con  $\mathbf{Y}_i = (Y_{i1}, Y_{i2}, \dots, Y_{iN_i})^t$ ,  $i = 1, 2, \dots, K$  para el cual asociada a cada variable  $Y_{ij}$  tenemos una información auxiliar representada por un vector  $\mathbf{x}_{ij} = (x_{ij1}, \dots, x_{ijp})$  conocido,  $j = 1, 2, \dots, N_i$ ,  $i = 1, 2, \dots, K$ , llamaremos modelo Bayesiano de regresión en dos etapas con una jerarquía con cociente entre varianzas desconocido al definido por:*

$$\begin{aligned} \mathbf{Y}_i | \beta_i, \sigma_i^2 &\sim N(\mathbf{X}_i \beta_i, \sigma_i^2 \mathbf{I}_{N_i}), \text{ independientes, } i = 1, 2, \dots, K, \\ \sigma_i^2 &\sim Ga^{-1}(a_0, b_0), \text{ independientes, } i = 1, 2, \dots, K, \\ \beta_i | \sigma_\beta^2 &\sim N_p(\beta, \sigma_\beta^2 \mathbf{C}_\beta), \text{ i.i.d., } i = 1, 2, \dots, K, \\ \sigma_\beta^2 &\sim Ga^{-1}(a_\beta, b_\beta), \end{aligned} \quad (2.46)$$

donde  $\mathbf{X}_i = (\mathbf{x}_{i1}^t, \mathbf{x}_{i2}^t, \dots, \mathbf{x}_{iN_i}^t)^t$ , es una matriz de rango  $p$ , para  $i = 1, 2, \dots, K$ ,  $\beta$  es un  $p$ -vector conocido,  $\mathbf{C}$  una matriz simétrica, definida positiva conocida de dimensión  $p \times p$  y  $a_0, b_0, a_\beta$  y  $b_\beta$  escalares positivos conocidos.

El análisis Bayes de este modelo dada la muestra, refleja la necesidad de la aplicación de técnicas de integración numérica como la aplicación de técnicas de Monte Carlo o basadas en el muestreo Gibbs para la obtención de las distribuciones predictivas de interés dada la muestra (en el caso del Gibbs de una muestra simulada de la misma) y los estimadores Bayes y varianzas dada la muestra.

El siguiente teorema proporciona las distribuciones predictivas condicionadas a  $(\lambda_1, \lambda_2, \dots, \lambda_K)$  y a la muestra y la distribución de  $\lambda_1, \dots, \lambda_K | (s, \mathbf{y}_s)$ , a partir de las cuales se puede obtener por integración numérica, las distribuciones predictivas del vector no observado de la población, media no observada de cada unidad y media no observada, así como las hipótesis suficientes para la existencia de los dos primeros momentos de dichas distribuciones y por lo tanto de los estimadores Bayes de la media de cada unidad y de la media poblacional, así como sus varianzas condicionadas a la muestra, que asimismo pueden calcularse obteniendo los correspondientes momentos respecto a  $\lambda_1, \dots, \lambda_K | (s, \mathbf{y}_s)$  de los que dependen, por integración numérica.

**Teorema 2.7** *Consideremos el modelo Bayes de Regresión en dos etapas con una jerarquía con cociente entre varianzas desconocido dado en (2.46); si definimos por  $\lambda_i = \frac{\sigma_i^2}{\sigma_\beta^2}$ , para  $i = 1, 2, \dots, K$ , entonces se tiene que:*

(a)  $\mathbf{Y}_{u_i} | (s, \mathbf{y}_s), \lambda_1, \lambda_2, \dots, \lambda_K$  se distribuye  $t(N-n)$ -variante con  $2a_1$  grados de libertad con, para  $a_1 > 1$ , media, para  $i = 1, 2, \dots, K$ :

$$E(\mathbf{Y}_{u_i} | (s, \mathbf{y}_s), \lambda_1, \dots, \lambda_K) = \mathbf{X}_{u_i} \hat{\boldsymbol{\beta}}_i^B, \quad (2.47)$$

y matriz de varianzas y covarianzas,

\* para  $i = 1, 2, \dots, K$ ,

$$\text{Var}(\mathbf{Y}_{u_i} | (s, \mathbf{y}_s), \lambda_1, \dots, \lambda_K) = \frac{b_1}{a_1 - 1} \{ \mathbf{I}_{N_i - n_i} \lambda_i + \mathbf{X}_{u_i} (\mathbf{I}_p - \boldsymbol{\alpha}) \mathbf{C}_\beta \mathbf{X}_{u_i}^t \}, \quad (2.48)$$

\* para  $i \neq j, i, j = 1, 2, \dots, K$ ,

$$\text{Cov}(\mathbf{Y}_{u_i}, \mathbf{Y}_{u_j} | (s, \mathbf{y}_s), \lambda_1, \dots, \lambda_K) = \mathbf{0}_{N_i - n_i \times N_j - n_j}. \quad (2.49)$$

(b)  $\bar{Y}_{u_i} | (s, \mathbf{y}_s), \lambda_i$  se distribuye  $t$  con  $2a_1$  grados de libertad y media y varianza, para  $a_1 > 1$ , dadas, para  $i = 1, 2, \dots, K$ , por:

$$\begin{aligned} E(\bar{Y}_{u_i} | (s, \mathbf{y}_s), \lambda_i) &= \frac{1}{N_i - n_i} \mathbf{1}_{N_i - n_i}^t \mathbf{X}_{u_i} \hat{\boldsymbol{\beta}}_i^B, \quad (2.50) \\ \text{Var}(\bar{Y}_{u_i} | (s, \mathbf{y}_s), \lambda_i) &= \frac{1}{(N_i - n_i)^2} \frac{b_1}{a_1 - 1} \{ \mathbf{1}_{N_i - n_i}^t [\mathbf{I}_{N_i - n_i} \lambda_i + \\ &\quad \mathbf{X}_{u_i} (\mathbf{I}_p - \boldsymbol{\alpha}) \mathbf{C}_\beta \mathbf{X}_{u_i}^t] \mathbf{1}_{N_i - n_i} \}, \end{aligned}$$

(c)  $\bar{Y}_{u_i} | (s, \mathbf{y}_s), \lambda_1, \dots, \lambda_K$  se distribuye  $t$  con  $2a_1$  grados de libertad y media y varianza, para  $a_1 > 1$ , dadas por:

$$\begin{aligned} E(\bar{Y}_{u_i} | (s, \mathbf{y}_s), \{\lambda_i\}_{i=1}^K) &= \frac{1}{N - n} \sum_{i=1}^K \mathbf{1}_{N_i - n_i}^t \mathbf{X}_{u_i} \hat{\boldsymbol{\beta}}_i^B, \quad (2.51) \\ \text{Var}(\bar{Y}_{u_i} | (s, \mathbf{y}_s), \{\lambda_i\}_{i=1}^K) &= \frac{1}{(N - n)^2} \frac{b_1}{a_1 - 1} \sum_{i=1}^K [ (N_i - n_i) \lambda_i + \\ &\quad \mathbf{1}_{N_i - n_i}^t \mathbf{X}_{u_i} (\mathbf{I}_p - \boldsymbol{\alpha}) \mathbf{C}_\beta \mathbf{X}_{u_i}^t \mathbf{1}_{N_i - n_i} ], \end{aligned}$$

(d)  $\lambda_1, \dots, \lambda_K | (s, \mathbf{y}_s)$  se distribuye con función de densidad dada por:

$$\pi(\lambda_1, \dots, \lambda_K | (s, \mathbf{y}_s)) \propto \prod_{i=1}^K |\mathbf{I}_{n_i} \lambda_i + \mathbf{X}_{s_i} \mathbf{C}_\beta \mathbf{X}_{s_i}^t|^{-\frac{1}{2}} \prod_{i=1}^K \lambda_i^{-(a_0+1)} b_1^{-a_1}, \quad (2.52)$$

(e) el estimador Bayes para la media poblacional de la unidad  $i$ -ésima,  $i = 1, 2, \dots, K$  y su varianza condicionada a la muestra vienen dados para  $\frac{n}{K} \geq 2$ ,  $a_0 > 1$  por:

$$\begin{aligned}\widehat{Y}_i &= \frac{n_i}{N_i} \bar{y}_{s_i} + \frac{1}{N_i} \mathbf{1}_{N_i-n_i}^t \mathbf{X}_{u_i} E^{\lambda_i|(s, \mathbf{y}_s)} [\widehat{\beta}_i^B], \quad (2.53) \\ \text{Var}(\widehat{Y}_i|(s, \mathbf{y}_s)) &= \frac{1}{N_i^2} \left\{ \mathbf{1}_{N_i-n_i}^t \mathbf{X}_{u_i} \text{Var}^{\lambda_i|(s, \mathbf{y}_s)} [\widehat{\beta}_i^B] \mathbf{X}_{u_i}^t \mathbf{1}_{N_i-n_i} + \right. \\ &\quad \left. + \frac{1}{a_1 - 1} E^{\lambda_i|(s, \mathbf{y}_s)} \{ b_1 [(N_i - n_i) \lambda_i + \right. \\ &\quad \left. \mathbf{1}_{N_i-n_i}^t \mathbf{X}_{u_i} (\mathbf{I}_p - \alpha_i) \mathbf{C}_\beta \mathbf{X}_{u_i}^t \mathbf{1}_{N_i-n_i} \} \right\},\end{aligned}$$

y (f) el estimador Bayes para la media poblacional y su varianza condicionada a la muestra vienen dados para  $\frac{n}{K} \geq 2$ ,  $a_0 > 1$  por:

$$\begin{aligned}\widehat{Y} &= \frac{n}{N} \bar{y}_s + \frac{1}{N} \sum_{i=1}^K \mathbf{1}_{N_i-n_i}^t \mathbf{X}_{u_i} E^{\lambda_i|(s, \mathbf{y}_s)} [\widehat{\beta}_i^B], \quad (2.54) \\ \text{Var}(\widehat{Y}|(s, \mathbf{y}_s)) &= \frac{1}{N^2} \left\{ \mathbf{1}_{N-n}^t \mathbf{X}_u \text{Var}^{\lambda_1, \lambda_2, \dots, \lambda_K|(s, \mathbf{y}_s)} [\widehat{\beta}_B] \mathbf{X}_u^t \mathbf{1}_{N-n} + \right. \\ &\quad \left. + \frac{1}{a_1 - 1} \sum_{i=1}^K E^{\lambda_i|(s, \mathbf{y}_s)} \{ b_1 [(N_i - n_i) \lambda_i + \right. \\ &\quad \left. + \mathbf{1}_{N_i-n_i}^t \mathbf{X}_{u_i} (\mathbf{I}_p - \alpha_i) \mathbf{C}_\beta \mathbf{X}_{u_i}^t \mathbf{1}_{N_i-n_i} \} \right\},\end{aligned}$$

donde  $E^{\lambda_i|(s, \mathbf{y}_s)}[\cdot]$ ,  $\text{Var}^{\lambda_i|(s, \mathbf{y}_s)}[\cdot]$ ,  $E^{\lambda_1, \dots, \lambda_K|(s, \mathbf{y}_s)}[\cdot]$  y  $\text{Var}^{\lambda_1, \dots, \lambda_K|(s, \mathbf{y}_s)}[\cdot]$  son respectivamente, los operadores esperanza y varianza respecto a la distribución de  $\lambda_i|(s, \mathbf{y}_s)$ ,  $i = 1, 2, \dots, K$  (marginal de  $\lambda_1, \dots, \lambda_K|(s, \mathbf{y}_s)$  dada en (2.52)) y respecto a la distribución de  $\lambda_1, \dots, \lambda_K|(s, \mathbf{y}_s)$  dada en (2.52), respectivamente,  $\widehat{\beta}_B = (\widehat{\beta}_1^B, \widehat{\beta}_2^B, \dots, \widehat{\beta}_K^B)^t$ ,  $\bar{y}_s = \frac{\sum_{(i,j) \in s} y_{ij}}{n}$  es la media muestral de la población,  $\bar{y}_{s_i}$  y  $\widehat{\beta}_i = (\mathbf{X}_{s_i}^t \mathbf{V}_{s_i}^{-1} \mathbf{X}_{s_i})^{-1} \mathbf{X}_{s_i}^t \mathbf{V}_{s_i}^{-1} \mathbf{y}_{s_i}$  son, respectivamente la media muestra y el estimador de regresión clásico para la unidad  $i$  para  $i = 1, 2, \dots, K$ , y donde, para  $i = 1, 2, \dots, K$ :

$$\begin{aligned}\alpha_i &= (\mathbf{X}_{s_i}^t \mathbf{X}_{s_i} + \lambda_i \mathbf{C}_\beta^{-1})^{-1} \mathbf{X}_{s_i}^t \mathbf{X}_{s_i}, \\ \widehat{\beta}_i^B &= E(\beta_i|(s, \mathbf{y}_s)) = \alpha_i \widehat{\beta}_i + (\mathbf{I}_p - \alpha_i) \beta, \\ a_1 &= a_\beta + K a_0 + \frac{n}{2}, \quad (2.55) \\ b_1 &= b_\beta + b_0 \sum_{i=1}^K \frac{1}{\lambda_i} + \frac{1}{2} \sum_{i=1}^K (\mathbf{y}_{s_i} - \mathbf{X}_{s_i} \beta)^t (\mathbf{I}_{n_i} \lambda_i + \mathbf{X}_{s_i} \mathbf{C}_\beta \mathbf{X}_{s_i}^t)^{-1} (\mathbf{y}_{s_i} - \mathbf{X}_{s_i} \beta),\end{aligned}$$

**Demostración.** La demostración de este teorema se apoya en el teorema 2.5 y los lemas B.8 y B.9; para su mejor desarrollo la dividiremos en los siguientes pasos:

1.- A partir de los resultados (a), (b) y (d) del teorema 2.5, tomando  $\mathbf{V} = \text{diag}_{i=1}^K \{\lambda_i \sigma_\beta^2 \mathbf{I}_{N_i}\}$  y  $\mathbf{B} = \sigma_\beta^2 \mathbf{C}_\beta$ , se obtienen las distribuciones de  $\mathbf{Y}_u|(s, \mathbf{y}_s), \sigma_\beta^2, \{\lambda_i\}_{i=1}^K, \bar{Y}_u|(s, \mathbf{y}_s), \sigma_\beta^2, \lambda_i, i = 1, 2, \dots, K$ , y  $\bar{Y}_u|(s, \mathbf{y}_s), \sigma_\beta^2, \{\lambda_i\}_{i=1}^K$ , que son:

\*  $\mathbf{Y}_u|(s, \mathbf{y}_s), \sigma_\beta^2, \{\lambda_i\}_{i=1}^K$ , independientes para  $i = 1, 2, \dots, K$ , con distribución normal dada por:

$$\mathbf{Y}_u|(s, \mathbf{y}_s), \sigma_\beta^2, \lambda_i \sim N_{N_i-n_i}(\mathbf{X}_u \hat{\beta}_i^B, \sigma_\beta^2 [\mathbf{I}_{N_i-n_i} \lambda_i + \mathbf{X}_u (\mathbf{I}_p - \alpha) \mathbf{C}_\beta \mathbf{X}_u^t]), \quad (2.56)$$

\*  $\bar{Y}_u|(s, \mathbf{y}_s), \sigma_\beta^2, \lambda_i$ , independientes para  $i = 1, 2, \dots, K$ , con distribución normal dada por:

$$\bar{Y}_u|(s, \mathbf{y}_s), \sigma_\beta^2, \lambda_i \sim N\left(\frac{1}{N_i - n_i} \mathbf{1}_{N_i-n_i}^t \mathbf{X}_u \hat{\beta}_i^B, \frac{\sigma_\beta^2 (a_1 - 1)}{b_1} \text{Var}(\bar{Y}_u|(s, \mathbf{y}_s), \lambda_i)\right), \quad (2.57)$$

\*  $\bar{Y}_u|(s, \mathbf{y}_s), \sigma_\beta^2, \{\lambda_i\}_{i=1}^K$ , se distribuye según una normal dada por:

$$\bar{Y}_u|(s, \mathbf{y}_s), \sigma_\beta^2, \{\lambda_i\}_{i=1}^K \sim N\left(\frac{1}{N - n} \mathbf{1}_{N-n}^t \mathbf{X}_u \hat{\beta}_B, \frac{\sigma_\beta^2 (a_1 - 1)}{b_1} \text{Var}(\bar{Y}_u|(s, \mathbf{y}_s), \{\lambda_i\}_{i=1}^K)\right), \quad (2.58)$$

2.- A partir de las distribuciones iniciales de  $\mathbf{Y}|\{\beta_i\}_{i=1}^K, \{\sigma_i^2\}_{i=1}^K$  y de  $\beta_i|\sigma_\beta^2, i = 1, 2, \dots, K$ , se obtiene aplicando, sobre la primera las propiedades de las distribuciones marginales de una normal y después, sobre el resultado y las segundas, el apartado (a) del lema B.8, la distribución de  $\mathbf{Y}_s|\sigma_\beta^2, \{\lambda_i\}_{i=1}^K$ , que es normal con  $\mathbf{Y}_s|\sigma_\beta^2, \{\lambda_i\}_{i=1}^K$ , independientes, dadas para  $i = 1, 2, \dots, K$  por:

$$\mathbf{Y}_s|\sigma_\beta^2, \{\lambda_i\}_{i=1}^K \sim N_{n_i}(\mathbf{X}_s \beta, \sigma_\beta^2 [\mathbf{I}_{n_i} \lambda_i + \mathbf{X}_s \mathbf{C}_\beta \mathbf{X}_s^t]). \quad (2.59)$$

3.- A partir de las distribuciones iniciales de  $\{\sigma_i^2\}_{i=1}^K$  y  $\sigma_\beta^2$  se obtiene, aplicando el método de obtención de una f.d.p por el método directo (para una transformación 1 a 1), la distribución inicial conjunta de  $(\sigma_\beta^2, \lambda_1, \lambda_2, \dots, \lambda_K)$ , que es:

$$\pi(\sigma_\beta^2, \lambda_1, \lambda_2, \dots, \lambda_K) \propto [\sigma_\beta^2]^{-(K a_0 + a_\beta + 1)} \exp\left\{-\frac{1}{\sigma_\beta^2} (b_\beta + b_0 \sum_{i=1}^K \frac{1}{\lambda_i})\right\} \prod_{i=1}^K \lambda_i^{-(a_0 + 1)}. \quad (2.60)$$

4.- A partir de (2.59) y (2.60), aplicando el teorema de Bayes e identificando los términos se obtienen las distribuciones de  $\sigma_\beta^2 | \{\lambda_i\}_{i=1}^K, (s, \mathbf{y}_s)$ , que viene dada por:

$$\sigma_\beta^2 | \{\lambda_i\}_{i=1}^K, (s, \mathbf{y}_s) \sim Ga^{-1}(a_1, b_1), \quad (2.61)$$

y de  $\lambda_1, \lambda_2, \dots, \lambda_K | (s, \mathbf{y}_s)$  dada en (2.52).

5.- A partir de (2.56) y (2.61) aplicando el apartado (a) del lema B.9 se obtiene la distribución de  $\mathbf{Y}_u | (s, \mathbf{y}_s), \{\lambda_i\}_{i=1}^K$  dada en (2.47), (2.48) y (2.49).

6.- A partir de (2.57) - (2.61), y (2.58) - (2.61), aplicando el apartado (a) del lema B.9 se obtienen respectivamente, la distribución de  $\bar{Y}_{u_i} | (s, \mathbf{y}_s), \{\lambda_i\}_{i=1}^K, i = 1, 2, \dots, K$ , dada en (2.50) y de  $\bar{Y}_u | (s, \mathbf{y}_s), \{\lambda_i\}_{i=1}^K$ , dada en (2.51).

7.- A partir de (2.50) y (2.52), se obtiene, aplicando la definición de estimador Bayes, el lema B.4, y que:

$$\begin{aligned} E(\bar{Y}_{u_i} | (s, \mathbf{y}_s)) &= E^{\lambda_i | (s, \mathbf{y}_s)} E(\bar{Y}_{u_i} | (s, \mathbf{y}_s), \lambda_i), \\ Var(\bar{Y}_{u_i} | (s, \mathbf{y}_s)) &= E^{\lambda_i | (s, \mathbf{y}_s)} Var(\bar{Y}_{u_i} | (s, \mathbf{y}_s), \lambda_i) + Var^{\lambda_i | (s, \mathbf{y}_s)} E(\bar{Y}_{u_i} | (s, \mathbf{y}_s), \lambda_i), \end{aligned}$$

el resultado (e), dado en (2.53), de este teorema.

8.- A partir de (2.51) y (2.52), se obtiene de forma totalmente análoga al apartado 7 de esta demostración, el resultado (f), dado en (2.54), de este teorema.

C.S.Q.D.

El teorema 2.7 permite la obtención de los resultados inferenciales de interés pero requiere del cálculo de algunas integrales multivariantes que en principio pueden ser complicadas por la forma de la distribución de  $\lambda_1, \lambda_2, \dots, \lambda_K | (s, \mathbf{y}_s)$ , dada en (2.52), (que no es producto de marginales independientes del tipo obtenido en el teorema 2.3 para el correspondiente modelo de una etapa). Un procedimiento alternativo de análisis y obtención de las distribuciones predictivas y estimadores Bayes es la aplicación de los procedimientos de muestreo Gibbs (ver sección C.1 del apéndice C). Las distribuciones condicionales necesarias para la aplicación del muestreo Gibbs vienen dadas en el siguiente teorema:

**Teorema 2.8** *El modelo Bayes de Regresión en dos etapas con una jerarquía con cociente entre varianzas desconocido dado en (2.46), es un modelo DAG (figura 2.2) con distribuciones condicionales dadas por:*

$$\begin{aligned} f((\mathbf{y}_{u_1}^t, \dots, \mathbf{y}_{u_K}^t)^t | \mathbf{y}_s, \{\beta_i\}_{i=1}^K, \{\sigma_i^2\}_{i=1}^K, \sigma_\beta^2) = \\ f((\mathbf{y}_{u_1}^t, \dots, \mathbf{y}_{u_K}^t)^t | \{\beta_i\}_{i=1}^K, \{\sigma_i^2\}_{i=1}^K) = \end{aligned}$$

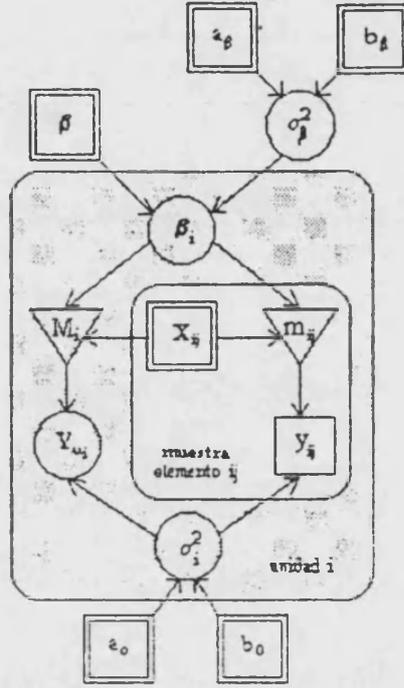


Figura 2.2: Representación gráfica del modelo Bayes de Regresión en dos etapas con una jerarquía con cociente entre varianzas desconocido

$$\prod_{i=1}^K N_{N_i, -n_i}(y_{u_i} | X_{u_i}, \beta_i, \sigma_i^2 \mathbf{I}_{N_i, -n_i}), \quad (2.62)$$

$$\pi((\beta_1^t, \dots, \beta_K^t)^t | y, \{\sigma_i^2\}_{i=1}^K, \sigma_\beta^2) =$$

$$\prod_{i=1}^K N_p(\beta_i | \alpha_i^* \hat{\beta}_i^* + (\mathbf{I}_p - \alpha_i^*) \beta_i, (\mathbf{I}_p - \alpha_i^*) \mathbf{C}_\beta \sigma_\beta^2), \quad (2.63)$$

$$\pi(\sigma_\beta^2 | y, \{\beta_i\}_{i=1}^K, \{\sigma_i^2\}_{i=1}^K) = \pi(\sigma_\beta^2 | \{\beta_i\}_{i=1}^K) =$$

$$Ga^{-1}(\sigma_\beta^2 | a_\beta^*, b_\beta^*), \quad (2.64)$$

$$\pi(\sigma_1^2, \dots, \sigma_K^2)^t | y, \{\beta_i\}_{i=1}^K, \sigma_\beta^2) = \pi(\sigma_1^2, \dots, \sigma_K^2)^t | y, \{\beta_i\}_{i=1}^K) =$$

$$\prod_{i=1}^K Ga^{-1}(\sigma_i^2 | a_i^*, b_i^*), \quad (2.65)$$

$$(2.66)$$

donde:

$$\hat{\beta}_i^* = (\mathbf{X}_i^t \mathbf{X}_i)^{-1} \mathbf{X}_i^t \mathbf{y}_i, \quad i = 1, 2, \dots, K \quad (2.67)$$

$$\alpha_i^* = (\mathbf{X}_i^t \mathbf{X}_i + \mathbf{C}_\beta^{-1} \frac{\sigma_i^2}{\sigma_\beta^2})^{-1} \mathbf{X}_i^t \mathbf{X}_i \quad i = 1, 2, \dots, K,$$

$$a_i^* = \frac{N_i}{2} + a_0, \quad i = 1, 2, \dots, K,$$

$$b_i^* = \frac{1}{2} (\mathbf{y}_i - \mathbf{X}_i \beta_i)^t (\mathbf{y}_i - \mathbf{X}_i \beta_i) + b_0, \quad i = 1, 2, \dots, K,$$

$$\begin{aligned}
a_{\beta}^* &= \frac{Kp}{2} + a_{\beta}, \\
b_{\beta}^* &= \frac{1}{2} \sum_{i=1}^K (\beta_i - \beta)^t C_{\beta}^{-1} (\beta_i - \beta) + b_{\beta}.
\end{aligned}$$

**Demostración.** De la definición 2.8 del modelo en estudio se deduce directamente la representación gráfica de la figura 2.2, que corresponde a un modelo DAG. Los resultados de las distribuciones condicionales se obtienen a partir de la distribución conjunta de de  $(\mathbf{Y}_{u_1}^t, \dots, \mathbf{Y}_{u_K}^t, \mathbf{Y}_{s_1}^t, \dots, \mathbf{Y}_{s_K}^t, \beta_1^t, \dots, \beta_K^t, \sigma_1^2, \dots, \sigma_K^2, \sigma_{\beta}^2)^t$ , que viene dada por:

$$\begin{aligned}
f((\mathbf{y}_{u_1}^t, \dots, \mathbf{y}_{u_K}^t, \mathbf{y}_{s_1}^t, \dots, \mathbf{y}_{s_K}^t, \beta_1^t, \dots, \beta_K^t, \sigma_1^2, \dots, \sigma_K^2, \sigma_{\beta}^2)^t) = \\
\prod_{i=1}^K N(\mathbf{y}_{u_i} | \mathbf{X}_{u_i}, \beta_i, \sigma_i^2 \mathbf{I}_{N_i - n_i}) \times \prod_{i=1}^K N(\mathbf{y}_{s_i} | \mathbf{X}_{s_i}, \beta_i, \sigma_i^2 \mathbf{I}_{n_i}) \times \prod_{i=1}^K N(\beta_i | \beta, \sigma_{\beta}^2 C_{\beta}) \times \\
\prod_{i=1}^K Ga^{-1}(\sigma_i^2 | a_0, b_0) \times Ga^{-1}(\sigma_{\beta}^2 | a_{\beta}, b_{\beta}),
\end{aligned}$$

separando los términos correspondientes a cada condicional (que contienen la variable cuya condicional estamos hallando), de manera análoga a como lo hacíamos en el teorema 2.4.

C.S.Q.D.

Notemos que las distribuciones condicionales necesarias para la simulación de las distribuciones finales dada la muestra son distribuciones estándar conocidas y por lo tanto, para las que existen algoritmos suficientemente probados para su simulación. Además, el modelo descrito es DAG y cumple las hipótesis de independencia inicial de todas las variables necesarias, por lo que puede aplicarse, nuevamente el programa BUGS (ver subsección C.2.1 del apéndice C).

### 2.3.4 Modelos Bayes en dos etapas con una jerarquía mínimo informativos

Para concluir el estudio de los modelos Bayes de Regresión en dos etapas con una jerarquía, incluimos a continuación la definición y resultados inferenciales habituales dada la muestra de dos modelos mínimo informativos: el modelo Bayes de Regresión en dos etapas con una jerarquía normal mínimo informativo y el modelo Bayes de Regresión en dos etapas con una jerarquía normal-gamma mínimo informativo; estos modelos se aplican para obtener inferencias sobre una característica continua de una población biotápica con las características comentadas al principio de la sección cuando no deseamos introducir información en la asignación de las distribuciones iniciales sobre los parámetros, en las situaciones respectivas de varianza de la población conocida y desconocida. En estos modelos, la condición sobre la muestra  $k = K$ , alcanza una

mayor significación que en el resto de la sección, ya que es condición para que, con las mínimo informativas que estamos empleando, las distribuciones predictivas sean propias.

### Modelo Bayes de Regresión en dos etapas con una jerarquía Normal mínimo informativo

**Definición 2.9** Llamaremos modelo Bayes de Regresión en dos etapas con una jerarquía Normal mínimo informativo a un modelo Bayes de Regresión en dos etapas con una jerarquía Normal en el que la distribución que asumimos sobre  $\beta_i$  es una mínimo informativa para  $i = 1, 2, \dots, K$  de la forma  $\pi(\beta_i) \propto 1$ ,  $i = 1, \dots, K$ .

**Corolario 2.3** Consideremos el modelo Bayes de Regresión en dos etapas con una jerarquía Normal mínimo informativo dado en la definición 2.9, entonces se tiene que las distribuciones predictivas para el vector poblacional no observado, la media no observada para cada unidad y la media poblacional no observada, y los estimadores de la media poblacional de cada unidad y de la media poblacional y su varianza condicionada a la muestra son los obtenidos de (2.30) a (2.34), tomando  $\alpha_i = \mathbf{I}_p$ ,  $i = 1, 2, \dots, K$ .

**Demostración.** La obtención de las distribuciones predictiva de  $\mathbf{Y}_u|(s, \mathbf{y}_s)$ , y de  $\bar{Y}_u|(s, \mathbf{y}_s)$ , así como del estimador de la media poblacional y su varianza condicionada a la muestra es inmediata a partir del corolario 3.3 (capítulo 3), tomando:  $m = 1$ ,  $\mathbf{Y} = (\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_K)^t$ ,  $\mathbf{Y}_i = (Y_{i1}, Y_{i2}, \dots, Y_{iN_i})^t$ ,  $i = 1, 2, \dots, K$ ,  $\theta_1 = (\beta_1^t, \dots, \beta_K^t)^t$ ,  $\mathbf{A}_1 = \text{diag}_{i=1}^K \{\mathbf{X}_i\}$ ,  $\mathbf{X}_i = (\mathbf{x}_{i1}^t, \mathbf{x}_{i2}^t, \dots, \mathbf{x}_{iN_i}^t)^t$ ,  $i = 1, 2, \dots, K$ ,  $\mathbf{x}_{ij} = (x_{ij1}, x_{ij2}, \dots, x_{ijp})$ ,  $j = 1, 2, \dots, N_i$ ,  $i = 1, 2, \dots, K$ ,  $\mathbf{A}_2 = \underbrace{(\mathbf{I}_p, \dots, \mathbf{I}_p)}_K \beta$ ,  $\mathbf{V}_1 =$

$\text{diag}_{i=1}^K \{\mathbf{V}_i\}$ , y denotando por  $\hat{\beta}_i^B$  el  $p$ -vector de  $\hat{\theta}_1^B$  correspondiente a la unidad  $i$ . Los restantes resultados son inmediatos a partir de la distribución predictiva del vector poblacional no observado dada la muestra aplicando que la marginal de una normal y la combinación lineal de normales es normal y la definición de estimador Bayes adaptada a la estimación de la media de la unidad  $i$  con  $i = 1, 2, \dots, K$ .

C.S.Q.D.

### Modelo Bayes de Regresión en dos etapas con una jerarquía Normal-Gamma mínimo informativo

**Definición 2.10** Llamaremos modelo Bayes de Regresión en dos etapas con una jerarquía Normal-Gamma mínimo informativo a un modelo Bayes de Regresión en dos etapas con una jerarquía Normal-Gamma en el que la distribución que asumimos sobre  $(\beta_1, \beta_2, \dots, \beta_K, \sigma^2)$  es una mínimo informativa<sup>5</sup> de la forma  $\pi(\beta_1, \beta_2, \dots, \beta_K, \sigma^2) \propto \frac{1}{\sigma^2}$ .

<sup>5</sup>ver nota 7 a pie de página de la subsección 3.4.3



**Corolario 2.4** Consideremos el modelo Bayesiano de regresión en dos etapas con una jerarquía normal-gamma mínimo informativo dado en la definición 2.10, entonces se tiene que las distribuciones predictivas para el vector poblacional no observado, la media no observada para cada unidad y la media poblacional no observada, y los estimadores de la media poblacional de cada unidad y de la media poblacional y su varianza condicionada a la muestra son los obtenidos de (2.38) a (2.44), tomando:

$$\begin{aligned} \alpha_i &= \mathbf{I}_p, \quad i = 1, 2, \dots, K, \\ a_1 &= \frac{n - Kp}{2}, \\ b_1 &= \frac{1}{2} \sum_{i=1}^K \mathbf{y}_{s_i}^t [\mathbf{W}_{s_i}^{-1} + \mathbf{W}_{s_i}^{-1} \mathbf{X}_{s_i} (\mathbf{X}_{s_i}^t \mathbf{W}_{s_i}^{-1} \mathbf{X}_{s_i})^{-1} \mathbf{X}_{s_i}^t \mathbf{W}_{s_i}^{-1}] \mathbf{y}_{s_i}. \end{aligned} \quad (2.68)$$

**Demostración.** La obtención de las distribuciones predictiva de  $\mathbf{Y}_u | (s, \mathbf{y}_s)$ , y de  $\bar{Y}_u | (s, \mathbf{y}_s)$ , así como del estimador de la media poblacional y su varianza condicionada a la muestra es inmediata a partir del corolario 3.4 (capítulo 3), tomando  $m = 1$ ,  $\mathbf{Y} = (\mathbf{Y}_1^t, \mathbf{Y}_2^t, \dots, \mathbf{Y}_K^t)^t$ ,  $\mathbf{Y}_i = (Y_{i1}, Y_{i2}, \dots, Y_{iN_i})^t$ ,  $i = 1, 2, \dots, K$ ,  $\theta_1 = (\beta_1^t, \dots, \beta_K^t)^t$ ,  $\mathbf{A}_1 = \text{diag}_{i=1}^K \{\mathbf{X}_i\}$ ,  $\mathbf{X}_i = (\mathbf{x}_{i1}^t, \mathbf{x}_{i2}^t, \dots, \mathbf{x}_{iN_i}^t)^t$ ,  $i = 1, 2, \dots, K$ ,  $\mathbf{x}_{ij} = (x_{ij1}, x_{ij2}, \dots, x_{ijp})$ ,  $j = 1, 2, \dots, N_i$ ,  $i = 1, 2, \dots, K$ ,  $\mathbf{A}_2 = (\mathbf{I}_p, \dots, \mathbf{I}_p)$   $\beta$  y  $\mathbf{W}_1 = \text{diag}_{i=1}^K \{\mathbf{W}_i\}$ , y denotando por  $\hat{\beta}_i^B$  el  $p$ -vector de  $\hat{\theta}_1^B$  correspondiente a la unidad  $i$ . La distribución predictiva de  $\bar{Y}_u | (s, \mathbf{y}_s)$ ,  $i = 1, 2, \dots, K$ , se obtiene<sup>6</sup> aplicando el resultado (a) del lema B.9 sobre la distribución de  $\bar{Y}_u | (s, \mathbf{y}_s)$ ,  $\sigma^2$  que se obtiene directamente de (2.31) tomando  $\alpha_i = \mathbf{I}_p$ ,  $i = 1, 2, \dots, K$  y  $\mathbf{V}_i = \sigma^2 \mathbf{W}_i$ ,  $i = 1, 2, \dots, K$  y sobre la distribución (3.84) de la demostración del corolario 3.4, y el estimador Bayes de la media poblacional de cada unidad y su varianza condicionada a la muestra se obtienen a partir de la distribución predictiva de la media no observada de cada unidad, aplicando la definición de estimador Bayes adaptada a la estimación de la media de cada unidad.

C.S.Q.D.

A partir de los resultados inferenciales para el modelo Bayes de Regresión en dos etapas con una jerarquía normal mínimo informativo se recuperan los estimadores Bayes para el correspondiente modelo clásico con matriz de varianzas-covarianzas conocido; y para el modelo Bayes de Regresión en dos etapas con una jerarquía normal-gamma, los estimadores para el correspondiente modelo clásico con matriz de varianzas-covarianzas desconocidas (ver apartado estimadores Bayes de la media poblacional de la subsección 1.4.2).

<sup>6</sup>Una manera alternativa de comprobar el resultado consiste en aplicar sobre la distribución predictiva del vector no observado que la marginal de una  $t$  y la combinación lineal de  $t$  es  $t$

## 2.4 Ejemplos para modelos Bayesianos en una jerarquía

Esta sección se dedicará a complementar los estudios sobre modelos, realizados en las dos secciones anteriores mediante su aplicación en la modelización de una población. Para ello, apoyándonos en los datos de Mosteller y Tukey(77)[75][pág 549-551], sobre la fertilidad en Suiza por los años 1888, hemos considerado la siguiente población, que llamaremos población 1, y describiremos a continuación.

### Población 1

La población 1, la asumiremos constituida por los datos sobre fertilidad de todas las “provincias” de Suiza por los años 1888; para analizar esta característica se dispone de los datos de 5 indicadores socioeconómicos: 1) porcentaje de la población de la provincia que trabaja en el sector agrario, 2) porcentaje de la provincia con nivel de educación superior a los estudios primarios, 3) porcentaje de la población de la provincia que es católica, 4) porcentaje de nacidos que viven menos de un año (mortalidad infantil) y 5) porcentaje de jóvenes de la provincia en edad de servicio militar que reciben la máxima calificación en el examen militar. Para el estudio de esta población disponemos de los datos de fertilidad y de los 5 indicadores socioeconómicos de 47 provincias de lengua francesa ( $n = 47$ ).

Para adaptar esta población a nuestro estudio, a partir de estos datos hemos considerado las siguientes dos poblaciones: población 1A y población 1B. Estas dos poblaciones consisten en la población 1 eliminando la variable explicativa 5) sobre los resultados de los jóvenes de cada provincia en el examen militar, diferenciándose entre sí, en la no división y división, respectivamente, de la población en estratos según los resultados de los jóvenes en el examen militar (la variable socioeconómica excluida en ambas poblaciones). A los datos facilitados por el banco de estas dos poblaciones les hemos añadido los restantes datos necesarios para nuestro análisis, concretamente:

- La población 1A, está constituida por  $N = 250$  provincias, de las que se muestrean  $n = 47$ ; el vector de medias de las características socioeconómicas de las provincias no muestreadas viene dado por  $\bar{X}_u = (22, 27, 24, 19)$ .
- La población 1B, se obtiene a partir de la división de la población 1 en 3 estratos, según el indicador socioeconómico sobre los resultados de las quintas de cada provincia en el examen militar. Los estratos quedan constituidos de la siguiente manera: Estrato 1) provincias con porcentajes entre 0 % y 30 % (nivel bajo), Estrato 2) provincias con porcentajes entre 30 % y 70 % (nivel medio) y Estrato 3) provincias con porcentajes entre 70 % y 100 % (nivel alto). Otros datos adicionales son:
  - \* Estrato 1:  $N_1 = 176$ ,  $n_1 = 33$ ,  $\bar{X}_{u_1} = (9.73, 11.63, 11.15, 7.77)$ .
  - \* Estrato 2:  $N_2 = 37$ ,  $n_2 = 7$ ,  $\bar{X}_{u_2} = (44.73, 40.63, 29.13, 43.77)$ .
  - \* Estrato 3:  $N_3 = 37$ ,  $n_3 = 7$ ,  $\bar{X}_{u_3} = (57.73, 86.63, 80.13, 47.77)$ .

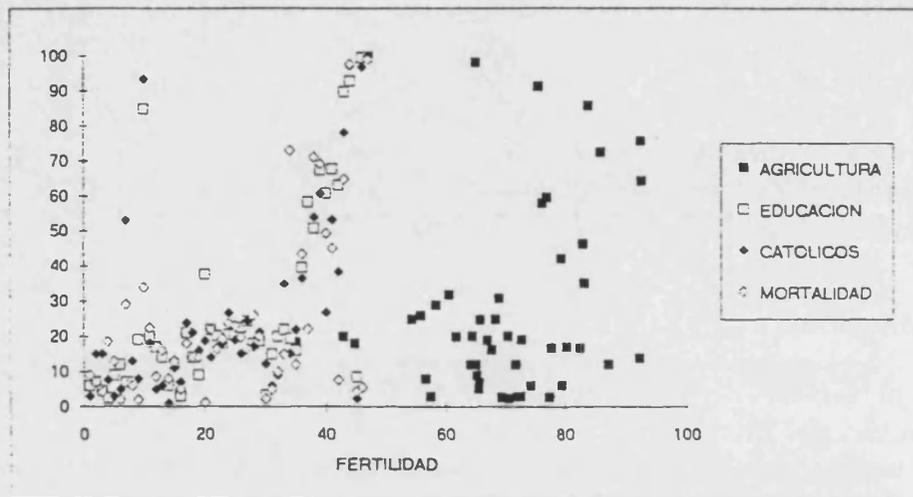


Figura 2.3: Relación entre la variable fertilidad y los indicadores socio-económicos para la Población 1A

(Nótese que se han muestreado todos los estratos,  $k = K$ .)

Para un mejor estudio previo de las dos poblaciones hemos construido a partir de los datos muestrales, las gráficas de las figuras 2.3 y 2.4, que representan para la población considerada en su globalidad (población 1A) y la población estructurada en estratos (población 1B), respectivamente los valores de la característica en estudio, fertilidad (en el eje de abscisas), en relación a los valores de las cuatro variables socioeconómicas del estudio. Notemos, a partir de la figura 2.3, la existencia para cada una de las variables socioeconómicas de un número considerable de observaciones que se separan de la nube de puntos mayoritaria (esto es, desviaciones en la hipótesis de una relación lineal entre la variable respuesta y las explicativas), que disminuirá la precisión de nuestras inferencias respecto a la media poblacional en la población 1A, al alterar los componentes de la pendiente de regresión estimada. Estas desviaciones en la hipótesis de linealidad se reducen algo al realizar la estratificación de la población antes propuesta (población 1B) sobre todo en el estrato 1, aunque siguen detectándose desviaciones considerables en los estratos 2 y 3 en relación a la covariable agricultura, tal y como muestra la figura 2.4.

A continuación, describiremos los modelos aplicados para realizar inferencias sobre la población 1A (modelos en una etapa) y la población 1B, cuando no establecemos una relación entre los estratos por medio de una jerarquía más en el modelo (modelos en dos etapas con una jerarquía). En la sección 3.3 del próximo capítulo se analizará la población 1B cuando sí se establece una relación de intercambiabilidad entre los estratos (modelos en dos etapas con dos jerarquías).

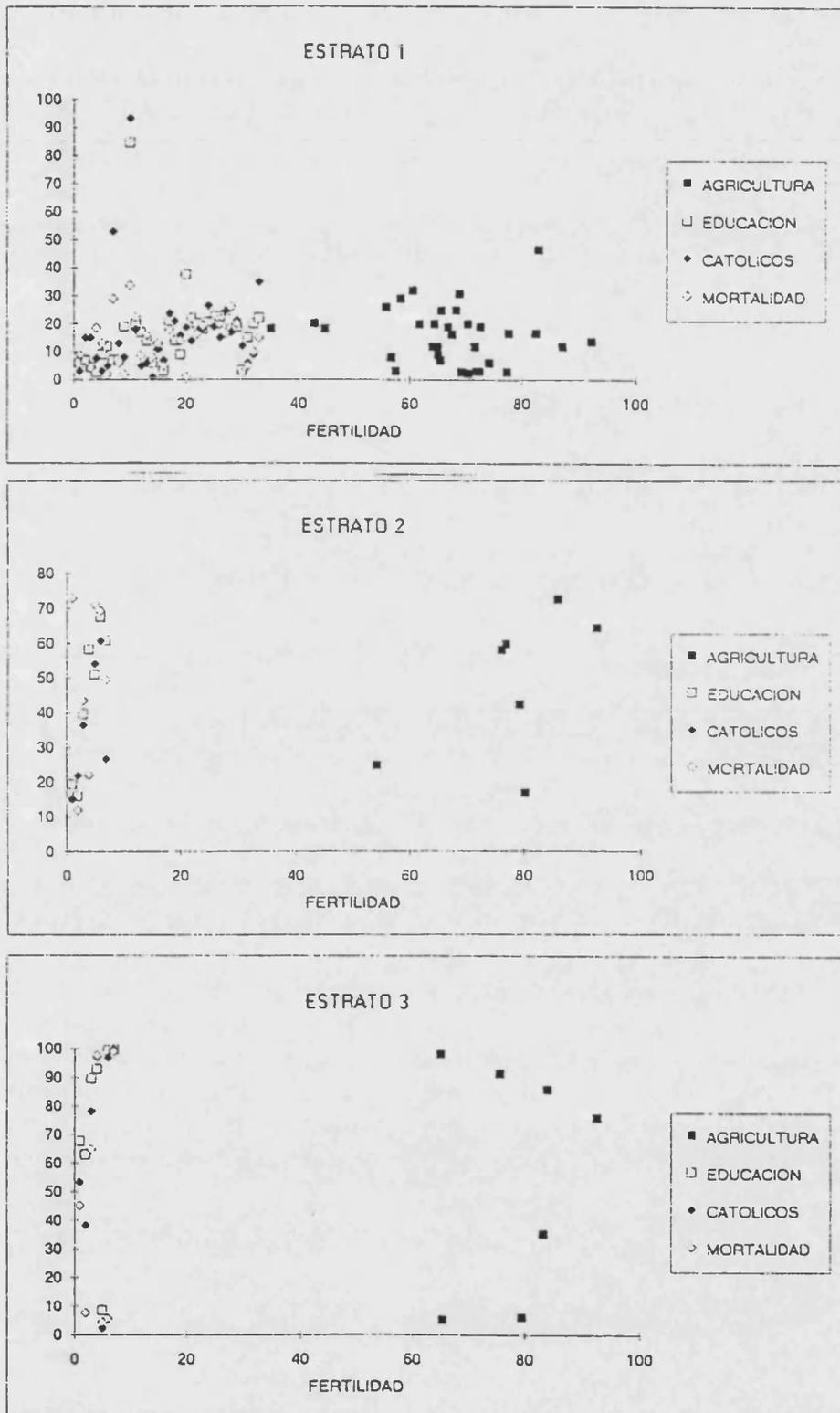


Figura 2.4: Relación entre la variable fertilidad y los indicadores socio-económicos para cada estrato de la Población 1B

### 2.4.1 Población 1A: modelos Bayesianos en una etapa

Para realizar inferencias sobre la media poblacional de la población 1A, emplearemos los modelos Bayesianos en una etapa descritos en la sección 2.2.

Como paso preliminar, para adecuar la característica en estudio a la hipótesis de normalidad, hemos realizado la transformación de todos los porcentajes  $p$  (tanto del vector de características, como de los datos de las variables auxiliares socioeconómicas), según la siguiente fórmula:  $\log\left(\frac{p}{100-p}\right)$ . Para el estudio de esta población transformada, se aplicaron 5 modelos:

1. El modelo Bayes de Regresión Normal, con parámetros  $\mathbf{V} = 0.45\mathbf{I}_{250}$ ,  $\mathbf{B} = 4\mathbf{I}_4$  y  $\mathbf{b} = (-1, 2, -2, 3)^t$ .
2. El modelo Bayes de Regresión Normal-Gamma, con parámetros  $\mathbf{W} = \mathbf{I}_{250}$ ,  $\mathbf{C} = \mathbf{I}_4$ ,  $\mathbf{b} = (-1, 2, -2, 3)^t$ ,  $a_0 = 3$  y  $b_0 = 8$  ( $E(\sigma^2) = 4$ ,  $Var(\sigma^2) = 16$ ).
3. El modelo Bayes de Regresión con cociente entre varianzas desconocido, con parámetros:  $\mathbf{W} = \mathbf{I}_{250}$ ,  $\mathbf{C} = \mathbf{I}_4$ ,  $\mathbf{b} = (-1, 2, -2, 3)^t$ ,  $a_0 = 3$  y  $b_0 = 2$  ( $E(\sigma^2) = 1$ ,  $Var(\sigma^2) = 1$ ),  $a_\beta = 3$  y  $b_\beta = 8$  ( $E(\sigma_\beta^2) = 4$ ,  $Var(\sigma_\beta^2) = 16$ ).

Para implementar este modelo se recurrió al programa BUGS. El análisis con el programa BUGS constó de tres fases: la traducción en BUGS del modelo (figura 2.5), el estudio de la convergencia del muestreo Gibbs y una vez asumida la estabilidad, la simulación de 1000 observaciones para cada una de las variables del problema (simulaciones de las distribuciones predictivas y finales). En el proceso de simulación por Gibbs se realizaron 1000 primeras simulaciones para obtener un buen punto de arranque para el algoritmo de Gibbs, y a partir de dicho punto de arranque se realizaron 5000 simulaciones tras las que se consideró alcanzada la estabilidad. Esta estabilidad se valoró a través del estadístico aproximado de Geweke(92)[39] (obtenido por el comando `diag` del BUGS) de las variables  $\bar{Y}_u$ ,  $\beta_1$ ,  $\sigma^2$  y  $\sigma_\beta^2$ , que tomaba valores inferiores en módulo a 1.96. A partir de la simulación del 1000 observaciones más, una vez asumida la estabilidad, se obtuvieron (con el comando `stats` del BUGS), las aproximaciones a la media y varianza de las distribuciones predictivas y finales (que sabemos que existen por el teorema 2.3), y una representación gráfica de las densidades de estas distribuciones predictivas y finales (con el programa de Splus, `drawdat`, incluido en el BUGS). (Para más detalles, sobre cómo se ha fundamentado el análisis a través del programa BUGS, ver la sección C.3 del apéndice C.)

4. El modelo Bayes de Regresión Normal mínimo informativo, con parámetro  $\mathbf{V} = 0.45\mathbf{I}_{250}$ .
5. El modelo Bayes de Regresión Normal-Gamma mínimo informativo, con parámetro  $\mathbf{W} = \mathbf{I}_{250}$ .

A partir de los datos de la muestra (los datos correspondientes a las 47 provincias de lengua francesa) y los modelos anteriores se obtuvieron, aplicando los teoremas 2.1, 2.2, para los modelos normal y normal-gamma, el programa BUGS (apoyado en los teoremas 2.3 y 2.4), para el modelo con cociente entre varianzas desconocido, y los

Modelo Bayes con cociente entre varianzas desconocido

```

model{
  const
  N=250, n=47, p=4
  var
  y[n], Y.mdu, Y.med, x[n,p], X.mdu[p],
  beta[p], b[p], sigma2, sigma2.b, m[n],
  M, tau, tau.b, tau.m, lambda
  data in "file.dat",
  {
    for(i in 1:n){
      y[i]~dnorm(m[i],tau)
      m[i]~-x[i,1]*beta[1]-x[i,2]*beta[2]-
        x[i,3]*beta[3]-x[i,4]*beta[4]
    }
    for(i in 1:p){
      beta[i]~dnorm(b[i],tau.b)
    }
    Y.mdu~dnorm(M,tau.m)
    tau.m<-(N-n)*tau
    M<-X.mdu[1]*beta[1]-X.mdu[2]*beta[2]-
      X.mdu[3]*beta[3]-X.mdu[4]*beta[4]
    tau~dgamma(3.0,2.0)
    sigma2<-1/tau
    tau.b~dgamma(3.0,3.0)
    sigma2.b<-1/tau.b
    lambda<-sigma2/sigma2.b
    Y.med<-sum(y[,])N-(N-n)*Y.mdu/N
  }
}

```

"file.data" (contenido)  
 list(y=c(-datos de la muestra separados por comas),  
 x=c(-datos variable explicativa de la muestra separados  
 por comas-1, b=c(-1,2,-2,3), X.mdu=c(-vector medias  
 no observadas-1))

Figura 2.5: Traducción en BUGS del modelo Bayes de Regresión con cociente entre varianzas desconocido

	MODELO NORMAL	MODELO NORM-GAMMA	MODELO COCIENTE VAR. DESCONOCIDO	MODELO NORMAL MIN. INFOR.	MODELO NORM-GAMMA MIN. INFOR.
$\bar{Y}$	0.5696 (63.87)	0.5687 (63.85)	-0.3347 (41.71)	0.5696 (63.87)	0.5696 (63.87)
$\text{Var}(\bar{Y}   s, y_s)$	0.0051	0.0163	0.0668	0.0051	0.0227

Las estimaciones de la media poblacional en porcentajes figuran en la tabla entre parentesis

Tabla 2.1: Estimadores Bayes de la media poblacional y su varianza para los modelos Bayesianos en una etapa

MOD. NORMAL	$\bar{Y}_u(s, y_s) \sim N(0.4844, 0.0077)$
MOD. NORM-GAMMA	$\bar{Y}_u(s, y_s) \sim U(53.0, 4832.0, 0.0237)$
MOD. NORMAL MIN. INF.	$\bar{Y}_u(s, y_s) \sim N(0.4844, 0.0077)$
MOD. NORM-GAMMA MIN. INF.	$\bar{Y}_u(s, y_s) \sim U(43.0, 4844.0, 0.0329)$

Tabla 2.2: Distribuciones predictivas de la media poblacional no observada para los modelos Bayesianos en una etapa conjugados

corolarios 2.1 y 2.2, para los modelos mínimo informativos: los estimadores Bayes de la media poblacional y su varianza condicionada a la muestra, para los cinco modelos (tabla 2.1), las distribuciones predictivas de la media poblacional no observada dada la muestra, para los modelos conjugados (tabla 2.2), y las aproximaciones (a través de 1000 simulaciones de las mismas) de las densidades de las distribuciones predictivas y finales dada la muestra (figura 2.6) de  $\bar{Y}_u$  ( $Y.mdu$ ) y  $\lambda$  (lambda), para el modelo Bayes de Regresión con cociente entre varianzas desconocido. También se obtuvieron para completar el análisis los estimadores Bayes (las medias de las distribuciones finales) para  $\beta$ , y en su caso,  $\sigma^2$ ,  $\sigma_\beta^2$  y  $\lambda$  (tabla 2.3), para los cinco modelos.

A partir de estos resultados, podemos señalar lo siguiente:

1.- En relación a la estimación de la media poblacional debemos destacar las coincidencias en valor estimado de los cuatro modelos conjugados (normal, normal-gamma, normal mínimo informativo y normal-gamma mínimo informativo), que sitúan la fertilidad alrededor del 63.8 %, en oposición al valor estimado por el modelo con cociente entre varianzas desconocido que la situa cerca del 41.7 %. Estas diferencias entre los valores estimados para los modelos conjugados y no conjugado se deben a que éste

### Output for MB1CD analysis

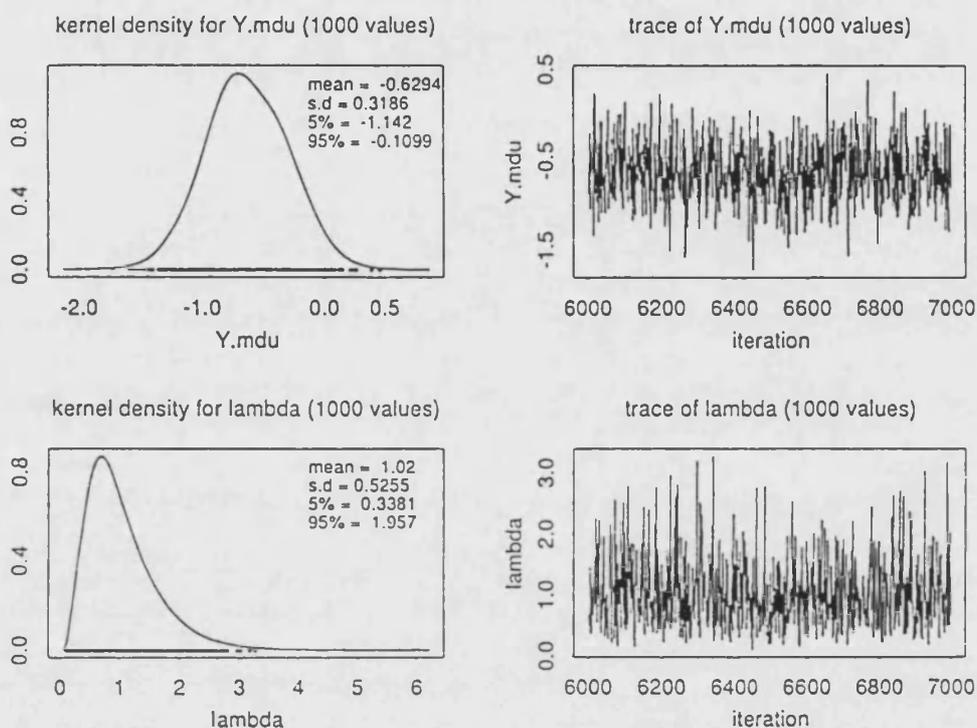


Figura 2.6: Aproximaciones a las densidades predictivas y finales para el modelo Bayes de Regresión con cociente entre varianzas desconocido

	MODELO NORMAL	MODELO NORM-GAMMA	MODELO COCIENTE VAR DESCONOCIDO	MODELO NORMAL MIN. INFOR.	MODELO NORM-GAMMA MIN. INFOR.
$\beta_1$	-0.2189	-0.2470	0.1541	-0.2151	-0.2151
$\beta_2$	0.3665	0.4581	0.4930	0.3541	0.3541
$\beta_3$	-0.2563	-0.3677	-0.5058	-0.2413	-0.2413
$\beta_4$	-0.1906	-0.1395	0.2552	-0.1973	-0.1973
$\sigma^2$		1.4416			2.0070
$\sigma_{\mu}^2$			3.772		
$\lambda$			1.019		

Tabla 2.3: Otras inferencias para los modelos Bayesianos en una etapa

último está detectando esos datos de la muestra alejados de la nube de puntos mayoritaria que ya señalábamos en la subsección anterior cuando analizábamos la figura 2.3, esto es, desviaciones en la hipótesis de linealidad entre la variable respuesta y las explicativas. Esta observación, se corrobora a partir de los resultados sobre la estimación de la pendiente de regresión que proporciona la tabla 2.3; notemos los cambios de signo entre los estimadores de  $\beta_1$  y  $\beta_4$ , para los modelos conjugados y no conjugado, que corresponden a los multiplicadores de la variable auxiliar agricultura y mortalidad infantil, que son las variables, que según la figura 2.3 presentan una mayor dispersión de las nubes de puntos.

2.- Respecto a la precisión de las estimaciones realizadas, notemos que los valores de las varianzas de la media poblacional condicionada a la muestra para los modelos conjugados, toman valores entre 0.0051 y 0.0227, siguiendo el orden de mayor a menor conocimiento sobre el modelo dentro de los conjugados (varianza conocida, varianza conocida mínimo informativo, varianza desconocida y varianza desconocida mínimo informativo), y en último lugar (con el valor más alto) el modelo con cociente entre varianzas desconocido.

3.- De los dos puntos anteriores se deriva la utilidad del modelo Bayes de Regresión con cociente entre varianzas desconocido, introducido en este capítulo de la tesis, ya que permite mediante su comparación con los modelos conjugados habituales, la detección de las desviaciones de la hipótesis de linealidad y consecuentemente, nos conduce a estudiar la posibilidad de una estructuración de la población que nos permita describir las distintas singularidades de la misma. Además, proporciona una valoración de la precisión de la estimación realizada más conservadora, y en este caso más creíble, atendiendo a las diferencias comentadas entre el modelo y los datos de la muestra.

4.- El modelo Bayes de Regresión con cociente entre varianzas desconocido, presenta el inconveniente de dar lugar a distribuciones predictivas no estándar para la media poblacional no observada y vector poblacional no observado. La figura 2.6, presenta las aproximaciones de kernel a las densidades de la distribución predictiva de la media poblacional no observada y para la distribución final de  $\lambda$  dada la muestra (definida en (2.52)), correspondiente a la muestra analizada de la población 1A. Si comparamos la densidad de la función predictiva para la media poblacional no observada, representada en la figura 2.6, con las correspondientes distribuciones predictivas para los cuatro modelos conjugados (tabla 2.2), podemos destacar lo siguiente:

- La forma acampanada de la densidad de la distribución predictiva para el modelo con cociente entre varianzas desconocido aunque ligeramente asimétrica; para los modelos conjugados todas las densidades de las distribuciones predictivas tienen forma acampanada y simétrica.
- Si comparamos el apuntamiento de la densidad de la distribución predictiva para el modelo con cociente entre varianzas desconocido con el correspondiente a una normal equivalente (en media y varianza), notamos que éste es inferior al de la normal (en el caso de la población estudiada y la muestra empleada).

- Si comparamos las 5 distribuciones predictivas podemos establecer el siguiente orden de mayor a menor apuntamiento de las correspondientes funciones de densidad (o de intervalos de confianza al 90 % menores a mayores): modelo normal, modelo normal mínimo informativo, modelo normal-gamma, modelo normal-gamma mínimo informativo y modelo con cociente entre varianzas desconocido.

(La figura 2.6 muestra a la derecha un gráfico que representa los valores de las simulaciones de las variables  $\bar{Y}_u$  y  $\lambda$ , en el que se aprecia la estabilidad (asumida) tras 5000 iteraciones Gibbs.)

## 2.4.2 Población 1B: modelos Bayesianos en dos etapas con una jerarquía

En esta subsección emplearemos los modelos descritos en 2.3 para realizar inferencias sobre la media poblacional y medias poblacionales de cada estrato de la población 1B. Los modelos concretamente aplicados previa transformación de los porcentajes  $p$ , de acuerdo con la fórmula:  $\log\left(\frac{p}{100-p}\right)$ , para adecuar nuestros datos a la hipótesis de normalidad, han sido los siguientes:

1. El modelo Bayes de Regresión en dos etapas con una jerarquía Normal, con parámetros  $\mathbf{V}_i = \sigma_i^2 \mathbf{I}_{N_i}$ ,  $i = 1, 2, 3$ ,  $\sigma_1^2 = 0.36$ ,  $\sigma_2^2 = 0.5$ ,  $\sigma_3^2 = 0.43$ ,  $\mathbf{B} = 4\mathbf{I}_4$  y  $\boldsymbol{\beta} = (-1, 2, -2, 3)^t$ .
2. El modelo Bayes de Regresión en dos etapas con una jerarquía Normal-Gamma, con parámetros  $\mathbf{W}_i = \mathbf{I}_{N_i}$ ,  $i = 1, 2, 3$ ,  $\mathbf{C}_\beta = \mathbf{I}_4$ ,  $\boldsymbol{\beta} = (-1, 2, -2, 3)^t$ ,  $a_0 = 3$  y  $b_0 = 8$  ( $E(\sigma^2) = 4$ ,  $Var(\sigma^2) = 16$ ).
3. El modelo Bayes de Regresión en dos etapas con una jerarquía con cociente entre varianzas desconocido, con parámetros  $\mathbf{W}_i = \mathbf{I}_{N_i}$ ,  $i = 1, 2, 3$ ,  $\mathbf{C}_\beta = \mathbf{I}_4$ ,  $\boldsymbol{\beta} = (-1, 2, -2, 3)^t$ ,  $a_0 = 3$ ,  $b_0 = 2$  ( $E(\sigma^2) = 1$ ,  $Var(\sigma^2) = 1$ ),  $a_\beta = 3$  y  $b_\beta = 8$  ( $E(\sigma_\beta^2) = 4$ ,  $Var(\sigma_\beta^2) = 16$ ).

Para obtener inferencias de la población a través de este modelo se aplicó el programa BUGS, en un proceso de tres fases: la traducción del modelo al BUGS (figura 2.7), la fase de simulación y una vez hemos supuesto alcanzada la estabilidad de la cadena, la obtención de la media y varianza de las distribuciones predictivas de las medias de cada estrato y de la población y finales de las pendientes de regresión y demás variables de interés (cuya existencia queda asegurada por el teorema 2.7) y de la aproximación de kernel a las densidades de las distribuciones predictivas de la media poblacional no observada y medias no observadas de cada estrato así como la de las distribuciones finales marginales de  $\lambda_i$ ,  $i = 1, 2, 3$ . La fase de simulación constó de la simulación de 1000 primeras observaciones que no se emplearon en el análisis para obtener un punto de arranque para la simulación bueno, la simulación a partir de dicho punto de arranque de 5000 observaciones tras las cuales se consideró alcanzada la estabilidad, una vez estudiado el estadístico aproximado de Geweke(92)[39] de las variables  $\bar{Y}_u$ ,  $\beta_{23}$ ,

$\sigma_3^2$  y  $\sigma_\beta^2$  (que se obtuvo con el comando **diag** del BUGS y alcanzó valores inferiores en módulo a 1.96), y por último la simulación de 1000 observaciones más, que consideradas simulaciones de las distribuciones predictivas y finales del modelo fueron empleadas para la estimación de la media y varianza de las distribuciones predictivas y finales (mediante el comando **stats** del BUGS) y la obtención de la aproximación de kernel a las densidades de las mismas (mediante el programa de Splus, **drawdat**, incluido en el BUGS). (Para más detalles, sobre cómo se ha fundamentado el análisis a través del programa BUGS, ver la sección C.3 del apéndice C.)

4. El modelo Bayes de Regresión en dos etapas con una jerarquía Normal mínimo informativo, con parámetro  $\mathbf{V}_i = \sigma_i^2 \mathbf{I}_{N_i}$ ,  $i = 1, 2, 3$  con  $\sigma_1^2 = 0.36$ ,  $\sigma_2^2 = 0.5$  y  $\sigma_3^2 = 0.43$ .
5. El modelo Bayes de Regresión en dos etapas con una jerarquía Normal-Gamma mínimo informativo, con parámetro  $\mathbf{W}_i = \mathbf{I}_{N_i}$ ,  $i = 1, 2, 3$ .

A partir de los datos de la muestra de la población 1B (en la que los elementos muestreados ya aparecen convenientemente separados por estratos) y los cinco modelos anteriores, se obtuvieron, aplicando los teoremas 2.5, 2.6, los teoremas 2.7 y 2.8 (que justifican la aplicación del programa BUGS) y los corolarios 2.3 y 2.4, los estimadores Bayes de la media de cada estrato y de la población y la varianza condicionada a la muestra de dichas medias (tabla 2.4), las distribuciones predictivas de la media no observada de cada estrato y de la media no observada poblacional para los modelos conjugados (tabla 2.5), y las aproximaciones (a través de 1000 simulaciones de las mismas) de las densidades de las distribuciones predictivas y finales dada la muestra de  $\bar{Y}_u$ , (Y.mdu[i]),  $i = 1, 2, 3$ ,  $\bar{Y}_u$  (Y.medu) y  $\lambda_i$  (lambda[i]),  $i = 1, 2, 3$ , para el modelo Bayes de Regresión en dos etapas con una jerarquía con cociente entre varianzas desconocido (figuras 2.8 y 2.9). También se obtuvieron los estimadores Bayes para  $\beta_i$ ,  $i = 1, 2, 3$ , y en su caso,  $\sigma^2$ ,  $\sigma_\beta^2$  y  $\lambda_i$ ,  $i = 1, 2, 3$ , para los cinco modelos (tabla 2.6).

A partir de estos resultados, podemos destacar lo siguiente:

1.- Respecto a la estimación de la media de cada estrato y de la población, señalemos:

- La semejanza de los estimadores de la media del estrato 1 (alrededor del 69.5 %) para los cinco modelos.
- Los estimadores de la media de los estratos 2 y 3 para los modelos normal y mínimo informativos son bastante similares (alrededor del 73 % y 76 %, respectivamente) y algo distintos para el modelo normal-gamma y sobre todo para el modelo con cociente entre varianzas desconocido, que llega a diferir en 10 tantos porcentuales respecto al modelo normal.

Nuevamente, tal y como sucedía en los modelos de una etapa, las razones de estas diferencias se encuentran en la naturaleza algo dispersa de las nubes de puntos (que suponen desviaciones en la hipótesis de linealidad) de la variable en estudio en relación con las variables socioeconómicas, en los estratos 2 y 3, sobre todo en relación con las variables agricultura y mortalidad infantil (ver figura 2.4), que

Modelo Bayes en dos etapas con una jerarquía con cociente entre varianzas desconocido	
<pre> model MB21CD. const NT=250, nt=47, p=4, K=3. var y[nt], y mds[K], Y mdu[K], Y md[k], Y med, Y medu, Y totu[K], x[nt,p], X mdu[K,p], beta[K,p], b[p], sigma2[K], sigma2 b, m[nt], M[K], tau[K], tau b, tau m[K], lambda[K], n[K], N[K]. data in "file.dat". ; for(i in 1:n[1]); y[i]~dnorm(m[i],tau[1]), m[i]~x[i,1]*beta[1,1]+x[i,2]*beta[1,2]+ x[i,3]*beta[1,3]+x[i,4]*beta[1,4]. ; for(i in n[1]+1:n[2]); y[i]~dnorm(m[i],tau[2]), m[i]~x[i,1]*beta[2,1]+x[i,2]*beta[2,2]+ x[i,3]*beta[2,3]+x[i,4]*beta[2,4]. ; for(i in n[1]+n[2]+1:n[3]); y[i]~dnorm(m[i],tau[3]), m[i]~x[i,1]*beta[3,1]+x[i,2]*beta[3,2]+ x[i,3]*beta[3,3]+x[i,4]*beta[3,4]. ; </pre>	<pre> for(i in 1:K); for(j in 1:p); beta[i,j]~dnorm(b[j],tau b). ; Y mdu[i,j]~dnorm(M[i],tau m[i]), tau m[i]~(N[i]-n[i])*tau[i], M[i]~X mdu[i,1]*beta[i,1]+X mdu[i,2]*beta[i,2]+ X mdu[i,3]*beta[i,3]+X mdu[i,4]*beta[i,4], tau[i]~dgamma(3 0,2 0), sigma2[i]~1/tau[i], lambda[i]~sigma2[i]/sigma2 b, Y totu[i]~(N[i]-n[i])*Y mdu[i], Y md[i]~(n[i]/N[i])*y mds[i]+Y totu[i]/N[i]. ; tau b~dgamma(3 0,8 0), sigma2 b~1/tau b, Y medu~sum(Y totu[])/(NT-nt), Y med~sum(y[,])/NT+sum(Y totu[])/NT. </pre>
<pre> "file data" (contenido) list(y=c(-datos de la muestra separados por comas-), x=c(-datos variable explicativa de la muestra separados por comas-), b=c(-1,2,-2,3), X mdu=c(-vector medias no observadas-), y mds=c(-vector de medias muestrales para cada estrato), n=c(33,7,7), N=c(170,37,37)) </pre>	

Figura 2.7: Traducción en BUGS del modelo Bayes de Regresión en dos etapas con una jerarquía con cociente entre varianzas desconocido

### Output for MB21CD analysis

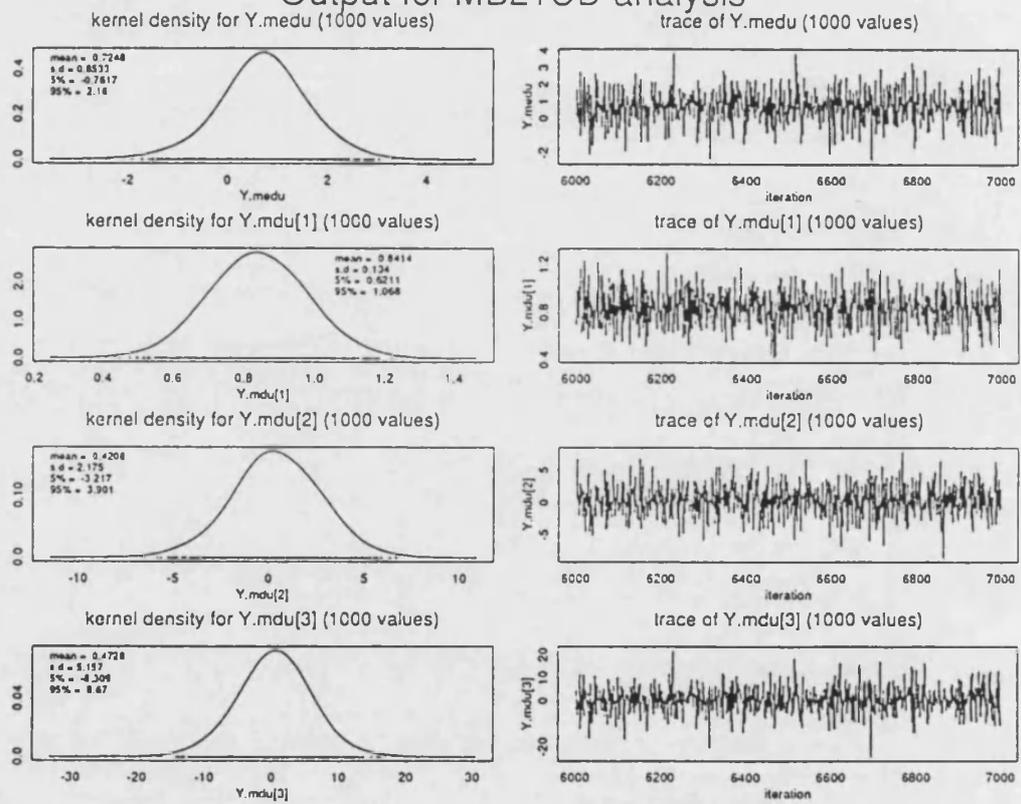


Figura 2.8: Aproximaciones a las densidades predictivas para el modelo Bayes de Regresión en dos etapas con una jerarquía con cociente entre varianzas desconocido

### Output for MB21CD (2) analysis

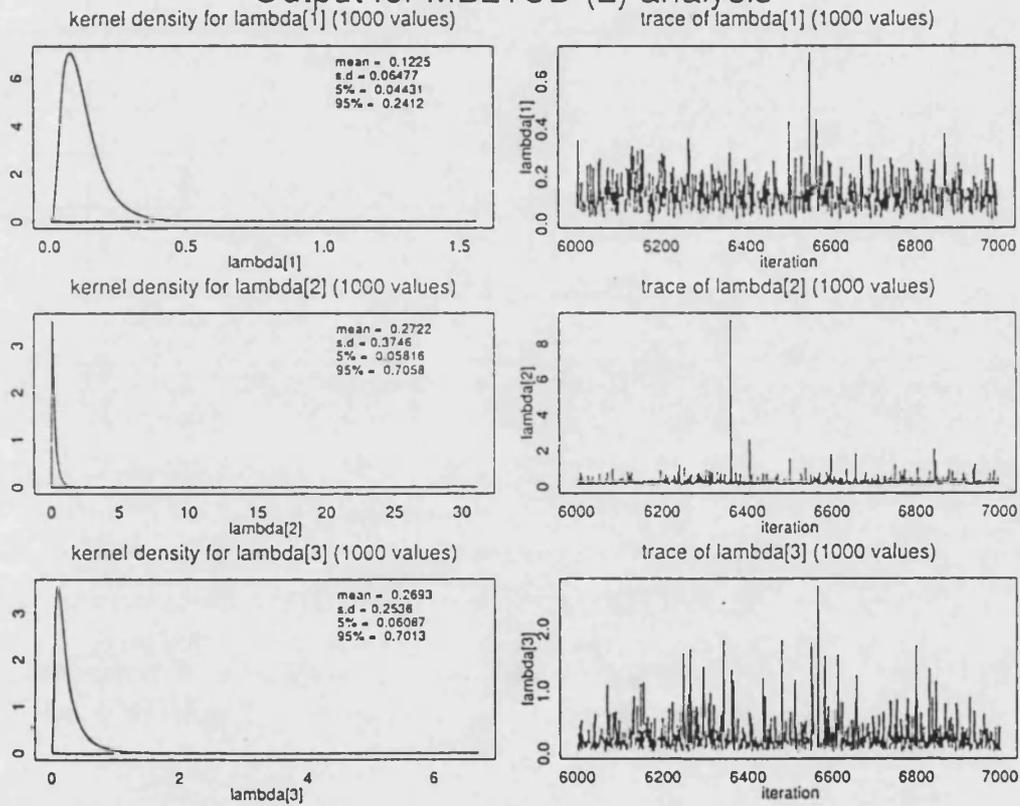


Figura 2.9: Aproximaciones a las densidades finales para el modelo Bayes de Regresión en dos etapas con una jerarquía con cociente entre varianzas desconocido

		MODELO NORMAL	MODELO NORM-GAMMA	MODELO COCIENTE VAR. DESCONOCIDO	MODELO NORMAL MIN. INFORM.	MODELO NORM-GAMMA MIN. INFORM.
E. 1	$\hat{\bar{Y}}_1$	0.8254 (69.54)	0.8019 (69.04)	0.8258 (69.55)	0.8278 (69.59)	0.8278 (69.59)
	$\text{Var}(\hat{\bar{Y}}_1   \{s, y_s\})$	0.0104	0.0373	0.0118	0.0104	0.0907
E. 2	$\hat{\bar{Y}}_2$	1.0292 (73.68)	1.1245 (75.48)	0.6002 (64.57)	1.0028 (73.16)	1.0028 (73.16)
	$\text{Var}(\hat{\bar{Y}}_2   \{s, y_s\})$	0.0500	0.1126	3.1046	0.0515	0.3238
E. 3	$\hat{\bar{Y}}_3$	1.1483 (75.92)	1.0125 (73.35)	0.6391 (65.45)	1.1818 (76.53)	1.1818 (76.53)
	$\text{Var}(\hat{\bar{Y}}_3   \{s, y_s\})$	0.0381	0.0899	17.4640	0.0412	0.3015
	$\hat{\bar{Y}}$	0.9053 (71.16)	0.8808 (70.70)	0.7648 (68.24)	0.9061 (71.22)	0.9061 (71.22)
	$\text{Var}(\hat{\bar{Y}}   \{s, y_s\})$	0.0071	0.0229	0.4796	0.0072	0.0586

Las estimaciones de la media poblacional de cada estrato y de toda la población en porcentajes figuran en la tabla entre paréntesis

Tabla 2.4: Estimadores Bayes de la media de cada estrato y poblacional y sus varianzas para los modelos Bayesianos en dos etapas con una jerarquía

conducen a estimaciones algo distintas de las pendientes de regresión asociadas a las mismas (ver tabla 2.6). Las diferencias, no obstante, entre los estimadores son mucho menores que en los modelos Bayesianos en una etapa.

- Los estimadores Bayes de la media poblacional son bastante similares (alrededor del 71 %), para los cinco modelos, y las diferencias mayores, de unos 3 puntos porcentuales corresponden al modelo con cociente entre varianzas desconocido y se explican por las diferencias en la estimación de la media de los estratos 2 y 3 de este modelo con los cuatro restantes.

Si comparamos estos resultados de estimación con los obtenidos para los modelos Bayesianos en una etapa se aprecia que en ellos se subestimaba el valor medio de la fertilidad para la totalidad de la población en unos 7 puntos en relación a los valores estimados actualmente. Ello, no obstante, la estimación mediante los modelos conjugados en una etapa era mucho mejor que la obtenida por el modelo en una etapa con cociente entre varianzas desconocido.

2.- Si nos fijamos en la precisión de los estimadores a partir de la varianza condicionada a la muestra de las medias de cada estrato y de la población, podemos establecer las siguientes observaciones, compatibles con las del párrafo anterior:

- Las varianzas condicionadas a la muestra de la media del estrato 1 son muy similares para los cinco modelos, siguiendo el siguiente orden de menor a mayor

MOD. NORMAL	$\bar{Y}_{U_1}(s, y_s) \sim N(0.8409, 0.0157)$ $\bar{Y}_{U_2}(s, y_s) \sim N(0.9498, 0.0761)$ $\bar{Y}_{U_3}(s, y_s) \sim N(1.1007, 0.0580)$ $\bar{Y}_U(s, y_s) \sim N(0.8954, 0.0107)$
MOD. NORM-GAMMA	$\bar{Y}_{U_1}(s, y_s) \sim t(53, 0.8120, 0.0543)$ $\bar{Y}_{U_2}(s, y_s) \sim t(53, 1.0673, 0.1649)$ $\bar{Y}_{U_3}(s, y_s) \sim t(53, 0.9333, 0.1316)$ $\bar{Y}_U(s, y_s) \sim t(53, 0.8676, 0.0334)$
MOD. NORMAL MIN. INF.	$\bar{Y}_{U_1}(s, y_s) \sim N(0.8439, 0.0157)$ $\bar{Y}_{U_2}(s, y_s) \sim N(0.9172, 0.0783)$ $\bar{Y}_{U_3}(s, y_s) \sim N(1.1421, 0.06268)$ $\bar{Y}_U(s, y_s) \sim N(0.8983, 0.0109)$
MOD. NORM-GAMMA MIN. INF.	$\bar{Y}_{U_1}(s, y_s) \sim t(35, 0.8439, 0.1295)$ $\bar{Y}_{U_2}(s, y_s) \sim t(35, 0.9172, 0.4644)$ $\bar{Y}_{U_3}(s, y_s) \sim t(35, 1.1421, 0.4324)$ $\bar{Y}_U(s, y_s) \sim t(35, 0.8983, 0.0838)$

Tabla 2.5: Distribuciones predictivas de la media de cada estrato y poblacional no observada para los modelos Bayesianos en dos etapas con una jerarquía conjugados

		MODELO NORMAL	MODELO NORM-GAMMA	MODELO COCIENTE VAR DESCONOCIDO	MODELO NORMAL MIN INFOR.	MODELO NORM-GAMMA MIN INFOR.
E S T R A T O 1	$\beta_{11}$	-0.0031	-0.1384	0.0115	0.0115	0.0115
	$\beta_{12}$	-0.0301	0.1141	-0.0485	-0.0458	-0.0458
	$\beta_{13}$	-0.0596	-0.1997	-0.0641	-0.0448	-0.0448
	$\beta_{14}$	-0.2623	-0.1296	-0.2559	-0.2763	-0.2763
	$\lambda_1$			0.1224		
E S T R A T O 2	$\beta_{21}$	-0.7581	-1.1112	-0.0986	-0.6623	-0.6623
	$\beta_{22}$	0.8449	1.1227	2.052	0.7791	0.7791
	$\beta_{23}$	-1.5739	-1.8506	-1.980	-1.5056	-1.5056
	$\beta_{24}$	1.1539	1.5442	3.094	1.0605	1.0605
	$\lambda_2$			0.2721		
E S T R A T O 3	$\beta_{31}$	-0.8512	-0.5222	-0.9747	-0.9221	-0.9221
	$\beta_{32}$	2.8404	2.5087	2.005	2.9603	2.9603
	$\beta_{33}$	-2.8081	-2.5431	-1.940	-2.9244	-2.9244
	$\beta_{34}$	0.2907	0.5115	2.966	0.2718	0.2718
	$\lambda_3$			0.2692		
	$\sigma_A^2$			4.460		
	$\sigma^2$		1.2980			3.1463

Tabla 2.6: Otras inferencias para los modelos Bayesianos en dos etapas con una jerarquía

varianza: modelos normal y normal mínimo informativo, modelo con cociente entre varianzas desconocido, modelo normal-gamma y modelo normal-gamma mínimo informativo.

- Las varianzas condicionadas de la media de los estratos 2 y 3 y de la población para los modelos conjugados son bastante similares (aunque algo más altas para el modelo normal-gamma mínimo informativo en el caso de las varianzas condicionadas de las medias de los estratos 2 y 3) y siguen el orden de menor a mayor varianza habitual para los modelos conjugados compatible con el nivel de conocimiento introducido en el modelo (varianza conocida, varianza conocida mínimo informativo, varianza desconocida y varianza desconocida mínimo informativo). Respecto a las varianzas para estas mismas medias para el modelo con cociente entre varianzas desconocido, destacar los altos valores correspondientes a los estratos 2 y 3.

Si comparamos la varianzas de la media poblacional condicionada a la muestra para estos cinco modelos en relación con la obtenida para los modelos Bayesianos en una etapa, se observa que los valores obtenidos son todos más altos, lo que confirma nuestra sospecha en la subsección anterior sobre que se estaba infravalorando el error cometido en la estimación.

3.- Las observaciones de los párrafos anteriores confirman la utilidad de los modelos con cociente entre varianzas desconocido para estudiar en qué medida la población se ajusta al modelo de regresión lineal. En concreto, notemos que el modelo Bayes de Regresión en dos etapas con una jerarquía con cociente en re varianzas desconocido actua de la siguiente manera:

- Cuando se satisfacen las hipótesis de linealidad (nube de puntos para cada variable explicativa mayoritariamente concentrada alrededor de una recta), caso del estrato 1; el modelo proporciona un estimador para la media del estrato similar al de los modelos conjugados, y un valor de la varianza condicionada a la muestra de la misma que se sitúa entre la obtenida para el modelo normal y para el modelo normal-gamma (notemos, que el modelo con cociente entre varianzas desconocido es “mejor” que el modelo normal-gamma porque modeliza una varianza distinta para la característica en estudio de la población y para las pendientes de los estratos).
- Cuando se aprecian desviaciones a la hipótesis de linealidad para una o varias variables explicativas, caso de los estratos 2 y 3; el estimador Bayes para la media del estrato para este modelo proporciona valores que difieren sustancialmente de los obtenidos para los modelos conjugados y un valor para la varianza condicionada muy alto en comparación con los obtenidos para los otros modelos.

4.- El modelo Bayes de Regresión en dos etapas con una jerarquía proporciona una herramienta útil para el análisis de una población estructurada en dos etapas, pero al no ser conjugado, las distribuciones predictivas y finales no son estándar. Las figuras 2.8 y 2.9 proporcionan una aproximación a las densidades de las distribuciones predictivas

para la media no observada de cada estrato y de la población y de las distribuciones finales marginales de los  $\lambda_i$ ,  $i = 1, 2, 3$ . Del análisis de estas figuras se desprende que:

- Las densidades de las distribuciones predictivas de la media de cada estrato y de la poblacional tienen forma acampanada aunque son ligeramente asimétricas, frente a las de los modelos conjugados que son acampanadas y simétricas, puesto que son normales y  $t$ 's (ver figura 2.9).
- Respecto al apuntamiento de las densidades de las distribuciones predictivas, señalemos que en el caso de la media poblacional no observada y la media no observada del estrato 1 éste es ligeramente inferior al de las densidades de las correspondientes distribuciones normales, siendo en cambio superior al de la densidad de la correspondiente normal en los casos de las densidades de las distribuciones predictivas de la media no observada del estrato 2 y 3 (en el caso de la población estudiada y para la muestra empleada).
- Si comparamos las cinco distribuciones predictivas (a partir de la tabla 2.5 y la figura 2.8), se observa en relación con la amplitud del intervalo de confianza al 90 % el siguiente orden de menor a mayor longitud: modelo normal, modelo normal mínimo informativo, modelo normal-gamma, modelo normal-gamma mínimo informativo y modelo con cociente entre varianzas desconocido, salvo en el caso de la distribución predictiva de la media no observada del estrato 1, para la que el modelo con cociente entre varianzas desconocido se sitúa entre el modelo normal mínimo informativo y el modelo normal-gamma.

Señalemos también que la densidad de la distribución predictiva para la media no observada del estrato 1 para el modelo con cociente entre varianzas desconocido es muy similar a la densidad de la correspondiente distribución predictiva para el modelo normal.

- Respecto a las densidades de las distribuciones finales marginales de  $\lambda_i$ ,  $i = 1, 2, 3$ , notemos comparando con la densidad de la distribución de  $\lambda|(s, \mathbf{y}_s)$  para el modelo Bayes de Regresión con cociente entre varianzas desconocido (en una etapa) representada en la figura 2.6, que son funciones de densidad con un mayor apuntamiento y una varianza menor sobre todo, en el caso de  $\lambda_1|(s, \mathbf{y}_s)$ .

(Las figuras 2.8 y 2.9 muestran a la derecha un gráfico que representa los valores de las simulaciones de las variables  $\bar{Y}_u$ ,  $\bar{Y}_{u_i}$ ,  $i = 1, 2, 3$ , y  $\lambda_i$ ,  $i = 1, 2, 3$ , en el que se aprecia gráficamente la estabilidad asumida tras la realización de 5000 iteraciones Gibbs.)

### 2.4.3 Conclusiones de la sección

Para terminar la sección debemos destacar las siguientes ideas:

- 1.- El análisis a través de los modelos Bayesianos en una etapa de la población no estratificada detectaba a través del modelo con cociente entre varianzas desconocido

las desviaciones a la hipótesis de linealidad entre las variables explicativas y la característica en estudio (ver figura 2.3) y nos hacía suponer que la precisión real del estimador era menor que la obtenida.

**2.-** Los resultados del análisis de los modelos Bayesianos en una etapa al no ser totalmente satisfactorios motivan a la realización de una estratificación de la población

**3.-** Los modelos Bayes en dos etapas con una jerarquía proporcionan resultados bastante similares entre sí, aunque el correspondiente modelo Bayes con cociente entre varianzas desconocido sigue detectando desviaciones de la linealidad en los estratos 2 y 3.

**4.-** De la comparación entre los estimadores de la media poblacional para los modelos Bayesianos en una etapa y dos etapas con una jerarquía se confirma la robustez de los estimadores obtenidos para los modelos Bayesianos en una etapa conjugados, ya que, aunque las hipótesis de linealidad no se cumplen de manera totalmente satisfactoria, sólo difieren de los obtenidos tras la estratificación en 7 puntos porcentuales.

## **2.5 Comentarios finales: limitaciones de los modelos en una jerarquía**

En este capítulo hemos planteado varios modelos jerárquicos con una jerarquía y hemos señalado en qué circunstancias podían ser adecuados estos modelos. En el caso de poblaciones no estructuradas (monoetápicas) estos modelos resultan generalmente adecuados, pero en cambio, en el caso de poblaciones bietápicas, en muchas circunstancias (incluso cuando la población se estratifica y se aplica un muestreo estratificado), es más sencillo, flexible y natural el invocar una jerarquía más para el modelo y establecer una relación de intercambiabilidad entre las pendientes de las distintas unidades, éstos nuevos modelos los llamaremos modelos Bayesianos en dos etapas con dos jerarquías y serán objeto de estudio en el próximo capítulo (sección 2.2).

Otra razón, para la introducción de más jerarquías es el establecimiento de modelos (los modelos  $m$ -jerárquicos), que generalicen los modelos normales y normales-gamma de este capítulo y los modelos Bayes de Regresión en dos etapas con dos jerarquías Normal y Normal-Gamma del próximo capítulo y sus correspondientes variantes mínimo informativas. La introducción de estos modelos  $m$ -jerárquicos y la obtención de los correspondientes resultados inferenciales habituales dada la muestra, permiten por una parte simplificar las demostraciones de los resultados inferenciales para los modelos que generalizan (notemos que las demostraciones de estos modelos presentadas se apoyan en los teoremas 3.5 y 3.6 y corolario 3.3 y 3.4 correspondientes a resultados inferenciales de modelos  $m$ -jerárquicos que se estudiarán en el próximo capítulo), y por otra el establecer un marco de análisis adecuado para poblaciones estructuradas en más de dos etapas.

## Capítulo 3

# Modelos de Superpoblación Bayesianos II: Modelos en dos o más jerarquias

### 3.1 Introducción

Dedicaremos este capítulo a completar el estudio sobre la modelización Bayes de una característica continua de una población finita, que iniciábamos en el capítulo 2. Este objetivo se cubrirá definiendo y analizando los modelos Bayesianos en dos etapas con dos jerarquías y los modelos Bayesianos  $m$ -jerárquicos.

Los modelos Bayesianos en dos etapas con dos jerarquías se aplican para modelizar características continuas en poblaciones con estructura bietápica, cuando se desea introducir una mayor relación entre las unidades mediante una jerarquía más en el modelo, de modo que, los elementos de la muestra correspondientes a unidades distintas proporcionen información acerca de una unidad dada concreta. El estudio analítico de estos modelos se realizará en la sección 3.2 y se completará en la sección 3.3 con la aplicación de los mismos para la realización de inferencias sobre la población 1B, que ya describíamos y estudiábamos mediante modelos Bayesianos en una jerarquía en el capítulo anterior.

Los modelos Bayesianos  $m$ -jerárquicos, que introduciremos y analizaremos en la sección 3.4 generalizan los modelos normales y normales-gamma del capítulo 2 y de la sección 3.2 del presente capítulo y se introducen en esta tesis con un doble objetivo, ya que por una parte, permiten una demostración más breve de los modelos normales y normales-gamma introducidos en el capítulo 2 y en la sección 3.2, (esta demostración ha sido la aplicada en esta tesis); y por otra, establecen el marco adecuado para el estudio de cualquier modelo normal o normal-gamma conjugado que tenga por objeto realizar inferencias sobre una característica continua de una población estructurada en tres o más

etapas dada una muestra de la misma. Los modelos  $m$ -jerárquicos que introduciremos en este capítulo son también útiles para la obtención de los resultados inferenciales habituales de modelos normal y normal-gamma similares a los introducidos en el capítulo 2 y en la próxima sección cuando el modelo de regresión se introduce en el segundo nivel del modelo (es decir, cuando las covariables del modelo proporcionan información acerca de la pendiente y no directamente sobre la característica en estudio), y cuando se introduce el modelo de regresión en uno o varios niveles cualesquiera del modelo de la población.

Para finalizar el capítulo, en la sección 3.5 se realizará un comentario del estudio realizado, su utilidad y sus limitaciones.

## 3.2 Modelos Bayesianos en dos etapas con dos jerarquías

Esta sección se dedicará al estudio de los modelos Bayesianos en dos etapas con dos jerarquías. Estos modelos son los más empleados en la modelización de una característica continua en poblaciones con una estructura en dos etapas al permitir (con la introducción de una jerarquía más), modelizar una relación entre las distintas unidades.

Como antecedentes de estos modelos, citemos a Scott y Smith(69)[106] y los trabajos en poblaciones finitas con estos modelos de Murgui(82)[76] (comentados en el apartado estimadores de la media poblacional de la subsección 1.4.2).

Los resultados inferenciales que estudiaremos en esta sección son a posteriori de la elección de una muestra de la población mediante un procedimiento de muestreo en dos etapas dado de tamaños fijos para  $k$  y  $n_i$ ,  $i \in \{i_1, i_2, \dots, i_k\}$  y no informativo, en el que los elementos y unidades seleccionados son distintos entre sí. Y la notación aplicada en la sección y los lemas previos en los que se apoyan las demostraciones de los teoremas aparecen descritos y estudiados en los apéndices A (en el apartado de estructura bietápica) y B respectivamente.

### 3.2.1 Modelo Bayes de Regresión en dos etapas con dos jerarquías normal

En primer lugar, estudiaremos el modelo Bayes de Regresión en dos etapas con dos jerarquías normal, este modelo que definiremos a continuación, se aplica para realizar inferencias sobre una característica continua en poblaciones estructuradas en dos etapas, cuando la varianza de la población es conocida y deseamos modelizar una relación de intercambiabilidad entre las unidades.

**Definición 3.1** *Dado el vector poblacional  $\mathbf{Y} = (\mathbf{Y}_1^t, \mathbf{Y}_2^t, \dots, \mathbf{Y}_K^t)^t$  con  $\mathbf{Y}_i = (Y_{i1}, Y_{i2}, \dots, Y_{iN_i})^t$ ,  $i = 1, 2, \dots, K$  para el cual asociada a cada variable  $Y_{ij}$  tenemos una información auxiliar representada por un vector  $\mathbf{x}_{ij} = (x_{ij1}, \dots, x_{ijp})$  conocido,  $j =$*

$1, 2, \dots, N_i, i = 1, 2, \dots, K$ , llamaremos modelo Bayes de Regresión en dos etapas con dos jerarquías Normal al definido por:

$$\begin{aligned} \mathbf{Y}_i | \mathbf{X}_i, \beta_i &\sim N(\mathbf{X}_i \beta_i, \mathbf{V}_i), \text{ independientes, } i = 1, 2, \dots, K, \\ \beta_i | \beta &\sim N_p(\beta, \mathbf{B}_\beta), \text{ i.i.d., } i = 1, 2, \dots, K, \\ \beta &\sim N_p(\beta_0, \mathbf{B}_{\beta_0}), \end{aligned} \quad (3.1)$$

donde  $\mathbf{X}_i = (\mathbf{x}_{i1}^t, \mathbf{x}_{i2}^t, \dots, \mathbf{x}_{iN_i}^t)^t$ , es una matriz de rango  $p$ , para  $i = 1, 2, \dots, K$ ,  $\beta_0$  es un  $p$ -vector conocido y  $\mathbf{V}_i, i = 1, 2, \dots, K, \mathbf{B}_\beta$  y  $\mathbf{B}_{\beta_0}$  son matrices simétricas definidas positivas conocidas de dimensiones  $N_i \times N_i, i = 1, 2, \dots, K, p \times p$  y  $p \times p$  respectivamente.

Este modelo presenta como caso particular para  $p = 1, \mathbf{X}_i = \mathbf{1}_{N_i}, i = 1, 2, \dots, K$ , el modelo Bayes Básico en dos etapas con dos jerarquías normal y como particularización de éste el modelo estudiado por Murgui(82)[76] (ver el apartado estimadores Bayes de la media poblacional de la subsección 1.4.2).

**Teorema 3.1** Consideremos el modelo Bayes de Regresión en dos etapas con dos jerarquías Normal dado en (3.1), entonces se tiene que:

(a)  $\mathbf{Y}_u | (s, \mathbf{y}_s)$  se distribuye normal  $(N - n)$ -variante con media,

$$E(\mathbf{Y}_u | (s, \mathbf{y}_s)) = \begin{cases} \mathbf{X}_{u_i} \hat{\beta}_i^B + \mathbf{V}_{su_i}^t \mathbf{V}_{s_i}^{-1} (\mathbf{y}_{s_i} - \mathbf{X}_{s_i} \hat{\beta}_i^B), & i \in \{i_1, i_2, \dots, i_k\}, \\ \mathbf{X}_{u_i} \hat{\beta}_i^B & i \notin \{i_1, i_2, \dots, i_k\}, \end{cases} \quad (3.2)$$

y matriz de varianzas y covarianzas,

\* para  $i \in \{i_1, i_2, \dots, i_k\}$

$$\begin{aligned} \text{Var}(\mathbf{Y}_{u_i} | (s, \mathbf{y}_s)) &= \mathbf{V}_{u_i} - \mathbf{V}_{su_i}^t \mathbf{V}_{s_i}^{-1} \mathbf{V}_{su_i} + \\ &+ (\mathbf{X}_{u_i} - \mathbf{V}_{su_i}^t \mathbf{V}_{s_i}^{-1} \mathbf{X}_{s_i}) \alpha_i \mathbf{S}_i^{-1} \mathbf{B}_\beta^{-1} \mathbf{S}_w^{-1} \mathbf{S}_i^{-1} \alpha_i^t (\mathbf{X}_{u_i} - \mathbf{V}_{su_i}^t \mathbf{V}_{s_i}^{-1} \mathbf{X}_{s_i})^t + \\ &+ (\mathbf{X}_{u_i} - \mathbf{V}_{su_i}^t \mathbf{V}_{s_i}^{-1} \mathbf{X}_{s_i}) \alpha_i \mathbf{S}_i^{-1} (\mathbf{X}_{u_i} - \mathbf{V}_{su_i}^t \mathbf{V}_{s_i}^{-1} \mathbf{X}_{s_i})^t, \end{aligned} \quad (3.3)$$

\* para  $i \notin \{i_1, i_2, \dots, i_k\}$ ,

$$\text{Var}(\mathbf{Y}_{u_i} | (s, \mathbf{y}_s)) = \mathbf{V}_{u_i} + \mathbf{X}_{u_i} \mathbf{B}_\beta \mathbf{X}_{u_i}^t + \mathbf{X}_{u_i} \mathbf{S}_w^{-1} \mathbf{B}_\beta \mathbf{X}_{u_i}^t, \quad (3.4)$$

\* para  $i \in \{i_1, i_2, \dots, i_k\}, j \notin \{i_1, i_2, \dots, i_k\}$ ,

$$\text{Cov}(\mathbf{Y}_{u_i}, \mathbf{Y}_{u_j} | (s, \mathbf{y}_s)) = (\mathbf{X}_{u_i} - \mathbf{V}_{su_i}^t \mathbf{V}_{s_i}^{-1} \mathbf{X}_{s_i}) \boldsymbol{\alpha}_i \mathbf{S}_i^{-1} \mathbf{B}_\beta^{-1} \mathbf{S}_w^{-1} \mathbf{B}_\beta \mathbf{X}_{u_j}^t, \quad (3.5)$$

\* para  $i \neq j$ ,  $i, j \in \{i_1, i_2, \dots, i_k\}$ ,

$$\begin{aligned} \text{Cov}(\mathbf{Y}_{u_i}, \mathbf{Y}_{u_j} | (s, \mathbf{y}_s)) = & \quad (3.6) \\ (\mathbf{X}_{u_i} - \mathbf{V}_{su_i}^t \mathbf{V}_{s_i}^{-1} \mathbf{X}_{s_i}) \boldsymbol{\alpha}_i \mathbf{S}_i^{-1} \mathbf{B}_\beta^{-1} \mathbf{S}_w^{-1} \mathbf{S}_j^{-1} \boldsymbol{\alpha}_j^t (\mathbf{X}_{u_j} - \mathbf{V}_{su_j}^t \mathbf{V}_{s_j}^{-1} \mathbf{X}_{s_j})^t, \end{aligned}$$

\* para  $i \neq j$ ,  $i, j \notin \{i_1, i_2, \dots, i_k\}$ ,

$$\text{Cov}(\mathbf{Y}_{u_i}, \mathbf{Y}_{u_j} | (s, \mathbf{y}_s)) = \mathbf{X}_{u_i} \mathbf{S}_w^{-1} \mathbf{B}_\beta \mathbf{X}_{u_j}^t, \quad (3.7)$$

(b)  $\bar{Y}_{u_i} | (s, \mathbf{y}_s)$ ,  $i = 1, 2, \dots, K$ , se distribuye normal con media y varianza dadas por:

\* para  $i \in \{i_1, i_2, \dots, i_k\}$

$$\begin{aligned} E(\bar{Y}_{u_i} | (s, \mathbf{y}_s)) &= \frac{1}{N_i - n_i} \mathbf{1}_{N_i - n_i}^t [\mathbf{X}_{u_i} \hat{\boldsymbol{\beta}}_i^B + \mathbf{V}_{su_i}^t \mathbf{V}_{s_i}^{-1} (\mathbf{y}_{s_i} - \mathbf{X}_{s_i} \hat{\boldsymbol{\beta}}_i^B)], \quad (3.8) \\ \text{Var}(\bar{Y}_{u_i} | (s, \mathbf{y}_s)) &= \frac{1}{(N_i - n_i)^2} \{ \mathbf{1}_{N_i - n_i}^t (\mathbf{V}_{u_i} - \mathbf{V}_{su_i}^t \mathbf{V}_{s_i}^{-1} \mathbf{V}_{su_i}) \mathbf{1}_{N_i - n_i} + \\ &+ \mathbf{1}_{N_i - n_i}^t (\mathbf{X}_{u_i} - \mathbf{V}_{su_i}^t \mathbf{V}_{s_i}^{-1} \mathbf{X}_{s_i}) \boldsymbol{\alpha}_i \mathbf{S}_i^{-1} \mathbf{B}_\beta^{-1} \mathbf{S}_w^{-1} \mathbf{S}_i^{-1} \boldsymbol{\alpha}_i^t (\mathbf{X}_{u_i} - \mathbf{V}_{su_i}^t \mathbf{V}_{s_i}^{-1} \mathbf{X}_{s_i})^t \mathbf{1}_{N_i - n_i} + \\ &+ \mathbf{1}_{N_i - n_i}^t (\mathbf{X}_{u_i} - \mathbf{V}_{su_i}^t \mathbf{V}_{s_i}^{-1} \mathbf{X}_{s_i}) \boldsymbol{\alpha}_i \mathbf{S}_i^{-1} (\mathbf{X}_{u_i} - \mathbf{V}_{su_i}^t \mathbf{V}_{s_i}^{-1} \mathbf{X}_{s_i})^t \mathbf{1}_{N_i - n_i} \}, \end{aligned}$$

\* para  $i \notin \{i_1, i_2, \dots, i_k\}$

$$\begin{aligned} E(\bar{Y}_{u_i} | (s, \mathbf{y}_s)) &= \frac{1}{N_i} \mathbf{1}_{N_i}^t \mathbf{X}_{u_i} \hat{\boldsymbol{\beta}}_i^B, \quad (3.9) \\ \text{Var}(\bar{Y}_{u_i} | (s, \mathbf{y}_s)) &= \frac{1}{N_i^2} \{ \mathbf{1}_{N_i}^t (\mathbf{V}_{u_i} + \mathbf{X}_{u_i} \mathbf{B}_\beta \mathbf{X}_{u_i}^t) \mathbf{1}_{N_i} + \\ &+ \mathbf{1}_{N_i}^t \mathbf{X}_{u_i} \mathbf{S}_w^{-1} \mathbf{B}_\beta \mathbf{X}_{u_i}^t \mathbf{1}_{N_i} \}, \end{aligned}$$

(c) el estimador Bayes para la media de la unidad  $i$ ,  $i = 1, 2, \dots, K$  y su varianza condicionada a la muestra, vienen dados por:

\* para  $i \in \{i_1, i_2, \dots, i_k\}$

$$\begin{aligned} \hat{Y}_i &= \frac{n_i}{N_i} \bar{y}_{s_i} + \frac{1}{N_i} \mathbf{1}_{N_i - n_i}^t [\mathbf{X}_{u_i} \hat{\boldsymbol{\beta}}_i^B + \mathbf{V}_{su_i}^t \mathbf{V}_{s_i}^{-1} (\mathbf{y}_{s_i} - \mathbf{X}_{s_i} \hat{\boldsymbol{\beta}}_i^B)], \quad (3.10) \\ \text{Var}(\bar{Y}_i | (s, \mathbf{y}_s)) &= \frac{1}{N_i^2} \{ \mathbf{1}_{N_i - n_i}^t (\mathbf{V}_{u_i} - \mathbf{V}_{su_i}^t \mathbf{V}_{s_i}^{-1} \mathbf{V}_{su_i}) \mathbf{1}_{N_i - n_i} + \\ &+ \mathbf{1}_{N_i - n_i}^t (\mathbf{X}_{u_i} - \mathbf{V}_{su_i}^t \mathbf{V}_{s_i}^{-1} \mathbf{X}_{s_i}) \boldsymbol{\alpha}_i \mathbf{S}_i^{-1} \mathbf{B}_\beta^{-1} \mathbf{S}_w^{-1} \mathbf{S}_i^{-1} \boldsymbol{\alpha}_i^t (\mathbf{X}_{u_i} - \mathbf{V}_{su_i}^t \mathbf{V}_{s_i}^{-1} \mathbf{X}_{s_i})^t \mathbf{1}_{N_i - n_i} + \\ &+ \mathbf{1}_{N_i - n_i}^t (\mathbf{X}_{u_i} - \mathbf{V}_{su_i}^t \mathbf{V}_{s_i}^{-1} \mathbf{X}_{s_i}) \boldsymbol{\alpha}_i \mathbf{S}_i^{-1} (\mathbf{X}_{u_i} - \mathbf{V}_{su_i}^t \mathbf{V}_{s_i}^{-1} \mathbf{X}_{s_i})^t \mathbf{1}_{N_i - n_i} \}, \end{aligned}$$

\* para  $i \notin \{i_1, i_2, \dots, i_k\}$

$$\begin{aligned}\widehat{Y}_i &= \frac{1}{N_i} \mathbf{1}_{N_i}^t \mathbf{X}_{u_i} \widehat{\beta}_i^B, \\ \text{Var}(\overline{Y}_i | (s, \mathbf{y}_s)) &= \frac{1}{N_i^2} \{ \mathbf{1}_{N_i}^t (\mathbf{V}_{u_i} + \mathbf{X}_{u_i} \mathbf{B}_\beta \mathbf{X}_{u_i}^t) \mathbf{1}_{N_i} + \\ &\quad + \mathbf{1}_{N_i}^t \mathbf{X}_{u_i} \mathbf{S}_w^{-1} \mathbf{B}_\beta \mathbf{X}_{u_i}^t \mathbf{1}_{N_i} \},\end{aligned}\quad (3.11)$$

(d)  $\overline{Y}_u | (s, \mathbf{y}_s)$  se distribuye normal con media y varianza dadas por:

$$\begin{aligned}E(\overline{Y}_u | (s, \mathbf{y}_s)) &= \frac{1}{N-n} \sum_{i \in \{i_1, i_2, \dots, i_k\}} \mathbf{1}_{N_i - n_i}^t [\mathbf{X}_{u_i} \widehat{\beta}_i^B + \mathbf{V}_{s u_i}^t \mathbf{V}_{s_i}^{-1} (\mathbf{y}_{s_i} - \mathbf{X}_{s_i} \widehat{\beta}_i^B)] + \\ &\quad + \frac{1}{N-n} \sum_{i \notin \{i_1, i_2, \dots, i_k\}} \mathbf{1}_{N_i}^t \mathbf{X}_{u_i} \widehat{\beta}_i^B, \\ \text{Var}(\overline{Y}_u | (s, \mathbf{y}_s)) &= \frac{1}{(N-n)^2} \{v_1 + v_2\},\end{aligned}\quad (3.12)$$

y (e) el estimador Bayes para la media poblacional y su varianza condicionada a la muestra, vienen dados por:

$$\begin{aligned}\widehat{Y} &= \frac{n}{N} \bar{y}_s + \frac{1}{N} \sum_{i \in \{i_1, \dots, i_k\}} \mathbf{1}_{N_i - n_i}^t [\mathbf{X}_{u_i} \widehat{\beta}_i^B + \mathbf{V}_{s u_i}^t \mathbf{V}_{s_i}^{-1} (\mathbf{y}_{s_i} - \mathbf{X}_{s_i} \widehat{\beta}_i^B)] + \\ &\quad + \frac{1}{N} \sum_{i \notin \{i_1, \dots, i_k\}} \mathbf{1}_{N_i}^t \mathbf{X}_{u_i} \widehat{\beta}_i^B, \\ \text{Var}(\overline{Y} | (s, \mathbf{y}_s)) &= \frac{1}{N^2} \{v_1 + v_2\},\end{aligned}\quad (3.13)$$

donde  $\mathbf{V}_{s_i} = \text{Var}_{\mathbf{Y} | \beta}(\mathbf{Y}_{s_i})$ ,  $\mathbf{V}_{u_i} = \text{Var}_{\mathbf{Y} | \beta}(\mathbf{Y}_{u_i})$ ,  $\mathbf{V}_{s u_i} = \text{Cov}_{\mathbf{Y} | \beta}(\mathbf{Y}_{s_i}, \mathbf{Y}_{u_i})$ ,  $\bar{y}_s = \frac{\sum_{(i,j) \in s} y_{ij}}{n}$  es la media muestral,  $\bar{y}_{s_i}$ , es la media muestral de la unidad  $i$ ,  $i \in \{i_1, i_2, \dots, i_k\}$ ,  $\mathbf{S}_i = \mathbf{X}_{s_i}^t \mathbf{V}_{s_i}^{-1} \mathbf{X}_{s_i}$ ,  $i \in \{i_1, i_2, \dots, i_k\}$ , y

$$\begin{aligned}\alpha_i &= (\mathbf{X}_{s_i}^t \mathbf{V}_{s_i}^{-1} \mathbf{X}_{s_i} + \mathbf{B}_\beta^{-1})^{-1} \mathbf{X}_{s_i}^t \mathbf{V}_{s_i}^{-1} \mathbf{X}_{s_i}, \quad i \in \{i_1, i_2, \dots, i_k\}, \\ \widehat{\beta}_i &= \mathbf{S}_i^{-1} \mathbf{X}_{s_i}^t \mathbf{V}_{s_i}^{-1} \mathbf{y}_{s_i}, \quad i \in \{i_1, i_2, \dots, i_k\}, \\ \mathbf{S}_w &= \sum_{i \in \{i_1, i_2, \dots, i_k\}} \alpha_i + \mathbf{B}_\beta \mathbf{B}_{\beta_0}^{-1}, \\ \widehat{\beta}_w &= \mathbf{S}_w^{-1} \left( \sum_{i \in \{i_1, i_2, \dots, i_k\}} \alpha_i \widehat{\beta}_i + \mathbf{B}_\beta \mathbf{B}_{\beta_0}^{-1} \beta_0 \right),\end{aligned}\quad (3.14)$$

$$\begin{aligned}
\widehat{\beta}_i^B &= E(\beta_i | (s, y_s)) = \begin{cases} \alpha_i \widehat{\beta}_i + (\mathbf{I}_p - \alpha_i) \widehat{\beta}_w, & i \in \{i_1, i_2, \dots, i_k\}, \\ \widehat{\beta}_w, & i \notin \{i_1, i_2, \dots, i_k\}, \end{cases} \\
\mathbf{Z}_w &= \sum_{i \in \{i_1, i_2, \dots, i_k\}} \mathbf{1}_{N_i - n_i}^t (\mathbf{X}_{u_i} - \mathbf{V}_{s_{u_i}}^t \mathbf{V}_{s_i}^{-1} \mathbf{X}_{s_i}) \alpha_i \mathbf{S}_i^{-1} \mathbf{B}_\beta^{-1} + \sum_{i \notin \{i_1, i_2, \dots, i_k\}} \mathbf{1}_{N_i}^t \mathbf{X}_{u_i}, \\
v_1 &= \sum_{i \in \{i_1, \dots, i_k\}} \mathbf{1}_{N_i - n_i}^t (\mathbf{V}_{u_i} - \mathbf{V}_{s_{u_i}}^t \mathbf{V}_{s_i}^{-1} \mathbf{V}_{s_{u_i}}) \mathbf{1}_{N_i - n_i} + \sum_{i \notin \{i_1, \dots, i_k\}} \mathbf{1}_{N_i}^t \mathbf{V}_{u_i} \mathbf{1}_{N_i} + \\
&+ \sum_{i \in \{i_1, \dots, i_k\}} \mathbf{1}_{N_i - n_i}^t (\mathbf{X}_{u_i} - \mathbf{V}_{s_{u_i}}^t \mathbf{V}_{s_i}^{-1} \mathbf{V}_{s_{u_i}}) \alpha_i \mathbf{S}_i^{-1} (\mathbf{X}_{u_i} - \mathbf{V}_{s_{u_i}}^t \mathbf{V}_{s_i}^{-1} \mathbf{V}_{s_{u_i}})^t \mathbf{1}_{N_i - n_i} + \\
&+ \sum_{i \in \{i_1, \dots, i_k\}} \mathbf{1}_{N_i - n_i}^t (\mathbf{X}_{u_i} - \mathbf{V}_{s_{u_i}}^t \mathbf{V}_{s_i}^{-1} \mathbf{V}_{s_{u_i}}) \alpha_i \mathbf{S}_i^{-1} \mathbf{S}_w^{-1} \mathbf{Z}_w^t, \\
v_2 &= \sum_{i \notin \{i_1, i_2, \dots, i_k\}} \mathbf{1}_{N_i}^t \mathbf{X}_{u_i} \mathbf{B}_\beta \mathbf{X}_{u_i}^t \mathbf{1}_{N_i} + \sum_{i \notin \{i_1, i_2, \dots, i_k\}} \mathbf{1}_{N_i}^t \mathbf{X}_{u_i} \mathbf{B}_\beta \mathbf{S}_w^{-1} \mathbf{Z}_w^t.
\end{aligned}$$

**Demostración.** La demostración de los apartados (a), (d) y (e) son inmediatas (salvo operaciones con matrices algo largas) a partir del teorema 3.5, tomando:  $m = 2$ ,  $\mathbf{Y} = (\mathbf{Y}_1^t, \mathbf{Y}_2^t, \dots, \mathbf{Y}_K^t)^t$ ,  $\mathbf{Y}_i = (Y_{i1}, Y_{i2}, \dots, Y_{iN_i})^t$ ,  $i = 1, 2, \dots, K$ ,  $\theta_1 = (\beta_1^t, \beta_2^t, \dots, \beta_K^t)^t$ ,  $\theta_2 = \beta$ ,  $\mathbf{A}_1 = \text{diag}_{i=1}^K \{\mathbf{X}_i\}$ ,  $\mathbf{X}_i = (\mathbf{x}_{i1}^t, \mathbf{x}_{i2}^t, \dots, \mathbf{x}_{iN_i}^t)^t$ ,  $i = 1, 2, \dots, K$ ,  $\mathbf{x}_{ij} = (x_{ij1}, x_{ij2}, \dots, x_{ijp})$ ,  $j = 1, 2, \dots, N_i$ ,  $i = 1, 2, \dots, K$ ,  $\mathbf{A}_2 = \underbrace{(\mathbf{I}_p, \dots, \mathbf{I}_p)}_K^t$ ,  $\mathbf{A}_3 = \beta_0$ ,  $\mathbf{V}_1 =$

$\text{diag}_{i=1}^K \{\mathbf{V}_i\}$ ,  $\mathbf{V}_2 = \text{diag}_{i=1}^K \{\mathbf{B}_\beta\}$  y  $\mathbf{V}_3 = \mathbf{B}_{\beta_0}$ , reordenando las filas de las matrices de elementos no observados colocando en las  $k$  primeras filas las filas correspondientes a las unidades muestreadas, aplicando el resultado (a) del lema B.1 sobre el vector  $\mathbf{d} = (d_1^t, d_2^t, \dots, d_K^t)^t$  y la matriz  $\mathbf{D}$ , de donde se obtiene que:

$$\begin{aligned}
\mathbf{d}_j &= \begin{cases} \mathbf{X}_{s_{ij}}^t \mathbf{V}_{s_{ij}}^{-1} \mathbf{y}_{s_{ij}} + \mathbf{B}_\beta^{-1} \beta_0 - \mathbf{B}_\beta^{-1} (\mathbf{K} \mathbf{B}_\beta^{-1} + \mathbf{B}_{\beta_0}^{-1})^{-1} \mathbf{K} \mathbf{B}_\beta^{-1} \beta_0, & j = 1, 2, \dots, k, \\ \mathbf{B}_\beta^{-1} \beta_0 - \mathbf{B}_\beta^{-1} (\mathbf{K} \mathbf{B}_\beta^{-1} + \mathbf{B}_{\beta_0}^{-1})^{-1} \mathbf{K} \mathbf{B}_\beta^{-1} \beta_0, & j = k + 1, \dots, K, \end{cases} \\
\mathbf{D} &= \text{diag}\{\text{diag}_{j=1}^k \{\mathbf{D}_{i_j}\}, \text{diag}_{j=k+1}^K \{\mathbf{B}_\beta\}\} - (\mathbf{D}_{i_1}, \mathbf{D}_{i_2}, \dots, \mathbf{D}_{i_k}, \underbrace{\mathbf{B}_\beta, \dots, \mathbf{B}_\beta}_{K-k})^t \times \\
&\quad \times \mathbf{B}_\beta^{-1} \mathbf{S} \mathbf{B}_\beta^{-1} (\mathbf{D}_{i_1}, \mathbf{D}_{i_2}, \dots, \mathbf{D}_{i_k}, \underbrace{\mathbf{B}_\beta, \dots, \mathbf{B}_\beta}_{K-k}),
\end{aligned}$$

con  $\mathbf{D}_i^{-1} = \mathbf{X}_{s_i}^t \mathbf{V}_{s_i}^{-1} \mathbf{X}_{s_i} + \mathbf{B}_\beta^{-1}$ ,  $i \in \{i_1, i_2, \dots, i_k\}$  y denotando por  $\widehat{\beta}_i^B$  el  $p$ -vector de  $\widehat{\theta}_1^B$  correspondiente a la unidad  $i$ . Los restantes resultados son inmediatos a partir de la distribución predictiva anterior aplicando que la marginal de una normal y la combinación lineal de normales es normal y la definición de estimador Bayes adaptados a la estimación de la media de la unidad  $i$  con  $i = 1, 2, \dots, K$ .

C.S.Q.D.

Notemos que la introducción de la segunda jerarquía en el modelo ha alterado sustancialmente la forma de las distribuciones predictivas (en comparación con las obtenidas para el modelo con una jerarquía Normal dadas en el teorema 2.5) para el vector poblacional no observado, y que los vectores no observados de cada unidad han dejado de ser independientes entre sí. Además la influencia de la segunda jerarquía también se detecta en la forma del estimador Bayes de la pendiente de regresión que es, para  $i \in \{i_1, i_2, \dots, i_k\}$ , una “combinación lineal” a través de  $\alpha_i$  del estimador clásico de regresión de cada unidad  $\hat{\beta}_i$  y de  $\hat{\beta}_w$  (en vez de  $\beta_0$ ), que es una ponderación de los  $\hat{\beta}_i$ ,  $i \in \{i_1, i_2, \dots, i_k\}$  (a través de  $\alpha_i$ ,  $i \in \{i_1, i_2, \dots, i_k\}$ ) y de  $\beta_0$ , y para  $i \notin \{i_1, i_2, \dots, i_k\}$  es igual a  $\hat{\beta}_w$ .

El factor  $\alpha_i$ ,  $i \in \{i_1, i_2, \dots, i_k\}$  de la combinación lineal anterior y de la ponderación de los  $\hat{\beta}_i$ ,  $i \in \{i_1, i_2, \dots, i_k\}$  para la obtención de  $\hat{\beta}_w$ , es un factor de encogimiento ( $\alpha_i = (\mathbf{I}_p - \alpha_i)\mathbf{B}_\beta\mathbf{S}_i$ ,  $i \in \{i_1, i_2, \dots, i_k\}$ ), que puede interpretarse como el porcentaje de variación de las pendientes de cada unidad muestreada no explicado por la muestra, ya que, para  $i \in \{i_1, i_2, \dots, i_k\}$ :

$$\alpha_i = [\text{Var}(\beta_i|\beta) - \text{Var}(\beta_i|(s, y_s), \beta_0)]\text{Var}^{-1}(\beta_i|\beta). \quad (3.15)$$

La expresión (3.15), se obtiene a partir de la definición de  $\alpha_i$ ,  $i \in \{i_1, i_2, \dots, i_k\}$ , dada en (3.14) y de la expresión de  $\text{Var}(\beta_i|(s, y_s), \beta_0) = \alpha_i\mathbf{S}_i^{-1}$ ,  $i \in \{i_1, i_2, \dots, i_k\}$ , obtenida de forma intermedia en la demostración anterior.

### 3.2.2 Modelo Bayes de Regresión en dos etapas con dos jerarquías normal-gamma

El modelo Bayes de Regresión en dos etapas con dos jerarquías Normal-Gamma, que definiremos a continuación, se aplica para modelizar una característica continua de una población estructurada en dos etapas cuando la varianza de la población es desconocida y el cociente entre varianzas conocido, y deseamos modelizar una relación de intercambiabilidad entre las distintas unidades.

**Definición 3.2** *Dado el vector poblacional  $\mathbf{Y} = (\mathbf{Y}_1^t, \mathbf{Y}_2^t, \dots, \mathbf{Y}_K^t)^t$  con  $\mathbf{Y}_i = (Y_{i1}, Y_{i2}, \dots, Y_{iN_i})^t$ ,  $i = 1, 2, \dots, K$  para el cual asociada a cada variable  $Y_{ij}$  tenemos una información auxiliar representada por un vector  $\mathbf{x}_{ij} = (x_{ij1}, \dots, x_{ijp})$  conocido,  $j = 1, 2, \dots, N_i$ ,  $i = 1, 2, \dots, K$ , llamaremos modelo Bayes de Regresión en dos etapas con dos jerarquías Normal-Gamma al definido por:*

$$\begin{aligned} \mathbf{Y}_i|\mathbf{X}_i, \beta_i, \sigma^2 &\sim N(\mathbf{X}_i\beta_i, \sigma^2\mathbf{W}_i), \text{ independientes, } i = 1, 2, \dots, K, \\ \beta_i|\beta, \sigma^2 &\sim N_p(\beta, \sigma^2\mathbf{C}_\beta), \text{ i.i.d., } i = 1, 2, \dots, K, \\ \beta|\sigma^2 &\sim N_p(\beta_0, \sigma^2\mathbf{C}_{\beta_0}), \\ \sigma^2 &\sim \text{Ga}^{-1}(a_0, b_0) \end{aligned} \quad (3.16)$$

donde  $\mathbf{X}_i = (\mathbf{x}_{i1}^t, \mathbf{x}_{i2}^t, \dots, \mathbf{x}_{iN_i}^t)^t$ , es una matriz de rango  $p$ ,  $i = 1, 2, \dots, K$ ,  $\beta_0$  es un  $p$ -vector conocido,  $\mathbf{W}_i$ ,  $i = 1, 2, \dots, K$ ,  $\mathbf{C}_\beta$  y  $\mathbf{C}_{\beta_0}$  son matrices simétricas y definidas positivas conocidas de dimensiones  $N_i \times N_i$ ,  $i = 1, 2, \dots, K$ ,  $p \times p$  y  $p \times p$  respectivamente y  $a_0$  y  $b_0$  escalares positivos conocidos.

Este modelo presenta como caso particular para  $p = 1$ ,  $\mathbf{X}_i = \mathbf{I}_{N_i}$ ,  $i = 1, 2, \dots, K$ , el modelo Bayes Básico en dos etapas con dos jerarquías normal-gamma y como una particularización de éste, el modelo estudiado por Murgui(82)[76] (ver el apartado estimadores Bayes de la media poblacional de la subsección 1.4.2).

El siguiente teorema obtiene los resultados inferenciales de interés para este modelo. Sobre estos resultados se pueden trasladar, de manera inmediata las reflexiones sobre la estimación de la pendiente de regresión que realizábamos sobre los resultados del teorema 3.1 para el correspondiente modelo normal.

**Teorema 3.2** Consideremos el modelo Bayes de Regresión en dos etapas con dos jerarquías Normal-Gamma dado en (3.16), entonces se tiene que:

(a)  $\mathbf{Y}_u|(s, \mathbf{y}_s)$  se distribuye  $t(N - n)$ -variante con  $2a_1$  grados de libertad, con para  $a_1 > 1$  media:

$$E(\mathbf{Y}_u|(s, \mathbf{y}_s)) = \begin{cases} \mathbf{X}_{u_i} \hat{\beta}_i^B + \mathbf{W}_{su_i}^t \mathbf{W}_{s_i}^{-1} (\mathbf{y}_{s_i} - \mathbf{X}_{s_i} \hat{\beta}_i^B), & i \in \{i_1, i_2, \dots, i_k\} \\ \mathbf{X}_{u_i} \hat{\beta}_i^B & i \notin \{i_1, i_2, \dots, i_k\} \end{cases} \quad (3.17)$$

y matriz de varianzas y covarianzas,

\* para  $i \in \{i_1, i_2, \dots, i_k\}$

$$\begin{aligned} Var(\mathbf{Y}_{u_i}|(s, \mathbf{y}_s)) &= \frac{b_1}{a_1 - 1} \{ \mathbf{W}_{u_i} - \mathbf{W}_{su_i}^t \mathbf{W}_{s_i}^{-1} \mathbf{W}_{su_i} + \\ &+ (\mathbf{X}_{u_i} - \mathbf{W}_{su_i}^t \mathbf{W}_{s_i}^{-1} \mathbf{X}_{s_i}) \alpha_i \mathbf{S}_i^{-1} \mathbf{C}_\beta^{-1} \mathbf{S}_w^{-1} \mathbf{S}_i^{-1} \alpha_i^t (\mathbf{X}_{u_i} - \mathbf{W}_{su_i}^t \mathbf{W}_{s_i}^{-1} \mathbf{X}_{s_i})^t + \\ &+ (\mathbf{X}_{u_i} - \mathbf{W}_{su_i}^t \mathbf{W}_{s_i}^{-1} \mathbf{X}_{s_i}) \alpha_i \mathbf{S}_i^{-1} (\mathbf{X}_{u_i} - \mathbf{W}_{su_i}^t \mathbf{W}_{s_i}^{-1} \mathbf{X}_{s_i})^t \}, \end{aligned} \quad (3.18)$$

\* para  $i \notin \{i_1, i_2, \dots, i_k\}$ ,

$$Var(\mathbf{Y}_{u_i}|(s, \mathbf{y}_s)) = \frac{b_1}{a_1 - 1} \{ \mathbf{W}_{u_i} + \mathbf{X}_{u_i} \mathbf{C}_\beta \mathbf{X}_{u_i}^t + \mathbf{X}_{u_i} \mathbf{S}_w^{-1} \mathbf{C}_\beta \mathbf{X}_{u_i}^t \}, \quad (3.19)$$

\* para  $i \in \{i_1, i_2, \dots, i_k\}$ ,  $j \notin \{i_1, i_2, \dots, i_k\}$ ,

$$Cov(\mathbf{Y}_{u_i}, \mathbf{Y}_{u_j}|(s, \mathbf{y}_s)) = \frac{b_1}{a_1 - 1} (\mathbf{X}_{u_i} - \mathbf{W}_{su_i}^t \mathbf{W}_{s_i}^{-1} \mathbf{X}_{s_i}) \alpha_i \mathbf{S}_i^{-1} \mathbf{C}_\beta^{-1} \mathbf{S}_w^{-1} \mathbf{C}_\beta \mathbf{X}_{u_j}^t, \quad (3.20)$$

\* para  $i \neq j, i, j \in \{i_1, i_2, \dots, i_k\}$ ,

$$\begin{aligned} \text{Cov}(\mathbf{Y}_{u_i}, \mathbf{Y}_{u_j} | (s, \mathbf{y}_s)) &= \\ \frac{b_1}{a_1 - 1} (\mathbf{X}_{u_i} - \mathbf{W}_{su_i}^t \mathbf{W}_{s_i}^{-1} \mathbf{X}_{s_i}) \alpha_i \mathbf{S}_i^{-1} \mathbf{C}_\beta^{-1} \mathbf{S}_w^{-1} \mathbf{S}_j^{-1} \alpha_j^t (\mathbf{X}_{u_j} - \mathbf{W}_{su_j}^t \mathbf{W}_{s_j}^{-1} \mathbf{X}_{s_j})^t, \end{aligned} \quad (3.21)$$

\* para  $i \neq j, i, j \notin \{i_1, i_2, \dots, i_k\}$ ,

$$\text{Cov}(\mathbf{Y}_{u_i}, \mathbf{Y}_{u_j} | (s, \mathbf{y}_s)) = \frac{b_1}{a_1 - 1} \mathbf{X}_{u_i} \mathbf{S}_w^{-1} \mathbf{C}_\beta \mathbf{X}_{u_j}^t, \quad (3.22)$$

(b)  $\bar{Y}_{u_i} | (s, \mathbf{y}_s)$ ,  $i = 1, 2, \dots, K$ , se distribuye  $t$  con  $2a_1$  grados de libertad y media y varianza, para  $a_1 > 1$ , dadas por:

\* para  $i \in \{i_1, i_2, \dots, i_k\}$

$$\begin{aligned} E(\bar{Y}_{u_i} | (s, \mathbf{y}_s)) &= \frac{1}{N_i - n_i} \mathbf{1}_{N_i - n_i}^t [\mathbf{X}_{u_i} \hat{\beta}_i^B + \mathbf{W}_{su_i}^t \mathbf{W}_{s_i}^{-1} (\mathbf{y}_{s_i} - \mathbf{X}_{s_i} \hat{\beta}_i^B)], \quad (3.23) \\ \text{Var}(\bar{Y}_{u_i} | (s, \mathbf{y}_s)) &= \frac{1}{(N_i - n_i)^2} \frac{b_1}{a_1 - 1} \{ \mathbf{1}_{N_i - n_i}^t (\mathbf{W}_{u_i} - \mathbf{W}_{su_i}^t \mathbf{W}_{s_i}^{-1} \mathbf{W}_{su_i}) \mathbf{1}_{N_i - n_i} + \\ &+ \mathbf{1}_{N_i - n_i}^t (\mathbf{X}_{u_i} - \mathbf{W}_{su_i}^t \mathbf{W}_{s_i}^{-1} \mathbf{X}_{s_i}) \alpha_i \mathbf{S}_i^{-1} \mathbf{C}_\beta^{-1} \mathbf{S}_w^{-1} \mathbf{S}_i^{-1} \alpha_i^t (\mathbf{X}_{u_i} - \mathbf{W}_{su_i}^t \mathbf{W}_{s_i}^{-1} \mathbf{X}_{s_i})^t \mathbf{1}_{N_i - n_i} + \\ &+ \mathbf{1}_{N_i - n_i}^t (\mathbf{X}_{u_i} - \mathbf{W}_{su_i}^t \mathbf{W}_{s_i}^{-1} \mathbf{X}_{s_i}) \alpha_i \mathbf{S}_i^{-1} (\mathbf{X}_{u_i} - \mathbf{W}_{su_i}^t \mathbf{W}_{s_i}^{-1} \mathbf{X}_{s_i})^t \mathbf{1}_{N_i - n_i} \}, \end{aligned}$$

\* para  $i \notin \{i_1, i_2, \dots, i_k\}$

$$\begin{aligned} E(\bar{Y}_{u_i} | (s, \mathbf{y}_s)) &= \frac{1}{N_i} \mathbf{1}_{N_i}^t \mathbf{X}_{u_i} \hat{\beta}_i^B, \quad (3.24) \\ \text{Var}(\bar{Y}_{u_i} | (s, \mathbf{y}_s)) &= \frac{1}{N_i^2} \frac{b_1}{a_1 - 1} \{ \mathbf{1}_{N_i}^t (\mathbf{W}_{u_i} + \mathbf{X}_{u_i} \mathbf{C}_\beta \mathbf{X}_{u_i}^t) \mathbf{1}_{N_i} + \\ &+ \mathbf{1}_{N_i}^t \mathbf{X}_{u_i} \mathbf{S}_w^{-1} \mathbf{C}_\beta \mathbf{X}_{u_i}^t \mathbf{1}_{N_i} \}, \end{aligned}$$

(c) el estimador Bayes para la media de la unidad  $i$ ,  $i = 1, 2, \dots, K$  y su varianza condicionada a la muestra, vienen dados, para  $a_1 > 1$ , por:

\* para  $i \in \{i_1, i_2, \dots, i_k\}$

$$\begin{aligned} \hat{Y}_i &= \frac{n_i}{N_i} \bar{y}_{s_i} + \frac{1}{N_i} \mathbf{1}_{N_i - n_i}^t [\mathbf{X}_{u_i} \hat{\beta}_i^B + \mathbf{W}_{su_i}^t \mathbf{W}_{s_i}^{-1} (\mathbf{y}_{s_i} - \mathbf{X}_{s_i} \hat{\beta}_i^B)], \quad (3.25) \\ \text{Var}(\bar{Y}_i | (s, \mathbf{y}_s)) &= \frac{1}{N_i^2} \frac{b_1}{a_1 - 1} \{ \mathbf{1}_{N_i - n_i}^t (\mathbf{W}_{u_i} - \mathbf{W}_{su_i}^t \mathbf{W}_{s_i}^{-1} \mathbf{W}_{su_i}) \mathbf{1}_{N_i - n_i} + \\ &+ \mathbf{1}_{N_i - n_i}^t (\mathbf{X}_{u_i} - \mathbf{W}_{su_i}^t \mathbf{W}_{s_i}^{-1} \mathbf{X}_{s_i}) \alpha_i \mathbf{S}_i^{-1} \mathbf{C}_\beta^{-1} \mathbf{S}_w^{-1} \mathbf{S}_i^{-1} \alpha_i^t (\mathbf{X}_{u_i} - \mathbf{W}_{su_i}^t \mathbf{W}_{s_i}^{-1} \mathbf{X}_{s_i})^t \mathbf{1}_{N_i - n_i} + \\ &+ \mathbf{1}_{N_i - n_i}^t (\mathbf{X}_{u_i} - \mathbf{W}_{su_i}^t \mathbf{W}_{s_i}^{-1} \mathbf{X}_{s_i}) \alpha_i \mathbf{S}_i^{-1} (\mathbf{X}_{u_i} - \mathbf{W}_{su_i}^t \mathbf{W}_{s_i}^{-1} \mathbf{X}_{s_i})^t \mathbf{1}_{N_i - n_i} \}, \end{aligned}$$

\* para  $i \notin \{i_1, i_2, \dots, i_k\}$

$$\begin{aligned}\widehat{Y}_i &= \frac{1}{N_i} \mathbf{1}_{N_i}^t \mathbf{X}_{u_i} \widehat{\beta}_i^B, \\ \text{Var}(\overline{Y}_i | (s, \mathbf{y}_s)) &= \frac{1}{N_i^2} \frac{b_1}{a_1 - 1} \{ \mathbf{1}_{N_i}^t (\mathbf{W}_{u_i} + \mathbf{X}_{u_i} \mathbf{C}_\beta \mathbf{X}_{u_i}^t) \mathbf{1}_{N_i} + \\ &\quad + \mathbf{1}_{N_i}^t \mathbf{X}_{u_i} \mathbf{S}_w^{-1} \mathbf{C}_\beta \mathbf{X}_{u_i}^t \mathbf{1}_{N_i} \},\end{aligned}\quad (3.26)$$

(d)  $\overline{Y}_u | (s, \mathbf{y}_s)$  se distribuye  $t$  con  $2a_1$  grados de libertad y media y varianza, para  $a_1 > 1$ , dadas por:

$$\begin{aligned}E(\overline{Y}_u | (s, \mathbf{y}_s)) &= \frac{1}{N-n} \sum_{i \in \{i_1, \dots, i_k\}} \mathbf{1}_{N_i - n_i}^t [\mathbf{X}_{u_i} \widehat{\beta}_i^B + \mathbf{W}_{s u_i}^t \mathbf{W}_{s_i}^{-1} (\mathbf{y}_{s_i} - \mathbf{X}_{s_i} \widehat{\beta}_i^B)] + \\ &\quad + \frac{1}{N-n} \sum_{i \notin \{i_1, \dots, i_k\}} \mathbf{1}_{N_i}^t \mathbf{X}_{u_i} \widehat{\beta}_i^B, \\ \text{Var}(\overline{Y}_u | (s, \mathbf{y}_s)) &= \frac{1}{(N-n)^2} \frac{b_1}{a_1 - 1} \{w_1 + w_2\},\end{aligned}\quad (3.27)$$

y (e) el estimador Bayes para la media poblacional y su varianza condicionada a la muestra, vienen dados, para  $a_1 > 1$ , por:

$$\begin{aligned}\widehat{Y} &= \frac{n}{N} \overline{y}_s + \frac{1}{N} \sum_{i \in \{i_1, \dots, i_k\}} \mathbf{1}_{N_i - n_i}^t [\mathbf{X}_{u_i} \widehat{\beta}_i^B + \mathbf{W}_{s u_i}^t \mathbf{W}_{s_i}^{-1} (\mathbf{y}_{s_i} - \mathbf{X}_{s_i} \widehat{\beta}_i^B)] + \\ &\quad + \frac{1}{N} \sum_{i \notin \{i_1, \dots, i_k\}} \mathbf{1}_{N_i}^t \mathbf{X}_{u_i} \widehat{\beta}_i^B, \\ \text{Var}(\overline{Y} | (s, \mathbf{y}_s)) &= \frac{1}{N^2} \frac{b_1}{a_1 - 1} \{w_1 + w_2\},\end{aligned}\quad (3.28)$$

donde  $\mathbf{W}_{s_i} = \sigma^{-2} \text{Var}_{\mathbf{Y} | \beta}(\mathbf{Y}_{s_i})$ ,  $\mathbf{W}_{u_i} = \sigma^{-2} \text{Var}_{\mathbf{Y} | \beta}(\mathbf{Y}_{u_i})$ ,  $\mathbf{W}_{s u_i} = \sigma^{-2} \text{Cov}_{\mathbf{Y} | \beta}(\mathbf{Y}_{s_i}, \mathbf{Y}_{u_i})$ ,  $\overline{y}_s = \frac{\sum_{(i,j) \in s} y_{ij}}{n}$  es la media muestral,  $\overline{y}_{s_i}$ , es la media muestral de la unidad  $i$ ,  $i \in \{i_1, i_2, \dots, i_k\}$ ,  $\mathbf{S}_i = \mathbf{X}_{s_i}^t \mathbf{W}_{s_i}^{-1} \mathbf{X}_{s_i}$ ,  $i \in \{i_1, i_2, \dots, i_k\}$ , y

$$\begin{aligned}\alpha_i &= (\mathbf{X}_{s_i}^t \mathbf{W}_{s_i}^{-1} \mathbf{X}_{s_i} + \mathbf{C}_\beta^{-1})^{-1} \mathbf{X}_{s_i}^t \mathbf{W}_{s_i}^{-1} \mathbf{X}_{s_i}, \quad i \in \{i_1, i_2, \dots, i_k\}, \\ \widehat{\beta}_i &= \mathbf{S}_i^{-1} \mathbf{X}_{s_i}^t \mathbf{W}_{s_i}^{-1} \mathbf{y}_{s_i}, \quad i \in \{i_1, i_2, \dots, i_k\}, \\ \mathbf{S}_w &= \sum_{i \in \{i_1, i_2, \dots, i_k\}} \alpha_i + \mathbf{C}_\beta \mathbf{C}_{\beta_0}^{-1},\end{aligned}\quad (3.29)$$

$$\begin{aligned}
\widehat{\beta}_w &= \mathbf{S}_w^{-1} \left( \sum_{i \in \{i_1, i_2, \dots, i_k\}} \alpha_i \widehat{\beta}_i + \mathbf{C}_\beta \mathbf{C}_{\beta_0}^{-1} \beta_0 \right), \\
\widehat{\beta}_i^B &= E(\beta_i | (s, y_s)) = \begin{cases} \alpha_i \widehat{\beta}_i + (\mathbf{I}_p - \alpha_i) \widehat{\beta}_w, & i \in \{i_1, i_2, \dots, i_k\}, \\ \widehat{\beta}_w, & i \notin \{i_1, i_2, \dots, i_k\}, \end{cases} \\
\mathbf{Z}_w &= \sum_{i \in \{i_1, i_2, \dots, i_k\}} \mathbf{1}_{N_i - n_i}^t (\mathbf{X}_{u_i} - \mathbf{W}_{s_{u_i}}^t \mathbf{W}_{s_i}^{-1} \mathbf{X}_{s_i}) \alpha_i \mathbf{S}_i^{-1} \mathbf{C}_\beta^{-1} + \sum_{i \notin \{i_1, i_2, \dots, i_k\}} \mathbf{1}_{N_i}^t \mathbf{X}_{u_i}, \\
a_1 &= a_0 + \frac{n}{2}, \\
b_1 &= b_0 + b_s + \frac{1}{2} \sum_{i \in \{i_1, i_2, \dots, i_k\}} (y_{s_i} - \mathbf{X}_{s_i} \beta_0)^t (\mathbf{W}_{s_i} + \mathbf{X}_{s_i} \mathbf{C}_\beta \mathbf{X}_{s_i}^t)^{-1} (y_{s_i} - \mathbf{X}_{s_i} \beta_0), \\
b_s &= \frac{1}{2} \left[ \sum_{i \in \{i_1, i_2, \dots, i_k\}} (y_{s_i} - \mathbf{X}_{s_i} \beta_0)^t (\mathbf{W}_{s_i} + \mathbf{X}_{s_i} \mathbf{C}_\beta \mathbf{X}_{s_i}^t)^{-1} \mathbf{X}_{s_i} \right] \times \\
&\quad \times \left[ \sum_{i \in \{i_1, i_2, \dots, i_k\}} \mathbf{X}_{s_i}^t (\mathbf{W}_{s_i} + \mathbf{X}_{s_i} \mathbf{C}_\beta \mathbf{X}_{s_i}^t)^{-1} \mathbf{X}_{s_i} + \mathbf{C}_{\beta_0}^{-1} \right]^{-1} \times \\
&\quad \times \left[ \sum_{i \in \{i_1, i_2, \dots, i_k\}} (y_{s_i} - \mathbf{X}_{s_i} \beta_0)^t (\mathbf{W}_{s_i} + \mathbf{X}_{s_i} \mathbf{C}_\beta \mathbf{X}_{s_i}^t)^{-1} \mathbf{X}_{s_i} \right]^t \\
w_1 &= \sum_{i \in \{i_1, \dots, i_k\}} \mathbf{1}_{N_i - n_i}^t (\mathbf{W}_{u_i} - \mathbf{W}_{s_{u_i}}^t \mathbf{W}_{s_i}^{-1} \mathbf{W}_{s_{u_i}}) \mathbf{1}_{N_i - n_i} + \sum_{i \notin \{i_1, \dots, i_k\}} \mathbf{1}_{N_i}^t \mathbf{W}_{u_i} \mathbf{1}_{N_i} + \\
&\quad + \sum_{i \in \{i_1, \dots, i_k\}} \mathbf{1}_{N_i - n_i}^t (\mathbf{X}_{u_i} - \mathbf{W}_{s_{u_i}}^t \mathbf{W}_{s_i}^{-1} \mathbf{W}_{s_{u_i}}) \alpha_i \mathbf{S}_i^{-1} (\mathbf{X}_{u_i} - \mathbf{W}_{s_{u_i}}^t \mathbf{W}_{s_i}^{-1} \mathbf{W}_{s_{u_i}})^t \mathbf{1}_{N_i - n_i} + \\
&\quad + \sum_{i \in \{i_1, \dots, i_k\}} \mathbf{1}_{N_i - n_i}^t (\mathbf{X}_{u_i} - \mathbf{W}_{s_{u_i}}^t \mathbf{W}_{s_i}^{-1} \mathbf{W}_{s_{u_i}}) \alpha_i \mathbf{S}_i^{-1} \mathbf{S}_w^{-1} \mathbf{Z}_w^t, \\
w_2 &= \sum_{i \notin \{i_1, i_2, \dots, i_k\}} \mathbf{1}_{N_i}^t \mathbf{X}_{u_i} \mathbf{C}_\beta \mathbf{X}_{u_i}^t \mathbf{1}_{N_i} + \sum_{i \notin \{i_1, i_2, \dots, i_k\}} \mathbf{1}_{N_i}^t \mathbf{X}_{u_i} \mathbf{C}_\beta \mathbf{S}_w^{-1} \mathbf{Z}_w^t.
\end{aligned}$$

**Demostración.** La demostración de los apartados (a), (d) y (e) son inmediatas (salvo operaciones con matrices algo largas) a partir del teorema 3.6, tomando:  $m = 2$ ,  $\mathbf{Y} = (\mathbf{Y}_1^t, \mathbf{Y}_2^t, \dots, \mathbf{Y}_K^t)^t$ ,  $\mathbf{Y}_i = (Y_{i1}, Y_{i2}, \dots, Y_{iN_i})^t$ ,  $i = 1, 2, \dots, K$ ,  $\theta_1 = (\beta_1^t, \beta_2^t, \dots, \beta_K^t)^t$ ,  $\theta_2 = \beta$ ,  $\mathbf{A}_1 = \text{diag}_{i=1}^K \{\mathbf{X}_i\}$ ,  $\mathbf{X}_i = (\mathbf{x}_{i1}^t, \mathbf{x}_{i2}^t, \dots, \mathbf{x}_{iN_i}^t)^t$ ,  $i = 1, 2, \dots, K$ ,  $\mathbf{x}_{ij} = (x_{ij1}, x_{ij2}, \dots, x_{ijp})$ ,  $j = 1, 2, \dots, N_i$ ,  $i = 1, 2, \dots, K$ ,  $\mathbf{A}_2 = \underbrace{(\mathbf{I}_p, \dots, \mathbf{I}_p)}_K^t$ ,  $\mathbf{A}_3 = \beta_0$ ,  $\mathbf{W}_1 =$

$\text{diag}_{i=1}^K \{\mathbf{W}_i\}$ ,  $\mathbf{W}_2 = \text{diag}_{i=1}^K \{\mathbf{C}_\beta\}$  y  $\mathbf{W}_3 = \mathbf{C}_{\beta_0}$ , reordenando las filas de las matrices de elementos no observados colocando en las  $k$  primeras filas las filas correspondientes a las unidades muestreadas, aplicando el resultado (a) del lema B.1 de manera totalmente análoga al teorema anterior sobre  $\mathbf{d}$ ,  $\mathbf{D}$  y  $b_1$ , realizando las operaciones de matrices necesarias y denotando por  $\widehat{\beta}_i^B$  el  $p$ -vector de  $\widehat{\theta}_1^B$  correspondiente a la unidad  $i$ . El resultado (b) se comprueba<sup>1</sup> aplicando el resultado (a) del lema B.9 sobre la

<sup>1</sup>Una demostración alternativa consiste en aplicar sobre la distribución predictiva del vector poblacional no observado obtenida en la primera parte de la demostración de la propiedad de que la marginal de una  $t$

distribución de  $\bar{Y}_{u_i} | (s, \mathbf{y}_s), \sigma^2$  que se obtiene directamente de (3.8) y (3.9), tomando  $\mathbf{V}_i = \{\sigma^2 \mathbf{W}_i\}$ ,  $i = 1, 2, \dots, K$  y  $\mathbf{B}_\beta = \{\sigma^2 \mathbf{C}_\beta\}$  y sobre la distribución (3.79) de la demostración del teorema 3.6 y el resultado (c) se obtiene a partir de (b) aplicando la definición de estimador Bayes adaptada a la estimación de la media de la unidad  $i$  con  $i = 1, 2, \dots, K$ .

C.S.Q.D.

### 3.2.3 Modelo Bayes de Regresión en dos etapas con dos jerarquías con cociente entre varianzas desconocido

El modelo Bayes de Regresión en dos etapas con dos jerarquías Normal-Gamma, que hemos estudiado en la subsección anterior, tiene la limitación de asumir que todas las varianzas (de la característica en estudio, de las pendientes de cada unidad y de la media de dichas pendientes), son proporcionales a un mismo escalar  $\sigma^2$ .

Esta limitación, necesaria para mantenernos en el terreno de los modelos conjugados, es superada por el siguiente modelo no conjugado, que se aplica para realizar inferencias sobre poblaciones estructuradas en dos etapas, cuando las varianzas son desconocidas y el cociente entre varianzas es asimismo desconocido, cuando deseamos introducir una relación de intercambiabilidad entre las pendientes de las distintas unidades.

**Definición 3.3** Dado el vector poblacional  $\mathbf{Y} = (\mathbf{Y}_1^t, \mathbf{Y}_2^t, \dots, \mathbf{Y}_K^t)^t$  con  $\mathbf{Y}_i = (Y_{i1}, Y_{i2}, \dots, Y_{iN_i})^t$ ,  $i = 1, 2, \dots, K$  para el cual asociada a cada variable  $Y_{ij}$  tenemos una información auxiliar representada por un vector  $\mathbf{x}_{ij} = (x_{ij1}, \dots, x_{ijp})$  conocido,  $j = 1, 2, \dots, N_i$ ,  $i = 1, 2, \dots, K$ , llamaremos modelo Bayes de Regresión en dos etapas con dos jerarquías con cociente entre varianzas desconocido al definido por:

$$\begin{aligned}
 \mathbf{Y}_i | \beta_i, \sigma_i^2 &\sim N(\mathbf{X}_i \beta_i, \sigma_i^2 \mathbf{I}_{N_i}), \text{ independientes, } i = 1, 2, \dots, K, \\
 \sigma_i^2 &\sim Ga^{-1}(a_0, b_0), \text{ i.i.d., } i = 1, 2, \dots, K, \\
 \beta_i | \beta, \sigma_\beta^2 &\sim N_p(\beta, \sigma_\beta^2 \mathbf{C}_\beta), \text{ i.i.d., } i = 1, 2, \dots, K, \\
 \sigma_\beta^2 &\sim Ga^{-1}(a_\beta, b_\beta), \\
 \beta | \sigma_{\beta_0}^2 &\sim N_p(\beta_0, \sigma_{\beta_0}^2 \mathbf{C}_{\beta_0}), \\
 \sigma_{\beta_0}^2 &\sim Ga^{-1}(a_{\beta_0}, b_{\beta_0}),
 \end{aligned} \tag{3.30}$$

donde  $\mathbf{X}_i = (\mathbf{x}_{i1}^t, \mathbf{x}_{i2}^t, \dots, \mathbf{x}_{iN_i}^t)^t$ , es una matriz de rango  $p$ ,  $i = 1, 2, \dots, K$ ,  $\beta$  es un  $p$ -vector conocido,  $\mathbf{C}_\beta$  y  $\mathbf{C}_{\beta_0}$  matrices simétricas definidas positivas conocidas de dimensión  $p \times p$  y  $a_0, b_0, a_\beta$  y  $b_\beta$  escalares positivos conocidos.

El modelo Bayes de Regresión en dos etapas con dos jerarquías que acabamos de definir se sitúa en la línea de los otros dos modelos Bayes con cociente entre varianzas

---

y la combinación lineal de  $t$  son  $t$

desconocido que estudiábamos en el capítulo anterior. Como aquellos, se trata de un modelo no conjugado y su análisis será realizado en dos fases: una primera fase en la que se obtendrá la distribución predictiva del vector poblacional no observado, la media no observada de cada unidad y la media poblacional no observada como resultado de un modelo en una jerarquía constituido por las distribuciones predictivas anteriores condicionadas al vector  $(\lambda_1, \lambda_2, \dots, \lambda_K, \lambda_\beta)$ , y la distribución final dada la muestra de este último y una segunda en la que se obtendrán las distribuciones condicionales necesarias para la aplicación de los métodos de Gibbs Sampler (ver sección C.1 en el apéndice C).

El siguiente teorema tiene por objetivo, esta primera fase del análisis del modelo.

**Teorema 3.3** *Consideremos el modelo Bayes de Regresión en dos etapas con dos jerarquías con cociente entre varianzas desconocido dado en (3.30); si definimos por  $\lambda_i = \frac{\sigma_i^2}{\sigma_{\beta_0}^2}$ , para  $i = 1, 2, \dots, K$  y por  $\lambda_\beta = \frac{\sigma_\beta^2}{\sigma_{\beta_0}^2}$ , entonces se tiene que:*

(a)  $\mathbf{Y}_{u_i} | (s, \mathbf{y}_s), \{\lambda_i\}_{i=1}^K, \lambda_\beta$  se distribuye  $t(N - n)$ -variante con  $2a_1$  grados de libertad con, para  $a_1 > 1$ , media, para  $i = 1, 2, \dots, K$ :

$$E(\mathbf{Y}_{u_i} | (s, \mathbf{y}_s), \{\lambda_i\}_{i=1}^K, \lambda_\beta) = \mathbf{X}_{u_i} \hat{\beta}_i^B \quad (3.31)$$

y matriz de varianzas y covarianzas,

\* para  $i \in \{i_1, i_2, \dots, i_k\}$

$$\begin{aligned} Var(\mathbf{Y}_{u_i} | (s, \mathbf{y}_s), \{\lambda_i\}_{i=1}^K, \lambda_\beta) = & \frac{b_1}{a_1 - 1} \{ \mathbf{I}_{N_i - n_i}, \lambda_i + \\ & + \mathbf{X}_{u_i}^t \alpha_i \mathbf{S}_i^{-1} \mathbf{C}_\beta^{-1} \mathbf{S}_w^{-1} \mathbf{S}_i^{-1} \alpha_i^t \mathbf{X}_{u_i}^t \lambda_i^2 \lambda_\beta^{-1} + \\ & + \mathbf{X}_{u_i} (\mathbf{I}_p - \alpha_i) \mathbf{C}_\beta \mathbf{X}_{u_i}^t \lambda_\beta \} , \end{aligned} \quad (3.32)$$

\* para  $i \notin \{i_1, i_2, \dots, i_k\}$ ,

$$Var(\mathbf{Y}_{u_i} | (s, \mathbf{y}_s), \{\lambda_i\}_{i=1}^K, \lambda_\beta) = \frac{b_1}{a_1 - 1} \{ \mathbf{I}_{N_i}, \lambda_i + \mathbf{X}_{u_i} \mathbf{C}_\beta \mathbf{X}_{u_i}^t \lambda_\beta + \mathbf{X}_{u_i} \mathbf{S}_w^{-1} \mathbf{C}_\beta \mathbf{X}_{u_i}^t \lambda_\beta \} , \quad (3.33)$$

\* para  $i \in \{i_1, i_2, \dots, i_k\}, j \notin \{i_1, i_2, \dots, i_k\}$ ,

$$Cov(\mathbf{Y}_{u_i}, \mathbf{Y}_{u_j} | (s, \mathbf{y}_s), \{\lambda_i\}_{i=1}^K, \lambda_\beta) = \frac{b_1}{a_1 - 1} \mathbf{X}_{u_i} \alpha_i \mathbf{S}_i^{-1} \mathbf{C}_\beta^{-1} \mathbf{S}_w^{-1} \mathbf{C}_\beta \mathbf{X}_{u_j}^t \lambda_i, \quad (3.34)$$

\* para  $i \neq j, i, j \in \{i_1, i_2, \dots, i_k\}$ ,

$$\text{Cov}(\mathbf{Y}_{u_i}, \mathbf{Y}_{u_j} | (s, \mathbf{y}_s), \{\lambda_i\}_{i=1}^K, \lambda_\beta) = \frac{b_1}{a_1 - 1} \mathbf{X}_{u_i} \boldsymbol{\alpha}_i \mathbf{S}_i^{-1} \mathbf{C}_\beta^{-1} \mathbf{S}_w^{-1} \mathbf{S}_j^{-1} \boldsymbol{\alpha}_j^t \mathbf{X}_{u_j}^t \lambda_i \lambda_j, \quad (3.35)$$

\* para  $i \neq j$ ,  $i, j \notin \{i_1, i_2, \dots, i_k\}$ ,

$$\text{Cov}(\mathbf{Y}_{u_i}, \mathbf{Y}_{u_j} | (s, \mathbf{y}_s), \{\lambda_i\}_{i=1}^K, \lambda_\beta) = \frac{b_1}{a_1 - 1} \mathbf{X}_{u_i} \mathbf{S}_w^{-1} \mathbf{C}_\beta \mathbf{X}_{u_j}^t \lambda_\beta, \quad (3.36)$$

(b)  $\bar{Y}_{u_i} | (s, \mathbf{y}_s), \{\lambda_i\}_{i=1}^K, \lambda_\beta$  se distribuye  $t$  con  $2a_1$  grados de libertad y media y varianza, para  $a_1 > 1$ , dadas por:

\* para  $i \in \{i_1, i_2, \dots, i_k\}$

$$\begin{aligned} E(\bar{Y}_{u_i} | (s, \mathbf{y}_s), \{\lambda_i\}_{i=1}^K, \lambda_\beta) &= \frac{1}{N_i - n_i} \mathbf{1}_{N_i - n_i}^t \mathbf{X}_{u_i} \hat{\boldsymbol{\beta}}_i^B, \quad (3.37) \\ \text{Var}(\bar{Y}_{u_i} | (s, \mathbf{y}_s), \{\lambda_i\}_{i=1}^K, \lambda_\beta) &= \frac{1}{(N_i - n_i)^2} \frac{b_1}{a_1 - 1} \{ (N_i - n_i) \lambda_i + \\ &+ \mathbf{1}_{N_i - n_i}^t (\mathbf{X}_{u_i} \boldsymbol{\alpha}_i \mathbf{S}_i^{-1} \mathbf{C}_\beta^{-1} \mathbf{S}_w^{-1} \mathbf{S}_i^{-1} \boldsymbol{\alpha}_i^t \mathbf{X}_{u_i}^t \mathbf{1}_{N_i - n_i} \lambda_i^2 \lambda_\beta^{-1} + \\ &+ \mathbf{1}_{N_i - n_i}^t \mathbf{X}_{u_i} (\mathbf{I}_p - \boldsymbol{\alpha}_i) \mathbf{C}_\beta \mathbf{X}_{u_i}^t \mathbf{1}_{N_i - n_i} \lambda_\beta \}, \end{aligned}$$

\* para  $i \notin \{i_1, i_2, \dots, i_k\}$

$$\begin{aligned} E(\bar{Y}_{u_i} | (s, \mathbf{y}_s), \{\lambda_i\}_{i=1}^K, \lambda_\beta) &= \frac{1}{N_i} \mathbf{1}_{N_i}^t \mathbf{X}_{u_i} \hat{\boldsymbol{\beta}}_i^B, \quad (3.38) \\ \text{Var}(\bar{Y}_{u_i} | (s, \mathbf{y}_s), \{\lambda_i\}_{i=1}^K, \lambda_\beta) &= \frac{1}{N_i^2} \frac{b_1}{a_1 - 1} \{ N_i \lambda_i + \mathbf{1}_{N_i}^t \mathbf{X}_{u_i} \mathbf{C}_\beta \mathbf{X}_{u_i}^t \mathbf{1}_{N_i} \lambda_\beta + \\ &+ \mathbf{1}_{N_i}^t \mathbf{X}_{u_i} \mathbf{S}_w^{-1} \mathbf{C}_\beta \mathbf{X}_{u_i}^t \mathbf{1}_{N_i} \lambda_\beta \}, \end{aligned}$$

(c)  $\bar{Y}_u | (s, \mathbf{y}_s), \{\lambda_i\}_{i=1}^K, \lambda_\beta$  se distribuye  $t$  con  $2a_1$  grados de libertad y media y varianza, para  $a_1 > 1$ , dadas por:

$$\begin{aligned} E(\bar{Y}_u | (s, \mathbf{y}_s)) &= \frac{1}{N - n} \sum_{i \in \{i_1, \dots, i_k\}} \mathbf{1}_{N_i - n_i}^t \mathbf{X}_{u_i} \hat{\boldsymbol{\beta}}_i^B + \frac{1}{N - n} \sum_{i \notin \{i_1, \dots, i_k\}} \mathbf{1}_{N_i}^t \mathbf{X}_{u_i} \hat{\boldsymbol{\beta}}_i^B, \\ \text{Var}(\bar{Y}_u | (s, \mathbf{y}_s)) &= \frac{1}{(N - n)^2} \frac{b_1}{a_1 - 1} \{w_1 + w_2\}, \quad (3.39) \end{aligned}$$

(d)  $\lambda_1, \dots, \lambda_K, \lambda_\beta | (s, \mathbf{y}_s)$  se distribuye con función de densidad propia dada por:

$$\begin{aligned} \pi(\lambda_1, \dots, \lambda_K, \lambda_\beta | (s, \mathbf{y}_s)) &\propto \prod_{i \in \{i_1, i_2, \dots, i_k\}} |\mathbf{I}_{n_i} \lambda_i + \mathbf{X}_{s_i} \mathbf{C}_\beta \mathbf{X}_{s_i}^t \lambda_\beta + \mathbf{X}_{s_i} \mathbf{C}_{\beta_0} \mathbf{X}_{s_i}^t|^{-\frac{1}{2}} \times \\ &\times \prod_{i=1}^K \lambda_i^{-(a_0+1)} \lambda_\beta^{-(a_\beta+1)} b_1^{-a_1}. \end{aligned} \quad (3.40)$$

donde  $\bar{y}_s = \frac{\sum_{(i,j) \in s} y_{ij}}{n}$  es la media muestral de la población,  $\bar{y}_{s_i}$  es la media muestral de la unidad  $i$ ,  $\mathbf{S}_i = \mathbf{X}_{s_i}^t \mathbf{X}_{s_i}$ ,  $i = 1, 2, \dots, K$ , y

$$\begin{aligned} \alpha_i &= (\mathbf{X}_{s_i}^t \mathbf{X}_{s_i} \lambda_i^{-1} + \mathbf{C}_\beta^{-1} \lambda_\beta^{-1})^{-1} \mathbf{X}_{s_i}^t \mathbf{X}_{s_i} \lambda_i^{-1}, \quad i \in \{i_1, i_2, \dots, i_k\}, \\ \hat{\beta}_i &= \mathbf{S}_i^{-1} \mathbf{X}_{s_i}^t \mathbf{y}_{s_i}, \quad i \in \{i_1, i_2, \dots, i_k\}, \\ \mathbf{S}_w &= \sum_{i \in \{i_1, i_2, \dots, i_k\}} \alpha_i + \mathbf{C}_\beta \mathbf{C}_{\beta_0}^{-1} \lambda_\beta, \\ \hat{\beta}_w &= \mathbf{S}_w^{-1} \left( \sum_{i \in \{i_1, i_2, \dots, i_k\}} \alpha_i \hat{\beta}_i + \mathbf{C}_\beta \mathbf{C}_{\beta_0}^{-1} \beta_0 \lambda_\beta \right), \\ \hat{\beta}_i^B &= E(\beta_i | (s, \mathbf{y}_s)) = \begin{cases} \alpha_i \hat{\beta}_i + (\mathbf{I}_p - \alpha_i) \hat{\beta}_w, & i \in \{i_1, i_2, \dots, i_k\}, \\ \hat{\beta}_w, & i \notin \{i_1, i_2, \dots, i_k\}, \end{cases} \\ \mathbf{Z}_w &= \sum_{i \in \{i_1, i_2, \dots, i_k\}} \mathbf{1}_{N_i - n_i}^t \mathbf{X}_{u_i} (\mathbf{I}_p - \alpha_i) + \sum_{i \notin \{i_1, i_2, \dots, i_k\}} \mathbf{1}_{N_i}^t \mathbf{X}_{u_i}, \\ a_1 &= K a_0 + a_\beta + a_{\beta_0} + \frac{n}{2}, \\ b_1 &= b_0 \sum_{i=1}^K \frac{1}{\lambda_i} + \frac{b_\beta}{\lambda_\beta} + b_{\beta_0} + \\ &+ \frac{1}{2} \sum_{i \in \{i_1, i_2, \dots, i_k\}} (\mathbf{y}_{s_i} - \mathbf{X}_{s_i} \beta_0)^t [\mathbf{I}_{n_i} \lambda_i + \mathbf{X}_{s_i} \mathbf{C}_\beta \mathbf{X}_{s_i}^t \lambda_\beta + \mathbf{X}_{s_i} \mathbf{C}_{\beta_0} \mathbf{X}_{s_i}^t]^{-1} \times \\ &\times (\mathbf{y}_{s_i} - \mathbf{X}_{s_i} \beta_0), \\ w_1 &= \sum_{i \in \{i_1, i_2, \dots, i_k\}} (N_i - n_i) \lambda_i + \sum_{i \notin \{i_1, i_2, \dots, i_k\}} N_i \lambda_i + \\ &+ \sum_{i \in \{i_1, i_2, \dots, i_k\}} \mathbf{1}_{N_i - n_i}^t \mathbf{X}_{u_i} \alpha_i \mathbf{S}_i^{-1} \mathbf{S}_w^{-1} \mathbf{Z}_w^t \lambda_i + \\ &+ \sum_{i \in \{i_1, i_2, \dots, i_k\}} \mathbf{1}_{N_i - n_i}^t \mathbf{X}_{u_i} \alpha_i \mathbf{S}_i^{-1} \mathbf{X}_{u_i}^t \mathbf{1}_{N_i - n_i} \lambda_i, \\ w_2 &= \sum_{i \notin \{i_1, i_2, \dots, i_k\}} \mathbf{1}_{N_i}^t \mathbf{X}_{u_i} \mathbf{C}_\beta \mathbf{X}_{u_i}^t \mathbf{1}_{N_i} \lambda_\beta + \sum_{i \notin \{i_1, i_2, \dots, i_k\}} \mathbf{1}_{N_i}^t \mathbf{X}_{u_i} \mathbf{C}_\beta \mathbf{S}_w^{-1} \mathbf{Z}_w^t \lambda_\beta. \end{aligned} \quad (3.41)$$

**Demostración.** La demostración de este teorema se apoya en el teorema 3.1 y los lemas B.8 y B.9; para su mejor desarrollo la dividiremos en los siguientes pasos:

1.- A partir de los resultados (a) y (b) del teorema 3.1, tomando  $\mathbf{V} = \text{diag}_{i=1}^K \{\lambda_i \sigma_{\beta_0}^2 \mathbf{I}_{N_i}\}$ ,  $\mathbf{B}_\beta = \lambda_\beta \sigma_{\beta_0}^2 \mathbf{C}_\beta$  y  $\mathbf{B}_{\beta_0} = \sigma_{\beta_0}^2 \mathbf{C}_{\beta_0}$ , se obtienen las distribuciones de  $\mathbf{Y}_u|(s, \mathbf{y}_s), \sigma_{\beta_0}^2, \{\lambda_i\}_{i=1}^K, \lambda_\beta$ , y  $\bar{Y}_{u_i}|(s, \mathbf{y}_s), \sigma_{\beta_0}^2, \{\lambda_i\}_{i=1}^K, \lambda_\beta, i = 1, 2, \dots, K$ , y  $\bar{Y}_u|(s, \mathbf{y}_s), \sigma_{\beta_0}^2, \{\lambda_i\}_{i=1}^K, \lambda_\beta$ , que son normales con media las señaladas en los apartados (a), (b) y (c) de este teorema y varianzas las resultantes de multiplicar las señaladas en estos mismos apartados de este teorema por  $\sigma_{\beta_0}^2$ .

2.- A partir de las distribuciones iniciales de  $\mathbf{Y}|\{\beta_i\}_{i=1}^K, \{\sigma_i^2\}_{i=1}^K, \beta_i|\beta, \sigma_\beta^2, i = 1, 2, \dots, K$  y  $\beta|\sigma_{\beta_0}^2$ , se obtiene aplicando, sobre la primera las propiedades de las distribuciones marginales de una normal y después, sobre el resultado y las segundas dos veces, el apartado (a) del lema B.8, la distribución de  $\mathbf{Y}_s|\sigma_{\beta_0}^2, \{\lambda_i\}_{i=1}^K, \lambda_\beta$ , que es normal con  $\mathbf{Y}_{s_i}|\sigma_{\beta_0}^2, \{\lambda_i\}_{i=1}^K, \lambda_\beta$  independientes cuyas distribuciones vienen dadas para  $i \in \{i_1, \dots, i_k\}$  por:

$$\mathbf{Y}_{s_i}|\sigma_{\beta_0}^2, \{\lambda_i\}_{i=1}^K, \lambda_\beta \sim N_{n_i}(\mathbf{X}_{s_i}\beta_0, \sigma_{\beta_0}^2[\mathbf{I}_{n_i}\lambda_i + \mathbf{X}_{s_i}\mathbf{C}_\beta\mathbf{X}_{s_i}^t\lambda_\beta + \mathbf{X}_{s_i}\mathbf{C}_{\beta_0}\mathbf{X}_{s_i}^t]). \quad (3.42)$$

3.- A partir de las distribuciones iniciales de  $\{\sigma_i^2\}_{i=1}^K, \sigma_\beta^2$  y  $\sigma_{\beta_0}^2$  se obtiene, aplicando el método de obtención de una f.d.p por el método directo (para una transformación 1 a 1) la distribución inicial conjunta de  $\sigma_{\beta_0}, \lambda_1, \lambda_2, \dots, \lambda_K, \lambda_\beta$ , que es:

$$\begin{aligned} \pi(\sigma_{\beta_0}^2, \lambda_1, \dots, \lambda_K, \lambda_\beta) &\propto [\sigma_{\beta_0}^2]^{-(K a_0 + a_\beta + a_{\beta_0} + 1)} \exp\left\{-\frac{1}{\sigma_{\beta_0}^2}\left(b_{\beta_0} + \frac{b_\beta}{\lambda_\beta} + b_0 \sum_{i=1}^K \frac{1}{\lambda_i}\right)\right\} \times \\ &\times \lambda_\beta^{-(a_\beta + 1)} \prod_{i=1}^K \lambda_i^{-(a_0 + 1)}. \end{aligned} \quad (3.43)$$

4.- A partir de (3.42) y (3.43), aplicando el teorema de Bayes e identificando los términos se obtienen las distribuciones de  $\sigma_\beta^2|\{\lambda_i\}_{i=1}^K, \lambda_\beta, (s, \mathbf{y}_s)$ , que viene dada por:

$$\sigma_\beta^2|\{\lambda_i\}_{i=1}^K, \lambda_\beta, (s, \mathbf{y}_s) \sim Ga^{-1}(a_1, b_1), \quad (3.44)$$

y de  $\lambda_1, \lambda_2, \dots, \lambda_K, \lambda_\beta|(s, \mathbf{y}_s)$  dada en (3.40), esta última distribución es propia por ser todas las distribuciones del modelo definido en (3.30) propias.

5.- A partir de la distribución de  $\mathbf{Y}_u|(s, \mathbf{y}_s), \sigma_{\beta_0}^2, \{\lambda_i\}_{i=1}^K, \lambda_\beta$  y de (3.44) aplicando el apartado (a) del lema B.9 se obtiene la distribución de  $\mathbf{Y}_u|(s, \mathbf{y}_s), \{\lambda_i\}_{i=1}^K, \lambda_\beta$  dada de (3.31) a (3.36).

6.- A partir de  $\bar{Y}_{u_i}|(s, \mathbf{y}_s), \sigma_{\beta_0}^2, \{\lambda_i\}_{i=1}^K, \lambda_\beta, i = 1, 2, \dots, K$ , con (3.44), y de  $\bar{Y}_u|(s, \mathbf{y}_s), \sigma_{\beta_0}^2, \{\lambda_i\}_{i=1}^K, \lambda_\beta$  con (3.44), aplicando el apartado (a) del lema B.9 se obtienen respectivamente, la distribución de  $\bar{Y}_{u_i}|(s, \mathbf{y}_s), \{\lambda_i\}_{i=1}^K, \lambda_\beta, i = 1, 2, \dots, K$ , dada en (3.37) y en (3.38) y de  $\bar{Y}_u|(s, \mathbf{y}_s), \{\lambda_i\}_{i=1}^K, \lambda_\beta$ , dada en (3.39).

C.S.Q.D.

Siguiendo el teorema 3.3, la obtención de las distribuciones predictivas del vector poblacional no observado, la media no observada de cada unidad y la media poblacional no observada, así como la obtención de los estimadores Bayes y las varianzas condicionadas a la muestra de la media de cada unidad y de la población, dependen de las propiedades de la distribución de  $(\lambda_1, \lambda_2, \dots, \lambda_K, \lambda_\beta)|(s, \mathbf{y}_s)$ , definida en (3.40). Esta distribución, apoyándonos (tal y como lo hacíamos en la demostración de este teorema) de que se obtiene a partir de un modelos en el que todas las distribuciones son propias y totalmente determinadas, es propia, si  $a_{\beta_0} > 1$  tiene media y si  $a_{\beta_0} > 2$  tiene varianza<sup>2</sup>; no obstante, la existencia de estos momentos no es suficiente para garantizar la existencia de los estimadores Bayes de la media de cada unidad y poblacional y de la varianza condicionada a la muestra de estas medias. Efectivamente, la forma de las distribuciones predictivas condicionadas a  $(\lambda_1, \lambda_2, \dots, \lambda_K, \lambda_\beta)$ , para la media poblacional no observada de cada unidad y de la población dadas en (3.37), (3.38) y (3.39), conducen a tener que garantizar para la existencia de los estimadores y varianzas anteriores la existencia de los siguientes momentos de la distribución de  $(\lambda_1, \lambda_2, \dots, \lambda_K, \lambda_\beta)|(s, \mathbf{y}_s)$ :  $E^{(\lambda_1, \dots, \lambda_K, \lambda_\beta)|(s, \mathbf{y}_s)}[\hat{\beta}^B]$ ,  $Var^{(\lambda_1, \dots, \lambda_K, \lambda_\beta)|(s, \mathbf{y}_s)}[\hat{\beta}^B]$ ,  $E^{(\lambda_1, \dots, \lambda_K, \lambda_\beta)|(s, \mathbf{y}_s)}[Var(\bar{Y}_{u_i}|(s, \mathbf{y}_s), \{\lambda_i\}_{i=1}^K, \lambda_\beta)]$  y  $E^{(\lambda_1, \dots, \lambda_K, \lambda_\beta)|(s, \mathbf{y}_s)}[b_\beta(w_1 + w_2)]$  con  $\hat{\beta}^B = (\hat{\beta}_1^{Bt}, \dots, \hat{\beta}_K^{Bt})^t$ , de estudio bastante complejo y que se deja para posteriores trabajos. Como alternativa a los estimadores Bayes de la media de cada unidad y de la media poblacional y sus varianzas condicionadas a la muestra, proponemos para este modelo como estimador, la mediana de las distribuciones predictivas de la media de cada unidad y de la media poblacional y como medidor de la varianza de la media de cada unidad y de la población dada la muestra, el cuadrado de la cuarta parte de la distancia entre los cuantiles 2.5 % y 97.5 % de las distribuciones predictivas anteriores. Aplicando el algoritmo Gibbs Sampling pueden obtenerse de manera sencilla muestras de las distribuciones predictivas anteriores (y con ellas estimaciones de las anteriores cantidades), ya que todas las distribuciones necesarias para esta técnica son estándares y por lo tanto “fáciles” de simular, tal y como demuestra el siguiente teorema.

**Teorema 3.4** *El modelo Bayes de Regresión en dos etapas con dos jerarquía y con cociente entre varianzas desconocido dado en (3.30), es un modelo DAG (figura 3.1) con distribuciones condicionales dadas por:*

$$\begin{aligned}
 & f((\mathbf{y}_{u_1}^t, \dots, \mathbf{y}_{u_K}^t)^t | \mathbf{y}_s, \{\beta_i\}_{i=1}^K, \{\sigma_i^2\}_{i=1}^K, \sigma_\beta^2, \sigma_{\beta_0}^2) = \\
 & f((\mathbf{y}_{u_1}^t, \dots, \mathbf{y}_{u_K}^t)^t | \{\beta_i\}_{i=1}^K, \{\sigma_i^2\}_{i=1}^K) = \\
 & \prod_{i=1}^K N_{N_i - n_i}(\mathbf{y}_{u_i} | \mathbf{X}_{u_i}, \beta_i, \sigma_i^2 \mathbf{I}_{N_i - n_i}), \quad (3.45) \\
 & \pi((\beta_1^t, \dots, \beta_K^t)^t | \mathbf{y}, \beta, \{\sigma_i^2\}_{i=1}^K, \sigma_\beta^2, \sigma_{\beta_0}^2) =
 \end{aligned}$$

<sup>2</sup>para demostrar estas dos últimas afirmaciones mediante el argumento anterior hay que considerar la generalización del modelo (3.30), consistente en tomar  $\sigma_i^2 \sim Ga^{-1}(a_{0i}, b_{0i})$ , con  $\{a_{0i}\}_{i=1}^K$  y  $\{b_{0i}\}_{i=1}^K$  conocidos y mayores que 0.

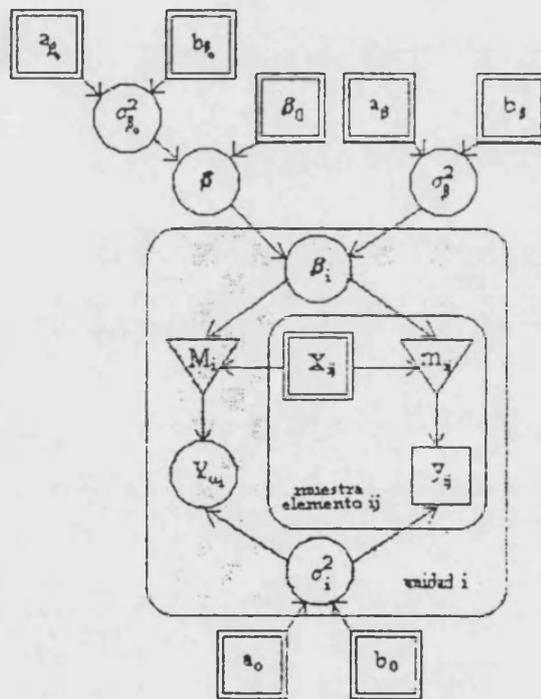


Figura 3.1: Representación gráfica del modelo Bayes de Regresión en dos etapas con dos jerarquías con cociente entre varianzas desconocido

$$\pi((\beta_1^t, \dots, \beta_K^t)^t | \mathbf{y}, \beta, \{\sigma_i^2\}_{i=1}^K, \sigma_\beta^2) = \prod_{i=1}^K N_p(\beta_i | \alpha_i^* \hat{\beta}_i^* + (\mathbf{I}_p - \alpha_i^*) \beta, (\mathbf{I}_p - \alpha_i^*) \mathbf{C}_\beta \sigma_\beta^2), \quad (3.46)$$

$$\pi(\beta | \mathbf{y}, \{\beta_i\}_{i=1}^K, \{\sigma_i^2\}_{i=1}^K, \sigma_\beta^2, \sigma_{\beta_0}^2) = \pi(\beta | \{\beta_i\}_{i=1}^K, \sigma_\beta^2, \sigma_{\beta_0}^2) = N_p(\beta | \delta^* \hat{\beta}^* + (\mathbf{I}_p - \delta^*) \beta_0, (\mathbf{I}_p - \delta^*) \mathbf{C}_{\beta_0} \sigma_{\beta_0}^2), \quad (3.47)$$

$$\pi(\sigma_{\beta_0}^2 | \mathbf{y}, \{\beta_i\}_{i=1}^K, \beta, \{\sigma_i^2\}_{i=1}^K, \sigma_\beta^2) = \pi(\sigma_{\beta_0}^2 | \beta) = Ga^{-1}(\sigma_{\beta_0}^2 | a_{\beta_0}^*, b_{\beta_0}^*), \quad (3.48)$$

$$\pi(\sigma_\beta^2 | \mathbf{y}, \{\beta_i\}_{i=1}^K, \beta, \{\sigma_i^2\}_{i=1}^K, \sigma_{\beta_0}^2) = \pi(\sigma_\beta^2 | \{\beta_i\}_{i=1}^K, \beta) = Ga^{-1}(\sigma_\beta^2 | a_\beta^*, b_\beta^*), \quad (3.49)$$

$$\begin{aligned} \pi((\sigma_1^2, \dots, \sigma_K^2)^t | \mathbf{y}, \{\beta_i\}_{i=1}^K, \beta, \sigma_\beta^2, \sigma_{\beta_0}^2) &= \\ \pi((\sigma_1^2, \dots, \sigma_K^2)^t | \mathbf{y}, \{\beta_i\}_{i=1}^K) &= \prod_{i=1}^K Ga^{-1}(\sigma_i^2 | a_i^*, b_i^*), \end{aligned} \quad (3.50)$$

donde:

$$\begin{aligned} \hat{\beta}_i^* &= (\mathbf{X}_i^t \mathbf{X}_i)^{-1} \mathbf{X}_i^t \mathbf{y}_i, \quad i = 1, 2, \dots, K, \\ \alpha_i^* &= (\mathbf{X}_i^t \mathbf{X}_i + \mathbf{C}_\beta^{-1} \frac{\sigma_i^2}{\sigma_\beta^2})^{-1} \mathbf{X}_i^t \mathbf{X}_i, \quad i = 1, 2, \dots, K, \\ \hat{\beta}^* &= \frac{\sum_{i=1}^K \beta_i}{K}, \\ \delta^* &= (K \mathbf{C}_\beta^{-1} \sigma_\beta^{-2} + \mathbf{C}_{\beta_0}^{-1} \sigma_{\beta_0}^{-2})^{-1} K \mathbf{C}_\beta^{-1} \sigma_\beta^{-2}, \\ a_i^* &= \frac{N_i}{2} + a_0, \quad i = 1, 2, \dots, K, \\ b_i^* &= \frac{1}{2} \sum_{i=1}^K (\mathbf{y}_i - \mathbf{X}_i \beta_i)^t (\mathbf{y}_i - \mathbf{X}_i \beta_i) + b_0, \quad i = 1, 2, \dots, K, \\ a_\beta^* &= \frac{Kp}{2} + a_\beta, \\ b_\beta^* &= \frac{1}{2} \sum_{i=1}^K (\beta_i - \beta)^t \mathbf{C}_\beta^{-1} (\beta_i - \beta) + b_\beta, \\ a_{\beta_0}^* &= \frac{p}{2} + a_{\beta_0}, \\ b_{\beta_0}^* &= \frac{1}{2} (\beta - \beta_0)^t \mathbf{C}_{\beta_0}^{-1} (\beta - \beta_0) + b_{\beta_0}. \end{aligned} \quad (3.51)$$

**Demostración.** De la definición 3.3 del modelo en estudio se deduce directamente la representación gráfica de la figura 3.1, que corresponde a un modelo DAG. Los resultados de las distribuciones condicionales se obtienen a partir de la distribución conjunta de  $(\mathbf{Y}_{u_1}^t, \dots, \mathbf{Y}_{u_1}^t, \mathbf{Y}_i^t, \beta_1^t, \dots, \beta_K^t, \sigma_1^2, \dots, \sigma_K^2, \sigma_\beta^2, \sigma_{\beta_0}^2)^t$ , que viene dada por:

$$\begin{aligned}
f((\mathbf{y}_{u_1}^t, \dots, \mathbf{y}_{u_K}^t, \mathbf{y}_s^t, \beta_1^t, \dots, \beta_K^t, \beta^t, \sigma_1^2, \dots, \sigma_K^2, \sigma_\beta^2, \sigma_{\beta_0}^2)^t) = \\
\prod_{i=1}^K N_{N_i-n_i}(\mathbf{y}_{u_i} | \mathbf{X}_{u_i}, \beta_i, \sigma_i^2 \mathbf{I}_{N_i-n_i}) \times \prod_{i \in \{i_1, \dots, i_k\}} N_{n_i}(\mathbf{y}_s | \mathbf{X}_s, \beta_i, \sigma_i^2 \mathbf{I}_{n_i}) \times \\
\times \prod_{i=1}^K N_p(\beta_i | \beta, \sigma_\beta^2 \mathbf{C}_\beta) \times N_p(\beta | \beta_0, \sigma_{\beta_0}^2 \mathbf{C}_{\beta_0}) \times \\
\times \prod_{i=1}^K Ga^{-1}(\sigma_i^2 | a_0, b_0) \times Ga^{-1}(\sigma_\beta^2 | a_\beta, b_\beta) \times Ga^{-1}(\sigma_{\beta_0}^2 | a_{\beta_0}, b_{\beta_0}).
\end{aligned}$$

separando los términos correspondientes a cada condicional (que contienen la variable cuya condicional estamos hallando), de modo que:

- $f((\mathbf{Y}_{u_1}^t, \mathbf{Y}_{u_2}^t, \dots, \mathbf{Y}_{u_K}^t)^t | \mathbf{y}_s, \{\beta_i\}_{i=1}^K, \{\sigma_i^2\}_{i=1}^K, \sigma_\beta^2, \sigma_{\beta_0}^2) = \prod_{i=1}^K N_{N_i-n_i}(\mathbf{Y}_{u_i} | \mathbf{X}_{u_i}, \beta_i, \sigma_i^2 \mathbf{I}_{N_i-n_i})$ , con lo que se obtiene (3.45).
- $f((\beta_1^t, \dots, \beta_K^t)^t | \mathbf{y}, \beta, \{\sigma_i^2\}_{i=1}^K, \sigma_\beta^2, \sigma_{\beta_0}^2) \propto \prod_{i=1}^K N_{N_i-n_i}(\mathbf{y}_{u_i} | \mathbf{X}_{u_i}, \beta_i, \sigma_i^2 \mathbf{I}_{N_i-n_i}) \times \prod_{i \in \{i_1, \dots, i_k\}} N_{n_i}(\mathbf{y}_s | \mathbf{X}_s, \beta_i, \sigma_i^2 \mathbf{I}_{n_i}) \times \prod_{i=1}^K N_p(\beta_i | \beta, \sigma_\beta^2 \mathbf{C}_\beta)$ , de donde se obtiene, completando cuadrados respecto a  $\beta_i$ ,  $i = 1, 2, \dots, K$ , la distribución dada en (3.46).
- $f(\beta | \mathbf{y}, \{\beta_i\}_{i=1}^K, \{\sigma_i^2\}_{i=1}^K, \sigma_\beta^2, \sigma_{\beta_0}^2) \propto \prod_{i=1}^K N_p(\beta_i | \beta, \sigma_\beta^2 \mathbf{C}_\beta) \times N_p(\beta | \beta_0, \sigma_{\beta_0}^2 \mathbf{C}_{\beta_0})$ , de donde se obtiene, completando cuadrados respecto a  $\beta$ , la distribución dada en (3.47).
- $\pi(\sigma_{\beta_0}^2 | \mathbf{y}, \{\beta_i\}_{i=1}^K, \beta, \{\sigma_i^2\}_{i=1}^K, \sigma_\beta^2) \propto Ga^{-1}(\sigma_{\beta_0}^2 | a_{\beta_0}, b_{\beta_0}) \times N_p(\beta | \beta_0, \sigma_{\beta_0}^2 \mathbf{C}_{\beta_0})$ , de donde se obtiene, juntando los términos correspondientes a potencias y funciones exponenciales de  $\sigma_{\beta_0}^2$ , la distribución dada en (3.48).
- $\pi(\sigma_\beta^2 | \mathbf{y}, \{\beta_i\}_{i=1}^K, \beta, \{\sigma_i^2\}_{i=1}^K, \sigma_{\beta_0}^2) \propto Ga^{-1}(\sigma_\beta^2 | a_\beta, b_\beta) \times \prod_{i=1}^K N_p(\beta_i | \beta, \sigma_\beta^2 \mathbf{C}_\beta)$ , de donde se obtiene, juntando los términos correspondientes a potencias y funciones exponenciales de  $\sigma_\beta^2$ , la distribución dada en (3.49).
- $\pi(\sigma_1^2, \dots, \sigma_K^2 | \mathbf{y}, \{\beta_i\}_{i=1}^K, \beta, \sigma_\beta^2, \sigma_{\beta_0}^2) \propto \prod_{i=1}^K Ga^{-1}(\sigma_i^2 | a_0, b_0) \times \prod_{i=1}^K N_{N_i-n_i}(\mathbf{Y}_{u_i} | \mathbf{X}_{u_i}, \beta_i, \sigma_i^2 \mathbf{I}_{N_i-n_i}) \times \prod_{i \in \{i_1, \dots, i_k\}} N_{n_i}(\mathbf{y}_s | \mathbf{X}_s, \beta_i, \sigma_i^2 \mathbf{I}_{n_i})$ , de donde se obtiene, juntando los términos correspondientes a potencias y funciones exponenciales de  $\sigma_i^2$ ,  $i = 1, 2, \dots, K$ , la distribución dada en (3.50).

C.S.Q.D.

Para finalizar esta subsección, notemos que, por la definición del modelo y por el teorema anterior, se cumplen las hipótesis para poder implementar el algoritmo Gibbs Sampling para este modelo a partir del programa BUGS (ver sección C.2).

### 3.2.4 Modelos Bayes en dos etapas con dos jerarquías mínimo informativos

Para terminar la sección definiremos y estudiaremos, las versiones mínimo informativas de los dos modelos anteriores: el modelo Bayes de Regresión en dos etapas con dos jerarquías Normal mínimo informativo y el modelo Bayes de Regresión en dos etapas con

dos jerarquías Normal - Gamma mínimo informativo, que se aplican para modelizar una característica continua en poblaciones estructuradas en dos etapas, cuando asumimos una relación de intercambiabilidad entre las unidades y no queremos introducir información inicial sobre los parámetros de las distribuciones iniciales en los casos de varianza conocida y desconocida respectivamente.

Las definiciones de los modelos y los resultados inferenciales habituales asociados, aparecen descritos en las siguientes definiciones y los siguientes corolarios.

### Modelo Bayes de Regresión en dos etapas con dos jerarquías Normal mínimo informativo

**Definición 3.4** *Llamaremos modelo Bayes de Regresión en dos etapas con dos jerarquías Normal mínimo informativo a un modelo Bayesiano de Regresión en dos etapas con dos jerarquías Normal en el que la distribución que asumimos sobre  $\beta$  es una mínimo informativa de la forma  $\pi(\beta) \propto 1$ .*

**Corolario 3.1** *Consideremos el modelo Bayes de Regresión en dos etapas con dos jerarquía Normal mínimo informativo dado en la definición 3.4, entonces se tiene que las distribuciones predictivas para el vector poblacional no observado, la media no observada para cada unidad y la media poblacional no observada, y los estimadores de la media poblacional de cada unidad y de la media poblacional son los obtenidos en de (3.2) a (3.13), tomando:*

$$\begin{aligned}
\alpha_i &= (\mathbf{X}_{s_i}^t \mathbf{V}_{s_i}^{-1} \mathbf{X}_{s_i} + \mathbf{B}_\beta^{-1})^{-1} \mathbf{X}_{s_i}^t \mathbf{V}_{s_i}^{-1} \mathbf{X}_{s_i}, \quad i \in \{i_1, i_2, \dots, i_k\}, \\
\hat{\beta}_i &= \mathbf{S}_i^{-1} \mathbf{X}_{s_i}^t \mathbf{V}_{s_i}^{-1} \mathbf{y}_{s_i}, \quad i \in \{i_1, i_2, \dots, i_k\}, \\
\mathbf{S}_w &= \sum_{i \in \{i_1, i_2, \dots, i_k\}} \alpha_i, \\
\hat{\beta}_w &= \mathbf{S}_w^{-1} \sum_{i \in \{i_1, i_2, \dots, i_k\}} \alpha_i \hat{\beta}_i, \\
\hat{\beta}_i^B &= E(\beta_i | (s, \mathbf{y}_s)) = \begin{cases} \alpha_i \hat{\beta}_i + (\mathbf{I}_p - \alpha_i) \hat{\beta}_w, & i \in \{i_1, i_2, \dots, i_k\}, \\ \hat{\beta}_w, & i \notin \{i_1, i_2, \dots, i_k\}, \end{cases} \\
\mathbf{Z}_w &= \sum_{i \in \{i_1, i_2, \dots, i_k\}} \mathbf{1}_{N_i - n_i}^t (\mathbf{X}_{u_i} - \mathbf{V}_{s_i}^t \mathbf{V}_{s_i}^{-1} \mathbf{X}_{s_i}) \alpha_i \mathbf{S}_i^{-1} \mathbf{B}_\beta^{-1} + \sum_{i \notin \{i_1, i_2, \dots, i_k\}} \mathbf{1}_{N_i}^t \mathbf{X}_{u_i},
\end{aligned} \tag{3.52}$$

**Demostración.** Las distribuciones predictivas del vector poblacional no observado y la media poblacional no observada, así como el estimador Bayes para la media poblacional y la varianza condicionada a la muestra de la media poblacional, se obtienen de forma inmediata (salvo operaciones con matrices algo largas) a partir del corolario 3.3 tomando:  $m = 2$ ,  $\mathbf{Y} = (\mathbf{Y}_1^t, \mathbf{Y}_2^t, \dots, \mathbf{Y}_K^t)^t$ ,  $\mathbf{Y}_i = (Y_{i1}, Y_{i2}, \dots, Y_{iN_i})^t$ ,  $i = 1, 2, \dots, K$ ,  $\theta_1 = (\beta_1^t, \beta_2^t, \dots, \beta_K^t)^t$ ,  $\theta_2 = \beta$ ,  $\mathbf{A}_1 = \text{diag}_{i=1}^K \{\mathbf{X}_i\}$ ,  $\mathbf{X}_i = (\mathbf{x}_{i1}^t, \mathbf{x}_{i2}^t, \dots, \mathbf{x}_{iN_i}^t)^t$ ,

$i = 1, 2, \dots, K$ ,  $\mathbf{x}_{ij} = (x_{ij1}, x_{ij2}, \dots, x_{ijp})$ ,  $j = 1, 2, \dots, N_i$ ,  $i = 1, 2, \dots, K$ ,  $\mathbf{A}_2 = (\mathbf{I}_p, \dots, \mathbf{I}_p)^t$ ,  $\mathbf{V}_1 = \text{diag}_{i=1}^K \{\mathbf{V}_i\}$  y  $\mathbf{V}_2 = \text{diag}_{i=1}^K \{\mathbf{B}_\beta\}$ , reordenando las filas de las

matrices de elementos no observados colocando en las  $k$  primeras filas las filas correspondientes a las unidades muestreadas, aplicando el resultado (a) del lema B.1 sobre el vector  $\mathbf{d} = (\mathbf{d}_1^t, \mathbf{d}_2^t, \dots, \mathbf{d}_K^t)^t$  y la matriz  $\mathbf{D}$ , con lo que se obtiene que:

$$\begin{aligned} \mathbf{d}_i &= \begin{cases} \mathbf{X}_{s_{ij}}^t \mathbf{V}_{s_{ij}}^{-1} \mathbf{y}_{s_{ij}}, & j = 1, 2, \dots, k, \\ 0_{N_i}, & j = k+1, \dots, K, \end{cases} \\ \mathbf{D} &= \text{diag}\{\text{diag}_{j=1}^k \{\mathbf{D}_{i_j}\}, \text{diag}_{j=k+1}^K \{\mathbf{B}_\beta\}\} - \\ &\quad - (\mathbf{D}_{i_1}, \mathbf{D}_{i_2}, \dots, \mathbf{D}_{i_k}, \mathbf{B}_\beta, \dots, \mathbf{B}_\beta)^t \mathbf{B}_\beta^{-1} \mathbf{S} \mathbf{B}_\beta^{-1} \times \\ &\quad \times (\mathbf{D}_{i_1}, \mathbf{D}_{i_2}, \dots, \mathbf{D}_{i_k}, \mathbf{B}_\beta, \dots, \mathbf{B}_\beta), \end{aligned}$$

realizando las operaciones de matrices necesarias y denotando por  $\hat{\beta}_i^B$  el  $p$ -vector de  $\hat{\theta}_1^B$  correspondiente a la unidad  $i$ ,  $i = 1, 2, \dots, K$ . Los restantes resultados son inmediatos a partir de la distribución predictiva para el vector poblacional no observado dada la muestra y aplicando que la marginal de normales y la combinación lineal de normales es normal y la definición del estimador Bayes de la media de cada unidad  $i$ .

C.S.Q.D.

### Modelo Bayes de Regresión en dos etapas con dos jerarquías Normal-Gamma mínimo informativo

**Definición 3.5** Llamaremos modelo Bayes de Regresión en dos etapas con dos jerarquías Normal-Gamma mínimo informativo a un modelo Bayes de Regresión en dos etapas con dos jerarquías Normal-Gamma en el que la distribución que asumimos sobre  $(\beta, \sigma^2)$  es una mínimo informativa<sup>3</sup> de la forma  $\pi(\beta, \sigma^2) \propto \frac{1}{\sigma^2}$ .

**Corolario 3.2** Consideremos el modelo Bayes de Regresión en dos etapas con dos jerarquías Normal-Gamma mínimo informativo dado en la definición 3.5, entonces se tiene que las distribuciones predictivas para el vector poblacional no observado, la media no observada para cada unidad y la media poblacional no observada, y los estimadores de la media poblacional de cada unidad y de la media poblacional son los obtenidos en de (3.17) a (3.28), tomando:

$$\begin{aligned} \alpha_i &= (\mathbf{X}_{s_i}^t \mathbf{W}_{s_i}^{-1} \mathbf{X}_{s_i} + \mathbf{C}_\beta^{-1})^{-1} \mathbf{X}_{s_i}^t \mathbf{W}_{s_i}^{-1} \mathbf{X}_{s_i}, \quad i \in \{i_1, i_2, \dots, i_k\}, \\ \hat{\beta}_i &= \mathbf{S}_i^{-1} \mathbf{X}_{s_i}^t \mathbf{W}_{s_i}^{-1} \mathbf{y}_{s_i}, \quad i \in \{i_1, i_2, \dots, i_k\}, \end{aligned}$$

<sup>3</sup>ver nota 7 a pie de página de la subsección 3.4.3

$$\begin{aligned}
\mathbf{S}_w &= \sum_{i \in \{i_1, i_2, \dots, i_k\}} \alpha_i, & (3.53) \\
\widehat{\beta}_w &= \mathbf{S}_w^{-1} \sum_{i \in \{i_1, i_2, \dots, i_k\}} \alpha_i \widehat{\beta}_i, \\
\widehat{\beta}_i^B &= E(\beta_i | (s, \mathbf{y}_s)) = \begin{cases} \alpha_i \widehat{\beta}_i + (\mathbf{I}_p - \alpha_i) \widehat{\beta}_w, & i \in \{i_1, i_2, \dots, i_k\}, \\ \widehat{\beta}_w, & i \notin \{i_1, i_2, \dots, i_k\}, \end{cases} \\
\mathbf{Z}_w &= \sum_{i \in \{i_1, i_2, \dots, i_k\}} \mathbf{1}_{N_i - n_i}^t (\mathbf{X}_{u_i} - \mathbf{W}_{s_{u_i}}^t \mathbf{W}_{s_i}^{-1} \mathbf{X}_{s_i}) \alpha_i \mathbf{S}_i^{-1} \mathbf{C}_\beta^{-1} + \sum_{i \notin \{i_1, i_2, \dots, i_k\}} \mathbf{1}_{N_i}^t \mathbf{X}_{u_i}, \\
a_1 &= \frac{n-p}{2}, \\
b_1 &= \frac{1}{2} \sum_{i \in \{i_1, i_2, \dots, i_k\}} \mathbf{y}_{s_i}^t (\mathbf{W}_{s_i} + \mathbf{X}_{s_i} \mathbf{C}_\beta \mathbf{X}_{s_i}^t)^{-1} \mathbf{y}_{s_i} - b_s, \\
b_s &= \frac{1}{2} \left[ \sum_{i \in \{i_1, i_2, \dots, i_k\}} \mathbf{y}_{s_i}^t (\mathbf{W}_{s_i} + \mathbf{X}_{s_i} \mathbf{C}_\beta \mathbf{X}_{s_i}^t)^{-1} \mathbf{X}_{s_i} \right] \times \\
&\quad \times \left[ \sum_{i \in \{i_1, i_2, \dots, i_k\}} \mathbf{X}_{s_i}^t (\mathbf{W}_{s_i} + \mathbf{X}_{s_i} \mathbf{C}_\beta \mathbf{X}_{s_i}^t)^{-1} \mathbf{X}_{s_i} \right]^{-1} \times \\
&\quad \times \left[ \sum_{i \in \{i_1, i_2, \dots, i_k\}} \mathbf{X}_{s_i}^t (\mathbf{W}_{s_i} + \mathbf{X}_{s_i} \mathbf{C}_\beta \mathbf{X}_{s_i}^t)^{-1} \mathbf{y}_{s_i} \right]
\end{aligned}$$

**Demostración.** Las distribuciones predictivas del vector poblacional no observado y la media poblacional no observada, así como el estimador Bayes para la media poblacional y la varianza condicionada a la muestra de la media poblacional, se obtienen de manera inmediata (salvo operaciones con matrices algo largas) a partir del corolario 3.4, tomando:  $m = 2$ ,  $\mathbf{Y} = (\mathbf{Y}_1^t, \mathbf{Y}_2^t, \dots, \mathbf{Y}_K^t)^t$ ,  $\mathbf{Y}_i = (Y_{i1}, Y_{i2}, \dots, Y_{iN_i})^t$ ,  $i = 1, 2, \dots, K$ ,  $\boldsymbol{\theta}_1 = (\beta_1^t, \beta_2^t, \dots, \beta_K^t)^t$ ,  $\boldsymbol{\theta}_2 = \beta$ ,  $\mathbf{A}_1 = \text{diag}_{i=1}^K \{\mathbf{X}_i\}$ ,  $\mathbf{X}_i = (\mathbf{x}_{i1}^t, \mathbf{x}_{i2}^t, \dots, \mathbf{x}_{iN_i}^t)^t$ ,  $i = 1, 2, \dots, K$ ,  $\mathbf{x}_{ij} = (x_{ij1}, x_{ij2}, \dots, x_{ijp})$ ,  $j = 1, 2, \dots, N_i$ ,  $i = 1, 2, \dots, K$ ,  $\mathbf{A}_2 = \underbrace{(\mathbf{I}_p, \dots, \mathbf{I}_p)}_K^t$ ,  $\mathbf{W}_1 = \text{diag}_{i=1}^K \{\mathbf{W}_i\}$  y  $\mathbf{W}_2 = \text{diag}_{i=1}^K \{\mathbf{C}_\beta\}$ , reordenando las filas

de las matrices de elementos no observados colocando en las  $k$  primeras filas las filas correspondientes a las unidades muestreadas, aplicando de manera totalmente similar al corolario anterior el lema B.1 sobre el vector  $\mathbf{d}$ , la matriz  $\mathbf{D}$  y el escalar  $b_1$ , realizando las operaciones de matrices necesarias y denotando por  $\widehat{\beta}_i^B$  el  $p$ -vector de  $\widehat{\boldsymbol{\theta}}_1^B$  correspondiente a la unidad  $i$ ,  $i = 1, 2, \dots, K$ . Las distribuciones predictivas de la media no observada de cada unidad  $i$ ,  $i = 1, 2, \dots, K$ , se obtienen<sup>4</sup> aplicando el resultado (a) del lema B.9 sobre la distribución de  $\overline{\mathbf{Y}}_{u_i} | (s, \mathbf{y}_s)$ ,  $\sigma^2$  que se obtiene directamente de (3.8) y (3.9) tomando las definiciones de (3.52) y  $\mathbf{V}_i = \sigma^2 \mathbf{W}_i$ ,  $i = 1, 2, \dots, K$  y sobre la distribución (3.84) de la demostración del corolario 3.4 y los estimadores Bayes

<sup>4</sup>Una demostración alternativa consiste en aplicar sobre la distribución predictiva del vector poblacional no observado obtenida en la primera parte de la demostración de la propiedad de que la marginal de una  $t$  y la combinación lineal de  $t$  son  $t$

para la media poblacional de cada unidad y las varianzas de la media de cada unidad condicionada a la muestra se obtienen a partir de las distribuciones predictivas de la media no observada de cada unidad aplicando la definición de estimador Bayes.

C.S.Q.D.

### 3.3 Ejemplos para modelos Bayesianos en dos etapas con dos jerarquías: otra modelización de la población 1B

En esta sección retomaremos el estudio de la población 1B, que ya definíamos y estudiábamos en la sección 2.4 del capítulo anterior.

La población 1B era el resultado de la estratificación de la población 1A (en 3 estratos), a través de un indicador socioeconómico sobre el porcentaje de jóvenes en edad de servicio militar que obtenían la máxima calificación en el examen militar. La característica en estudio es la fertilidad y las variables auxiliares para describirla de que disponemos son: el porcentaje de la provincia de ocupación en el sector agrario, el porcentaje de la provincia de personas con estudios superiores, el porcentaje de la provincia de católicos y el porcentaje de la provincia de mortalidad infantil. Las relaciones entre la característica en estudio y las variables auxiliares anteriores, por estrato se valoraban a partir de los datos muestrales en la figura 2.4.

Realizada la transformación de los datos de la población en porcentajes,  $p$ , de acuerdo con la fórmula:  $\log(\frac{p}{100-p})$ , para adecuarlos a las habituales hipótesis de normalidad, y considerada una muestra de 47 provincias de las 250 existentes (con datos asimismo transformados y clasificados según los estratos), se aplicaron, en la sección 2.4, para su análisis los modelos Bayesianos en dos etapas con una jerarquía. A continuación aplicaremos los correspondientes modelos Bayesianos en dos etapas con dos jerarquías para realizar ese mismo análisis (a posteriori de la transformación logarítmica antes citada), y compararemos los resultados obtenidos por los dos grupos de modelos. En concreto, los modelos aplicados en este segundo análisis de la población 1B, han sido los siguientes:

1. El modelo Bayes de Regresión en dos etapas con dos jerarquías Normal, con parámetros  $\mathbf{V}_i = \sigma_i^2 \mathbf{I}_{N_i}$ ,  $i = 1, 2, 3$ ,  $\sigma_1^2 = 0.36$ ,  $\sigma_2^2 = 0.5$ ,  $\sigma_3^2 = 0.43$ ,  $\mathbf{B}_\beta = 2\mathbf{I}_4$ ,  $\mathbf{B}_{\beta_0} = 4\mathbf{I}_4$  y  $\beta_0 = (-1, 2, -2, 3)^t$ .
2. El modelo Bayes de Regresión en dos etapas con dos jerarquías Normal-Gamma, con parámetros  $\mathbf{W}_i = \mathbf{I}_{N_i}$ ,  $i = 1, 2, 3$ ,  $\mathbf{C}_\beta = \mathbf{I}_4$ ,  $\mathbf{C}_{\beta_0} = \mathbf{I}_4$ ,  $\beta_0 = (-1, 2, -2, 3)^t$ ,  $a_0 = 3$  y  $b_0 = 8$  ( $E(\sigma^2) = 4$ ,  $Var(\sigma^2) = 16$ ).
3. El modelo Bayes de Regresión en dos etapas con dos jerarquías con cociente entre varianzas desconocido, con parámetros  $\mathbf{W}_i = \mathbf{I}_{N_i}$ ,  $i = 1, 2, 3$ ,  $\mathbf{C}_\beta = \mathbf{I}_4$ ,  $\mathbf{C}_{\beta_0} = \mathbf{I}_4$ ,  $\beta_0 = (-1, 2, -2, 3)^t$ ,  $a_0 = 3$ ,  $b_0 = 2$  ( $E(\sigma^2) = 1$ ,  $Var(\sigma^2) = 1$ ),  $a_\beta = 3$ ,  $b_\beta = 4$  ( $E(\sigma_\beta^2) = 2$ ,  $Var(\sigma_\beta^2) = 4$ ),  $a_{\beta_0} = 3$  y  $b_{\beta_0} = 8$  ( $E(\sigma_{\beta_0}^2) = 4$ ,  $Var(\sigma_{\beta_0}^2) = 16$ ).

Para estudiar este modelo se obtuvo una muestra de la distribución predictiva de la media no observada poblacional y de cada estrato mediante la aplicación del algoritmo Gibbs Sampling. Este método de análisis se implementó en el ordenador mediante el programa BUGS y constó de tres fases: programación del modelo según BUGS (figura 3.2), fase de simulación de todas las variables del modelo y análisis de convergencia del algoritmo y una vez asumida la estabilidad de la cadena de Markov, simulación de 1000 observaciones de las distribuciones predictivas y finales a partir de las cuales se obtienen los estimadores y una aproximación gráfica a las densidades de dichas distribuciones predictivas y finales. En la fase de simulación y análisis de convergencia, se simuló inicialmente 1000 observaciones que se emplearon únicamente para la obtención del punto inicial del muestreo Gibbs (y no en el análisis de convergencia del método), y después 5000 simulaciones más a partir del nuevo punto de arranque con las que se analizó la estabilidad mediante el estadístico aproximado de Geweke(92)[39] (que se obtiene con el comando `diag` del BUGS) de las variables  $\bar{Y}_{u_1}$ ,  $\beta_{23}$ ,  $\sigma_3^2$ ,  $\sigma_\beta^2$  y  $\sigma_{\beta_0}^2$ , obteniendo para todas ellas valores inferiores en módulo a 1.96, por lo que pareció razonable suponer alcanzada la estabilidad de la cadena. Asumida la estabilidad se simuló 1000 observaciones más (simulaciones, dada la supuesta estabilidad de la cadena, de las distribuciones predictivas y finales), a partir de las que se obtuvieron, mediante el comando `stats`, los estadísticos mediana y cuantiles 2.5 % y 97.5 %, para las variables sobre las medias de la población y pendientes y media y varianza para las variables  $\lambda$ 's, y mediante el programa de Splus, `drawdat`, incluido en el BUGS, las aproximaciones kernel a las densidades de las distribuciones predictivas de las medias y finales para las  $\lambda$ 's dada la muestra (consultar C.3 para ver más detalles sobre cómo se ha fundamentado el análisis a través del programa BUGS).

4. El modelo Bayes de Regresión en dos etapas con dos jerarquías Normal mínimo informativo, con parámetros  $\mathbf{V}_i = \sigma_i^2 \mathbf{I}_{N_i}$ ,  $i = 1, 2, 3$  con  $\sigma_1^2 = 0.36$ ,  $\sigma_2^2 = 0.5$  y  $\sigma_3^2 = 0.43$  y  $\mathbf{B}_\beta = 2\mathbf{I}_4$ .
5. El modelo Bayes de Regresión en dos etapas con dos jerarquías Normal-Gamma mínimo informativo, con parámetros  $\mathbf{W}_i = \mathbf{I}_{N_i}$ ,  $i = 1, 2, 3$  y  $\mathbf{C}_\beta = \mathbf{I}_4$ .

A partir de los datos de la muestra de la población y los cinco modelos anteriores, se obtuvieron aplicando los teoremas 3.1 y 3.2, en los dos primeros, el programa BUGS, en el modelo con cociente entre varianzas desconocido (apoyándonos en los teoremas 3.3 y 3.4), y los corolarios 3.1 y 3.2, para los modelos mínimo informativos, los estimadores Bayes de la media de cada estrato y la media poblacional así como la varianza condicionada a la muestra de dichas medias para los modelos conjugados y en su sustitución para el modelo con cociente entre varianzas desconocido la mediana y cuadrado de la cuarta parte de la longitud del intervalo de confianza al 95 % de las distribuciones predictivas de la media de cada estrato y media poblacional, estimados a partir de la muestra simulada por Gibbs (el empleo de este estimador y valor de la "varianza" alternativos se justificaba en la subsección 3.2.3), que aparecen reunidos en la tabla 3.1, y las distribuciones predictivas de la media no observada de cada estrato y la media poblacional no observada para los modelos conjugados (tabla 3.2). También se obtuvo,

Modelo Bayes en dos etapas con dos jerarquías con cociente entre varianzas desconocido	
<pre> model M022CD. const NT=250, m=47, p=4, K=3. var y[nt], yinds[K], Y mdu[K], Y md[K], Y med. Y medu, Y totu[K], x[nt,p], X mdu[K,p], beta[K,p], beta b[p], b[p], sigma2[K], sigma2 b. sigma2 0, m[nt], M[K], tau[K], tau b, tau 0. tau m[K], lambda[K], lambda b, n[K], N[K]. data in "file.dat". } for(i in 1:n[1]){ y[i]-dnorm(m[i],tau[1]), m[i]&lt;-x[i,1]*beta[1,1]+x[i,2]*beta[1,2]+ x[i,3]*beta[1,3]+x[i,4]*beta[1,4]. } for(i in n[1]+1:n[2]){ y[i]-dnorm(m[i],tau[2]), m[i]&lt;-x[i,1]*beta[2,1]+x[i,2]*beta[2,2]+ x[i,3]*beta[2,3]+x[i,4]*beta[2,4]. } for(i in n[1]+n[2]+1:n[3]){ y[i]-dnorm(m[i],tau[3]), m[i]&lt;-x[i,1]*beta[3,1]+x[i,2]*beta[3,2]+ x[i,3]*beta[3,3]+x[i,4]*beta[3,4]. } for(i in 1:K){ for(j in 1:p){ beta[i,j]-dnorm(beta b[j],tau b). } } </pre>	<pre> Y mdu[i]-dnorm(M[i],tau m[i]), tau m[i]&lt;-4*N[i]-n[i]*tau[i], M[i]-X mdu[i,1]*beta[i,1]-X mdu[i,2]*beta[i,2]- X mdu[i,3]*beta[i,3]-X mdu[i,4]*beta[i,4], tau[i]-dgamma(3 0,2 0), sigma2[i]&lt;-1+tau[i], lambda[i]&lt;-sigma2[i]/sigma2 0, Y totu[i]&lt;-(N[i]-n[i])*Y mdu[i], Y md[i]&lt;-(n[i]*N[i])*y mds[i]+Y totu[i]*N[i]. } for(i in 1:p){ beta b[i]-dnorm(b[i],tau 0), } tau b-dgamma(3 0,4 0), sigma2 b&lt;-1+tau b, tau 0-dgamma(3 0,5 0), sigma2 0&lt;-1+tau 0, lambda b&lt;-sigma2 b/sigma2 0, Y medu&lt;-sum(Y totu[])/(NT-nt), Y med&lt;-sum(y[,1])/NT-sum(Y totu[])/NT. } </pre>
<pre> "file data" (contenido) list(y=c(-datos de la muestra separados por comas-), x=c(-datos variable explicativa de la muestra separados por comas-), b=c(-1,2,-2,3), X mdu=c(-vector medias no observadas-), y mds=c(-vector de medias muestrales para cada estrato), n=c(33,7,7), N=c(176,37,37)) </pre>	

Figura 3.2: Traducción en BUGS del modelo Bayes de Regresión en dos etapas con dos jerarquías con cociente entre varianzas desconocido

		MODELO NORMAL	MODELO NORM-GAMMA	MODELO COCIENTE VAR. DESCONOCIDO	MODELO NORMAL MIN. INFORM.	MODELO NORM-GAMMA MIN. INFORM.
E 1	$\hat{\bar{Y}}_1$	0.8264 (69.56)	0.8178 (69.38)	0.8324 (69.69)	0.8270 (69.57)	0.8144 (69.30)
	Var( $\hat{\bar{Y}}_1 (s, y_s)$ )	0.0104	0.0471	0.0127	0.0104	0.0177
E 2	$\hat{\bar{Y}}_2$	0.9982 (73.07)	0.9911 (72.93)	0.3497 (58.65)	0.9879 (72.87)	1.0913 (74.86)
	Var( $\hat{\bar{Y}}_2 (s, y_s)$ )	0.0495	0.1479	1.6920	0.0496	0.0567
E 3	$\hat{\bar{Y}}_3$	1.0833 (74.72)	0.9515 (72.14)	0.4698 (61.53)	1.0753 (74.56)	1.0368 (73.82)
	Var( $\hat{\bar{Y}}_3 (s, y_s)$ )	0.0372	0.1185	10.0283	0.0374	0.0455
	$\hat{\bar{Y}}$	0.8899 (70.89)	0.8632 (70.33)	0.6996 (66.81)	0.8875 (70.84)	0.8883 (70.85)
	Var( $\hat{\bar{Y}} (s, y_s)$ )	0.0071	0.0294	0.2075	0.0071	0.0111

Las estimaciones de la media poblacional de cada estrato y de toda la población en porcentajes figuran en la tabla entre parentesis. En el modelo cociente varianzas desconocido se ha empleado como estimador la mediana y la varianza se ha estimado mediante el cuadrado de la cuarta parte de la distancia entre los cuantiles 0.75 y 0.25.

Tabla 3.1: Estimadores Bayes de la media de cada estrato y poblacional y sus varianzas para los modelos Bayesianos en dos etapas con dos jerarquías

la aproximación de kernel a las densidades de las distribuciones predictivas de la media poblacional no observada,  $\bar{Y}_u$  (Y.medu), las medias no observadas de cada estrato,  $\bar{Y}_{u_i}$  (Y.mdu[i]),  $i = 1, 2, 3$  y de las distribuciones finales marginales dada la muestra de  $\lambda_i$  (lambda[i]),  $i = 1, 2, 3$  y  $\lambda_\beta$  (lambda.b) para el modelo con cociente entre varianzas desconocido (figuras 3.3 y 3.4). Y para completar el análisis, los estimadores Bayes para las pendientes de regresión,  $\beta_i$ ,  $i = 1, 2, 3$  y  $\sigma^2$  (en su caso), para los modelos conjugados y los estimadores mediana para las pendientes de regresión,  $\beta_i$ ,  $i = 1, 2, 3$  y  $\sigma_{\beta_0}^2$ , y el estimador Bayes<sup>5</sup> para  $\lambda_i$ ,  $i = 1, 2, 3$  y  $\lambda_\beta$  para el modelo Bayes de Regresión en dos etapas con dos jerarquías con cociente entre varianzas desconocido (tabla 3.3).

A partir de estos resultados podemos destacar lo siguiente:

1.- Respecto a la estimación de la media de cada estrato y de la población, señalemos:

- La semejanza de los estimadores de la media del estrato 1 (alrededor del 69.5 %), para los cinco modelos estudiados.
- Los estimadores de la media de los estratos 2 y 3 y de la población son asimismo bastante semejantes para los modelos conjugados (alrededor del 73 %, para la

<sup>5</sup>la existencia de este estimador queda garantizada por los comentarios realizados sobre la existencia del primer momento de la distribución de  $(\lambda_1, \dots, \lambda_K, \lambda_\beta)$ , realizados en la subsección 3.2.3.

MOD. NORMAL	$\bar{Y}_{U_1}   (s, y_s) \sim N(0.8421, 0.0157)$ $\bar{Y}_{U_2}   (s, y_s) \sim N(0.9116, 0.0753)$ $\bar{Y}_{U_3}   (s, y_s) \sim N(1.0208, 0.0565)$ $\bar{Y}_U   (s, y_s) \sim N(0.8788, 0.0107)$
MOD. NORM-GAMMA	$\bar{Y}_{U_1}   (s, y_s) \sim t(53, 0.8316, 0.0687)$ $\bar{Y}_{U_2}   (s, y_s) \sim t(53, 0.9028, 0.2165)$ $\bar{Y}_{U_3}   (s, y_s) \sim t(53, 0.8580, 0.1735)$ $\bar{Y}_U   (s, y_s) \sim t(53, 0.8460, 0.0445)$
MOD. NORMAL MIN. INF.	$\bar{Y}_{U_1}   (s, y_s) \sim N(0.8429, 0.0157)$ $\bar{Y}_{U_2}   (s, y_s) \sim N(0.8988, 0.0755)$ $\bar{Y}_{U_3}   (s, y_s) \sim N(1.0107, 0.0569)$ $\bar{Y}_U   (s, y_s) \sim N(0.8759, 0.0107)$
MOD. NORM-GAMMA MIN. INF.	$\bar{Y}_{U_1}   (s, y_s) \sim t(43, 0.8274, 0.0256)$ $\bar{Y}_{U_2}   (s, y_s) \sim t(43, 1.0264, 0.0822)$ $\bar{Y}_{U_3}   (s, y_s) \sim t(43, 0.9632, 0.0660)$ $\bar{Y}_U   (s, y_s) \sim t(43, 0.8769, 0.0161)$

Tabla 3.2: Distribuciones predictivas de la media de cada estrato y poblacional no observada para los modelos Bayesianos en dos etapas con dos jerarquías conjugados

### Output for MB22CD analysis

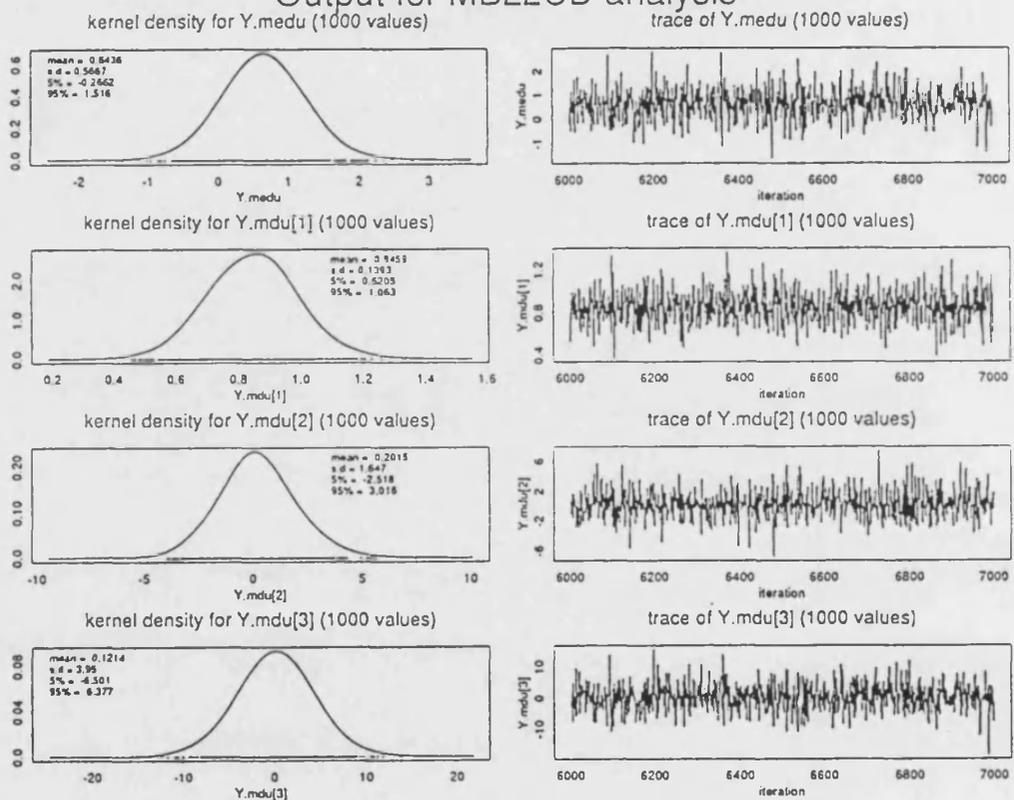


Figura 3.3: Aproximaciones a las densidades predictivas para el modelo Bayes de Regresión en dos etapas con dos jerarquías con cociente entre varianzas desconocido

### Output for MB22CD (2) analysis

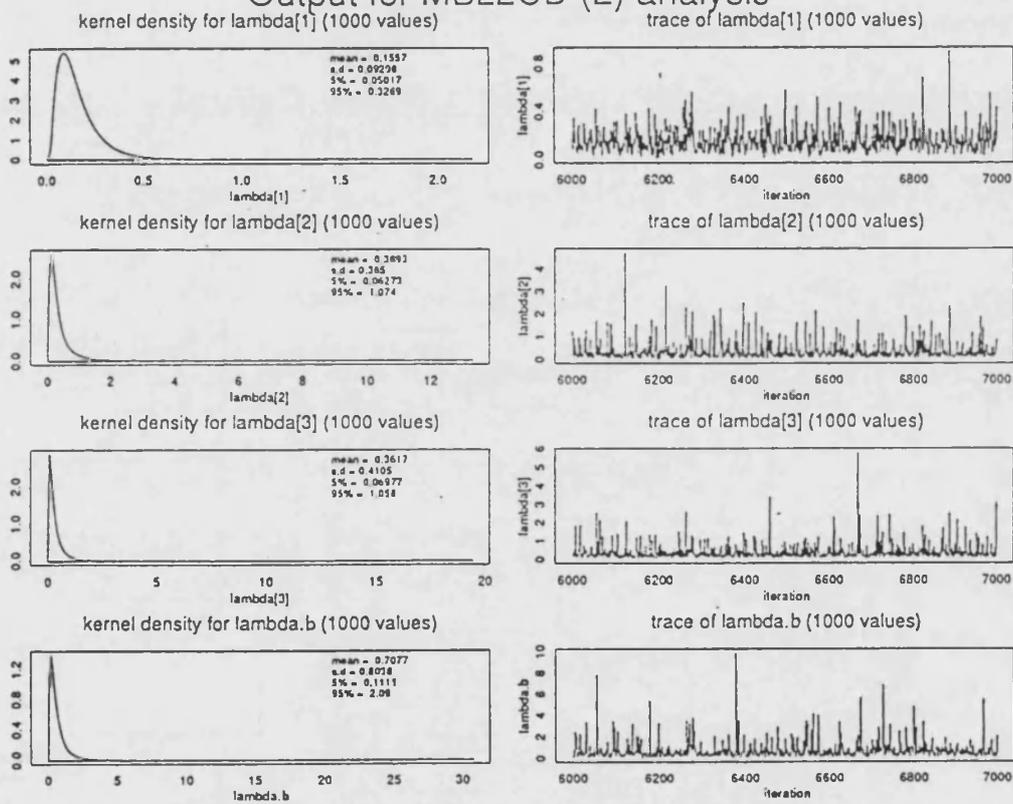


Figura 3.4: Aproximaciones a las densidades finales para el modelo Bayes de Regresión en dos etapas con dos jerarquías con cociente entre varianzas desconocido



		MODELO NORMAL	MODELO NORM-GAMMA	MODELO COCIENTE VAR DESCONOCIDO	MODELO NORMAL MIN. INFOR.	MODELO NORM-GAMMA MIN. INFOR.
E S T R A T O 1	$\beta_{11}$	-0.0019	-0.0635	0.0270	0.0011	-0.0989
	$\beta_{12}$	-0.0255	0.0572	-0.0395	-0.0277	0.1137
	$\beta_{13}$	-0.0620	-0.1364	-0.0583	-0.0595	-0.1837
	$\beta_{14}$	-0.2659	-0.2114	-0.2893	-0.2690	-0.1845
	$\lambda_1$			0.1557		
E S T R A T O 2	$\beta_{21}$	-0.6369	-0.6371	-0.2944	-0.5997	-0.8444
	$\beta_{22}$	0.8023	0.8557	0.6470	0.7824	1.0949
	$\beta_{23}$	-1.5064	-1.5183	-0.5895	-1.4818	-1.7616
	$\beta_{24}$	1.0291	1.0259	0.8335	0.9919	1.2087
	$\lambda_2$			0.3692		
E S T R A T O 3	$\beta_{31}$	-0.7696	-0.5313	-0.2339	-0.7633	-0.6455
	$\beta_{32}$	2.4725	1.9246	0.7233	2.4150	2.4169
	$\beta_{33}$	-2.3974	-1.8340	-0.7342	-2.3306	-2.3839
	$\beta_{34}$	0.1861	0.1728	0.7771	0.1620	0.3105
	$\lambda_3$			0.3617		
	$\lambda_0$			0.7076		
	$\sigma_0^2$			3.1100		
	$\sigma^2$		1.6387			0.6163

En el modelo con error varianzas desconocido se ha empleado como estimador la mediana, salvo en el caso de la estimación de las  $\lambda_i$  que han sido estimadas como es habitual con la media de la distribución a posteriori dada la muestra.

Tabla 3.3: Otras inferencias para los modelos Bayesianos en dos etapas con dos jerarquías

estimación de las medias de los dos estratos y del 71 %, para la media poblacional). Los estimadores para la media de los estratos 2 y 3 y de la población obtenidos con el modelo con cociente entre varianzas desconocido difieren de los anteriores sustancialmente, sobre todo, los referidos a las medias de los estratos.

2.- Respecto a la precisión de las estimaciones, podemos señalar lo siguiente:

- Las varianzas condicionadas a la muestra para la media del estrato 1 son muy similares para los cinco modelos, presentando el siguiente orden de menor a mayor valor: modelo normal, modelo normal mínimo informativo, modelo con cociente entre varianzas desconocido, modelo normal-gamma mínimo informativo y modelo normal-gamma.
- Las varianzas condicionadas a la muestra para las medias de los estratos 2 y 3, y poblacional son bastante similares para los cuatro modelos conjugados, siendo la correspondiente al modelo con cociente entre varianzas desconocido mucho más alta, sobre todo, para las medias de los estratos 2 y 3 (aunque menos que para el modelo Bayes de Regresión en dos etapas con una jerarquías con cociente entre varianzas desconocido). El orden entre los modelos según la varianza es el citado arriba sólo que ahora ocupa la última posición (la de varianza más alta), el modelo c cociente entre varianzas desconocido.

3.- A partir de las dos observaciones anteriores se revela el papel del modelo Bayes de Regresión en dos etapas con dos jerarquías con cociente entre varianzas desconocido, como valorador de las desviaciones de la hipótesis de linealidad, ya que como ya sucedía para los otros dos grupos de modelos con el correspondiente modelo para cociente entre varianzas desconocido, funciona de manera similar a los modelos conjugados (en estimación y varianza), en la estimación de la media del estrato 1 y detecta las desviaciones de esta hipótesis para los estratos 2 y 3, proporcionando estimadores más bajos y estimaciones de las varianzas más altas que las correspondientes a los modelos conjugados.

Si nos fijamos en las estimaciones de las pendientes de regresión, parece detectarse una mayor proximidad entre los modelos conjugados y cociente entre varianzas desconocido en dos etapas con dos jerarquías que la observada entre los modelos Bayesianos en dos etapas con una jerarquía.

4.- Conviene señalar también respecto a los modelos conjugados que:

- Los estimadores de la media de cada estrato y poblacional dejan de coincidir para los modelos normal y normal-gamma mínimo informativos, ya que la distribución mínimo informativa se introduce en la segunda jerarquía del modelo.
- El modelo normal-gamma señala para las cantidades estimadas, varianzas más altas que el modelo normal-gamma mínimo informativo, al contrario de lo que sucedía en los otros dos grupos de modelos. Esta diferencia puede deberse a dos circunstancias, por una parte que la distribución mínimo informativa sobre el

modelo normal-gamma mínimo informativo es sobre la segunda jerarquía (por lo que la primera jerarquía del modelo sigue proporcionando cierta información), y por otra parte que la distribución inicial asignada sobre  $\beta$ , en el modelo normal-gamma no es buena (si nos fijamos en las pendientes estimadas y en la asignada a priori, las diferencias son bastante grandes).

5.- Los resultados obtenidos para esta población con estos modelos y con los modelos en dos etapas con una jerarquía, son muy similares y conducen a las mismas conclusiones. No obstante, se aprecia, en general una varianza menor para las variables estimadas en el caso de los modelos con dos jerarquías y se detecta la mayor relación establecida entre los estratos en pendientes estimadas para cada estrato más similares entre sí en todos los modelos en dos etapas con dos jerarquías.

6.- Por último, realizaremos un breve estudio de las distribuciones predictivas de la media no observada de cada estrato y la media poblacional, y finales marginales de  $\lambda_i$ ,  $i = 1, 2, 3$ , y  $\lambda_\beta$ , para el modelo con cociente entre varianzas desconocido (recordemos que estas distribuciones no tienen forma estándar), a través de las densidades aproximadas de las mismas dadas en las figuras 3.3 y 3.4. Notemos que:

- Las densidades de las distribuciones predictivas para la media poblacional no observada y media no observada para cada estrato, son acampanadas y ligeramente asimétricas y con un apuntamiento superior a la correspondiente normal para las densidades predictivas de las variables  $\bar{Y}_u$ ,  $\bar{Y}_{u_1}$  y  $\bar{Y}_{u_3}$ , y menor en el caso de  $\bar{Y}_{u_2}$ .
- Si comparamos las distribuciones predictivas para los cinco modelos a través de la amplitud del intervalo de confianza del 90 %, podemos establecer el siguiente orden de menor a mayor (mayor a menor apuntamiento): modelos normal, modelo normal mínimo informativo, modelo normal-gamma mínimo informativo, modelo normal-gamma y modelo con cociente entre varianzas desconocido, que sólo es alterado en el caso de la distribución predictiva para la media del estrato 1 por el modelo con cociente entre varianzas desconocido, que en este caso se sitúa a continuación de los modelos normales.
- Respecto a las distribuciones finales marginales de  $\lambda_i$ ,  $i = 1, 2, 3$  y  $\lambda_\beta$ , señalemos que sus densidades presentan una forma menos apuntada que las correspondientes densidades de las distribuciones finales marginales de las  $\lambda$ 's para el modelo Bayes de Regresión en dos etapas con una jerarquía con cociente entre varianzas desconocido.

(Las figuras 3.3 y 3.4 muestran a la derecha un gráfico que representa los valores de las simulaciones de las variables  $\bar{Y}_u$ ,  $\bar{Y}_{u_i}$ ,  $i = 1, 2, 3$ ,  $\lambda_i$ ,  $i = 1, 2, 3$  y  $\lambda_\beta$ , que valora gráficamente la estabilidad asumida tras la realización de 5000 iteraciones Gibbs.)

### 3.4 Modelos Bayesianos $m$ -jerárquicos

El objetivo de esta sección es el análisis de los modelos Bayesianos  $m$ -jerárquicos normal y normal-gamma, así como sus correspondientes versiones mínimo informativas. Estos modelos, tal y como comentábamos en la Introducción, cumplen dos objetivos, por una parte, generalizan los modelos normales y normales-gamma, que hemos definido y analizado en el capítulo 2 y en la sección anterior, de modo que los resultados inferenciales para estos modelos pueden tomarse y demostrarse (tal y como hemos hecho), como casos particulares de los resultados inferenciales para modelos  $m$ -jerárquicos. Y además, presentan un modelo flexible para representar y estudiar otros modelos para poblaciones en una y dos etapas y para poblaciones estructuradas en tres o más etapas.

En esta sección, denotaremos por  $\mathbf{Y}$  al vector de características de la población (ya sea la población monoetápica, bietápica, trietápica o  $(m - 1)$ -etápica), y por  $N$ , el tamaño de la población, que asumiremos conocido y la muestra en la que se apoyan nuestras inferencias consiste en la selección mediante un procedimiento de muestreo no informativo y de tamaño fijo de  $n$  elementos distintos de la población.

Los lemas a los que hacen referencia las demostraciones de los siguientes teoremas y corolarios se encuentran enunciados y demostrados en el apéndice B.

#### 3.4.1 Modelo Bayes de Regresión $m$ -jerárquico Normal

El modelo Bayes de Regresión  $m$ -jerárquico Normal que definiremos a continuación, se aplica cuando la varianza de la población es conocida.

**Definición 3.6** *Llamaremos modelo Bayes de Regresión con  $m$  jerarquías normal al definido por el siguiente modelo jerárquico:*

$$\mathbf{Y}|\theta_1 \sim N(\mathbf{A}_1\theta_1, \mathbf{V}_1), \quad (3.54)$$

$$\theta_1|\theta_2 \sim N(\mathbf{A}_2\theta_2, \mathbf{V}_2), \quad (3.55)$$

$$\vdots \quad \vdots \quad \vdots$$

$$\theta_{m-1}|\theta_m \sim N(\mathbf{A}_m\theta_m, \mathbf{V}_m), \quad (3.56)$$

$$\theta_m \sim N(\mathbf{A}_{m+1}, \mathbf{V}_{m+1}), \quad (3.57)$$

donde  $\mathbf{Y}$  es el vector de características de la población (representando, tanto una estructura monoetápica como bietápica, trietápica, ..., o  $(m-1)$ -etápica de la población),  $\{\mathbf{A}_i\}_{i=1}^{m+1}$  son matrices conocidas y  $\{\mathbf{V}_i\}_{i=1}^{m+1}$  son matrices conocidas simétricas definidas positivas, todas ellas de dimensiones adecuadas. ( $\theta_m$  suelen ser variable aleatoria y por tanto  $\mathbf{A}_{m+1}$  y  $\mathbf{V}_{m+1}$  de dimensión 1.)

Este modelo fue estudiado por Malec y Sedransk(85)[72], que obtuvieron la distribución predictiva del vector no observado y particularizaron el modelo y sus resultados inferenciales para una población estructurada en tres etapas. Otros casos particulares del modelo son: el modelo Bayes de Regresión Normal (definición 2.1), el modelo Bayes de Regresión en dos etapas con una jerarquía Normal (definición 2.6) y el modelo Bayes de Regresión en dos etapas con dos jerarquías Normal (definición 3.1).

Las distribuciones predictivas del vector poblacional no observado,  $\mathbf{Y}_u$ , la media no observada,  $\bar{Y}_u$  y el estimador Bayes de la media poblacional,  $\hat{Y}$  y su varianza condicionada a la muestra  $(s, \mathbf{y}_s)$  ( $s$ , índices de los elementos muestreados e  $\mathbf{y}_s$ , el vector observado de la población), vienen dados por el siguiente teorema.

**Teorema 3.5** *Consideremos el modelo Bayesiano m-jerárquico normal dado en la definición 3.6, entonces se tiene que:*

(a)  $\mathbf{Y}_u|(s, \mathbf{y}_s)$  se distribuye normal  $(N - n)$ -variante con media y varianza dadas respectivamente por:

$$\begin{aligned} E(\mathbf{Y}_u|(s, \mathbf{y}_s)) &= \mathbf{A}_u \hat{\theta}_1^B + \mathbf{V}_{su}^t \mathbf{V}_s^{-1} (\mathbf{y}_s - \mathbf{A}_s \hat{\theta}_1^B), \\ \text{Var}(\mathbf{Y}_u|(s, \mathbf{y}_s)) &= \mathbf{V}_u - \mathbf{V}_{su}^t \mathbf{V}_s^{-1} \mathbf{V}_{su} + \\ &\quad + (\mathbf{A}_u - \mathbf{V}_{su}^t \mathbf{V}_s^{-1} \mathbf{A}_s) \mathbf{D} (\mathbf{A}_u - \mathbf{V}_{su}^t \mathbf{V}_s^{-1} \mathbf{A}_s)^t, \end{aligned} \quad (3.58)$$

(b)  $\bar{Y}_u|(s, \mathbf{y}_s)$  se distribuye normal con media y varianza dadas por:

$$\begin{aligned} E(\bar{Y}_u|(s, \mathbf{y}_s)) &= \frac{1}{N-n} \mathbf{1}_{N-n}^t [\mathbf{A}_u \hat{\theta}_1^B + \mathbf{V}_{su}^t \mathbf{V}_s^{-1} (\mathbf{y}_s - \mathbf{A}_s \hat{\theta}_1^B)], \\ \text{Var}(\bar{Y}_u|(s, \mathbf{y}_s)) &= \frac{1}{(N-n)^2} \{ \mathbf{1}_{N-n}^t [\mathbf{V}_u - \mathbf{V}_{su}^t \mathbf{V}_s^{-1} \mathbf{V}_{su}] \mathbf{1}_{N-n} + \\ &\quad + \mathbf{1}_{N-n}^t (\mathbf{A}_u - \mathbf{V}_{su}^t \mathbf{V}_s^{-1} \mathbf{A}_s) \mathbf{D} (\mathbf{A}_u - \mathbf{V}_{su}^t \mathbf{V}_s^{-1} \mathbf{A}_s)^t \mathbf{1}_{N-n} \}, \end{aligned} \quad (3.59)$$

y (c) el estimador Bayes para la media poblacional y su varianza condicionada a la muestra vienen dados por:

$$\begin{aligned} \hat{Y} &= \frac{1}{N} \mathbf{1}_n^t \mathbf{y}_s + \frac{1}{N} \mathbf{1}_{N-n}^t [\mathbf{A}_u \hat{\theta}_1^B + \mathbf{V}_{su}^t \mathbf{V}_s^{-1} (\mathbf{y}_s - \mathbf{A}_s \hat{\theta}_1^B)], \\ \text{Var}(\bar{Y}_u|(s, \mathbf{y}_s)) &= \frac{1}{(N-n)^2} \{ \mathbf{1}_{N-n}^t [\mathbf{V}_u - \mathbf{V}_{su}^t \mathbf{V}_s^{-1} \mathbf{V}_{su}] \mathbf{1}_{N-n} + \\ &\quad + \mathbf{1}_{N-n}^t (\mathbf{A}_u - \mathbf{V}_{su}^t \mathbf{V}_s^{-1} \mathbf{A}_s) \mathbf{D} (\mathbf{A}_u - \mathbf{V}_{su}^t \mathbf{V}_s^{-1} \mathbf{A}_s)^t \mathbf{1}_{N-n} \}, \end{aligned} \quad (3.60)$$

donde  $\mathbf{y}_s = (y_i : i \in s)^t$ ,  $\mathbf{Y}_u = (Y_i : i \notin s)^t$ ,  $\mathbf{A}_s$  es la matriz formada por las filas de la matriz  $\mathbf{A}_1$  correspondientes a índices  $i \in s$ ,  $\mathbf{A}_u$  es la matriz formada por las filas de la

matriz  $\mathbf{A}_1$  correspondientes a índices  $i \notin s$ ,  $\mathbf{V}_s = \text{Var}_{\mathbf{Y}|\theta_1}(\mathbf{Y}_s)$ ,  $\mathbf{V}_u = \text{Var}_{\mathbf{Y}|\theta_1}(\mathbf{Y}_u)$ ,  $\mathbf{V}_{su} = \text{Cov}_{\mathbf{Y}|\theta_1}(\mathbf{Y}_s, \mathbf{Y}_u)$  y

$$\begin{aligned}
\mathbf{D}^{-1} &= \mathbf{A}_s^t \mathbf{V}_s^{-1} \mathbf{A}_s + (\mathbf{V} + \mathbf{A} \mathbf{V}_{m+1} \mathbf{A}^t)^{-1}, \\
\mathbf{d} &= \mathbf{A}_s^t \mathbf{V}_s^{-1} \mathbf{y}_s + (\mathbf{V} + \mathbf{A} \mathbf{V}_{m+1} \mathbf{A}^t)^{-1} \mathbf{A} \mathbf{A}_{m+1}, \\
\hat{\theta}_1^B &= E(\theta_1 | (s, \mathbf{y}_s)) = \mathbf{D} \mathbf{d}, \\
\mathbf{A} &= \mathbf{A}_2 \mathbf{A}_3 \cdots \mathbf{A}_m, \\
\mathbf{V} &= \mathbf{V}_2 + \mathbf{A}_2 \mathbf{V}_3 \mathbf{A}_2^t + \cdots + (\mathbf{A}_2 \mathbf{A}_3 \cdots \mathbf{A}_{m-1}) \mathbf{V}_m (\mathbf{A}_2 \mathbf{A}_3 \cdots \mathbf{A}_{m-1})^t,
\end{aligned} \tag{3.61}$$

siendo para  $m = 1$   $\mathbf{A} = \mathbf{I}$  y  $\mathbf{V} = \mathbf{0}$  ( $\mathbf{I}$  matriz identidad y  $\mathbf{0}$  matriz de ceros de las dimensiones del vector  $\theta_1$ ).

**Demostración.** Dividiremos la demostración (que se apoya en el lema B.8) de los resultados del teorema en varios pasos:

1.- A partir de (3.55), ..., (3.56) y (3.57) obtenemos aplicando el resultado (a) del lema B.8 iterativamente la distribución de  $\theta_1$  que es:

$$\theta_1 \sim N(\mathbf{A} \mathbf{A}_{m+1}, \mathbf{V} + \mathbf{A} \mathbf{V}_{m+1} \mathbf{A}^t). \tag{3.62}$$

2.- A partir de (3.54) aplicando propiedades básicas de la distribución normal, tenemos que:

$$\mathbf{Y}_s | \theta_1 \sim N(\mathbf{A}_s \theta_1, \mathbf{V}_s). \tag{3.63}$$

3.- A partir de (3.63) y (3.62) obtenemos aplicando el resultado (b) del lema B.8 la distribución de  $\theta_1 | (s, \mathbf{y}_s)$  que es:

$$\theta_1 | (s, \mathbf{y}_s) \sim N(\mathbf{D} \mathbf{d}, \mathbf{D}). \tag{3.64}$$

4.- A partir de (3.54) aplicando propiedades básicas de la distribución normal, tenemos que:

$$\mathbf{Y}_u | (s, \mathbf{y}_s), \theta_1 \sim N(\mathbf{A}_u \theta_1 + \mathbf{V}_{su}^t \mathbf{V}_s^{-1} (\mathbf{y}_s - \mathbf{A}_s \theta_1), \mathbf{V}_u - \mathbf{V}_{su}^t \mathbf{V}_s^{-1} \mathbf{V}_{su}) \tag{3.65}$$

5.- A partir de (3.65) y (3.64) aplicando el resultado (a) del lema B.8 se obtiene la distribución de  $\mathbf{Y}_u | (s, \mathbf{y}_s)$  dada en (3.58).

6.- A partir de (3.58) y aplicando que  $\bar{\mathbf{Y}}_u = \frac{1}{N-n} \mathbf{1}_{N-n}^t \mathbf{Y}_u$  y que la combinación lineal de normales es normal se obtiene la distribución de  $\bar{\mathbf{Y}}_u | (s, \mathbf{y}_s)$  dada en (3.59).

7.- A partir de (3.59) y de la definición de estimador Bayes de la media poblacional, se obtiene el resultado (c) de este teorema dado en (3.60).

C.S.Q.D.

Notemos que el estimador Bayes de Regresión para este modelo puede escribirse, como una “combinación lineal” del estimador de regresión clásico,  $\hat{\theta}_1 = (\mathbf{A}_s^t \mathbf{V}_s^{-1} \mathbf{A}_s)^{-1} \mathbf{A}_s^t \mathbf{V}_s^{-1} \mathbf{y}_s$ , y de la pendiente a priori asignada a la población,  $\mathbf{A} \mathbf{A}_{m+1}$ , dado que:

$$\begin{aligned} \hat{\theta}_1^B &= \alpha_1 \hat{\theta}_1 + (\mathbf{I} - \alpha_1) \mathbf{A} \mathbf{A}_{m+1}, \\ \alpha_1 &= [\mathbf{A}_s^t \mathbf{V}_s^{-1} \mathbf{A}_s + (\mathbf{V} + \mathbf{A} \mathbf{V}_{m+1} \mathbf{A}^t)^{-1}]^{-1} \mathbf{A}_s^t \mathbf{V}_s^{-1} \mathbf{A}_s, \end{aligned} \quad (3.66)$$

con  $\alpha_1$  un factor de encogimiento ( $\alpha_1 = (\mathbf{I} - \alpha_1)(\mathbf{V} + \mathbf{A} \mathbf{V}_{m+1} \mathbf{A}^t) \mathbf{A}_s^t \mathbf{V}_s^{-1} \mathbf{A}_s$ ), e  $\mathbf{I}$ , una matriz identidad de las dimensiones del vector  $\theta_1$ . Estas expresiones se deducen de manera inmediata del teorema 3.5.

### 3.4.2 El modelo Bayes de Regresión $m$ -jerárquico Normal - Gamma

El modelo Bayes de Regresión  $m$ -jerárquico Normal-Gamma, cuya definición veremos a continuación, se aplica cuando la varianza de la población es desconocida.

Como casos particulares del modelo tenemos: el modelo Bayes de Regresión Normal-Gamma (definición 2.2), el modelo Bayes de Regresión en dos etapas con una jerarquía Normal-Gamma (definición 2.7) y el modelo Bayes de Regresión en dos etapas con dos jerarquías Normal-Gamma (definición 3.2). El modelo Bayes de Regresión  $m$ -jerárquico Normal - Gamma en el caso  $Cov(\mathbf{Y}_s, \mathbf{Y}_u | \theta_1, \sigma^2) = \mathbf{0}$  es un caso particular del modelo  $m$ -jerárquico estudiado en Bernardo y Girón(91)[7].

**Definición 3.7** *Llamaremos modelo Bayes de Regresión en  $m$  jerarquías Normal-Gamma al definido por el siguiente modelo jerárquico:*

$$\mathbf{Y} | \theta_1, \sigma^2 \sim N(\mathbf{A}_1 \theta_1, \sigma^2 \mathbf{W}_1), \quad (3.67)$$

$$\theta_1 | \theta_2, \sigma^2 \sim N(\mathbf{A}_2 \theta_2, \sigma^2 \mathbf{W}_2), \quad (3.68)$$

$\vdots$   
 $\vdots$   
 $\vdots$

$$\theta_{m-1} | \theta_m, \sigma^2 \sim N(\mathbf{A}_m \theta_m, \sigma^2 \mathbf{W}_m), \quad (3.69)$$

$$\theta_m | \sigma^2 \sim N(\mathbf{A}_{m+1}, \sigma^2 \mathbf{W}_{m+1}), \quad (3.70)$$

$$\sigma^2 \sim Ga^{-1}(a_0, b_0), \quad (3.71)$$

donde  $\mathbf{Y}$  es el vector de características de la población (representando, tanto una estructura monoetápica como bietápica, trietápica, ...,  $(m-1)$ -etápica),  $\{\mathbf{A}_i\}_{i=1}^{m+1}$  son matrices conocidas,  $\{\mathbf{W}_i\}_{i=1}^{m+1}$  son matrices conocidas simétricas definidas positivas, todas ellas de dimensiones adecuadas,  $a_0$  y  $b_0$  escalares positivos conocidos. ( $\theta_m$  suele ser variable aleatoria y por tanto  $\mathbf{A}_{m+1}$  y  $\mathbf{W}_{m+1}$  de dimensión 1).

Las distribuciones predictivas para el vector poblacional no observado, la media no observada y el estimador de la media poblacional y su varianza condicionada a la muestra, vienen dadas por el siguiente teorema.

**Teorema 3.6** Consideremos el modelo Bayes  $m$ -jerárquico normal-gamma dado en la definición 3.7, entonces se tiene que:

(a)  $\mathbf{Y}_u|(s, \mathbf{y}_s)$  se distribuye  $t$  ( $N-n$ )-variante con  $2a_1$  grados de libertad y media y varianza dadas, para  $a_1 > 1$ , respectivamente por:

$$\begin{aligned} E(\mathbf{Y}_u|(s, \mathbf{y}_s)) &= \mathbf{A}_u \hat{\theta}_1^B + \mathbf{W}_{su}^t \mathbf{W}_s^{-1} (\mathbf{y}_s - \mathbf{A}_s \hat{\theta}_1^B), \\ \text{Var}(\mathbf{Y}_u|(s, \mathbf{y}_s)) &= \frac{b_1}{a_1 - 1} \{ \mathbf{W}_u - \mathbf{W}_{su}^t \mathbf{W}_s^{-1} \mathbf{W}_{su} + \\ &\quad + (\mathbf{A}_u - \mathbf{W}_{su}^t \mathbf{W}_s^{-1} \mathbf{A}_s) \mathbf{D} (\mathbf{A}_u - \mathbf{W}_{su}^t \mathbf{W}_s^{-1} \mathbf{A}_s)^t \}, \end{aligned} \quad (3.72)$$

(b)  $\bar{Y}_u|(s, \mathbf{y}_s)$  se distribuye  $t$  univariante con  $2a_1$  grados de libertad y media y varianza dadas, para  $a_1 > 1$ , por:

$$\begin{aligned} E(\bar{Y}_u|(s, \mathbf{y}_s)) &= \frac{1}{N-n} \mathbf{1}_{N-n}^t [\mathbf{A}_s \hat{\theta}_1^B + \mathbf{W}_{su}^t \mathbf{W}_s^{-1} (\mathbf{y}_s - \mathbf{A}_s \hat{\theta}_1^B)], \\ \text{Var}(\bar{Y}_u|(s, \mathbf{y}_s)) &= \frac{1}{(N-n)^2} \frac{b_1}{a_1 - 1} \{ \mathbf{1}_{N-n}^t [\mathbf{W}_u - \mathbf{W}_{su}^t \mathbf{W}_s^{-1} \mathbf{W}_{su}] \mathbf{1}_{N-n} + \\ &\quad + \mathbf{1}_{N-n}^t (\mathbf{A}_u - \mathbf{W}_{su}^t \mathbf{W}_s^{-1} \mathbf{A}_s) \mathbf{D} (\mathbf{A}_u - \mathbf{W}_{su}^t \mathbf{W}_s^{-1} \mathbf{A}_s)^t \mathbf{1}_{N-n} \}, \end{aligned} \quad (3.73)$$

y (c) el estimador Bayes para la media poblacional y su varianza condicionada a la muestra vienen dados, para  $a_1 > 1$ , por:

$$\begin{aligned} \hat{Y} &= \frac{1}{N} \mathbf{1}_n \mathbf{y}_s + \frac{1}{N} \mathbf{1}_{N-n}^t [\mathbf{A}_s \hat{\theta}_1^B + \mathbf{W}_{su}^t \mathbf{W}_s^{-1} (\mathbf{y}_s - \mathbf{A}_s \hat{\theta}_1^B)], \\ \text{Var}(\hat{Y}|(s, \mathbf{y}_s)) &= \frac{1}{N^2} \frac{b_1}{a_1 - 1} \{ \mathbf{1}_{N-n}^t [\mathbf{W}_u - \mathbf{W}_{su}^t \mathbf{W}_s^{-1} \mathbf{W}_{su}] \mathbf{1}_{N-n} + \\ &\quad + \mathbf{1}_{N-n}^t (\mathbf{A}_u - \mathbf{W}_{su}^t \mathbf{W}_s^{-1} \mathbf{A}_s) \mathbf{D} (\mathbf{A}_u - \mathbf{W}_{su}^t \mathbf{W}_s^{-1} \mathbf{A}_s)^t \mathbf{1}_{N-n} \}, \end{aligned} \quad (3.74)$$

donde  $\mathbf{y}_s = (y_i : i \in s)^t$ ,  $\mathbf{Y}_u = (Y_i : i \notin s)^t$ ,  $\mathbf{A}_s$  es la matriz formada por las filas de la matriz  $\mathbf{A}_1$  correspondientes a índices  $i \in s$ ,  $\mathbf{A}_u$  es la matriz formada por

las filas de la matriz  $\mathbf{A}_1$  correspondientes a índices  $i \notin s$ ,  $\mathbf{W}_s = \sigma^{-2} \text{Var}_{\mathbf{Y}|\theta_1}(\mathbf{Y}_s)$ ,  $\mathbf{W}_u = \sigma^{-2} \text{Var}_{\mathbf{Y}|\theta_1}(\mathbf{Y}_u)$ ,  $\mathbf{W}_{su} = \sigma^{-2} \text{Cov}_{\mathbf{Y}|\theta_1}(\mathbf{Y}_s, \mathbf{Y}_u)$  y

$$\begin{aligned}
\mathbf{D}^{-1} &= \mathbf{A}_s^t \mathbf{W}_s^{-1} \mathbf{A}_s + (\mathbf{W} + \mathbf{A} \mathbf{W}_{m+1} \mathbf{A}^t)^{-1}, \\
\mathbf{d} &= \mathbf{A}_s^t \mathbf{W}_s^{-1} \mathbf{y}_s + (\mathbf{W} + \mathbf{A} \mathbf{W}_{m+1} \mathbf{A}^t)^{-1} \mathbf{A} \mathbf{A}_{m+1}, \\
\hat{\theta}_1^B &= E(\theta_1 | (s, \mathbf{y}_s)) = \mathbf{D} \mathbf{d}, \\
\mathbf{E}^{-1} &= \mathbf{W}_s + \mathbf{A}_s (\mathbf{W} + \mathbf{A} \mathbf{W}_{m+1} \mathbf{A}^t) \mathbf{A}_s^t, \\
\mathbf{A} &= \mathbf{A}_2 \mathbf{A}_3 \cdots \mathbf{A}_m, \\
\mathbf{W} &= \mathbf{W}_2 + \mathbf{A}_2 \mathbf{W}_3 \mathbf{A}_2^t + \dots + (\mathbf{A}_2 \mathbf{A}_3 \cdots \mathbf{A}_{m-1}) \mathbf{W}_m (\mathbf{A}_2 \mathbf{A}_3 \cdots \mathbf{A}_{m-1})^t, \\
a_1 &= a_0 + \frac{n}{2}, \\
b_1 &= b_0 + \frac{1}{2} (\mathbf{y}_s - \mathbf{A}_s \mathbf{A} \mathbf{A}_{m+1})^t \mathbf{E} (\mathbf{y}_s - \mathbf{A}_s \mathbf{A} \mathbf{A}_{m+1}),
\end{aligned} \tag{3.75}$$

siendo, para  $m = 1$ ,  $\mathbf{A} = \mathbf{I}$  y  $\mathbf{W} = \mathbf{0}$  ( $\mathbf{I}$  matriz identidad y  $\mathbf{0}$  matriz de ceros de las dimensiones del vector  $\theta_1$ ).

**Demostración.** La demostración de este teorema se basa en los resultados obtenidos en el teorema 3.5 y en la aplicación de los lemas B.8 y B.9. En primer lugar obtendremos la distribución de  $\sigma^2 | (s, \mathbf{y}_s)$  para lo cual realizaremos los siguientes pasos:

1.- A partir de (3.68), ..., (3.69) y (3.70), obtenemos aplicando el resultado (a) del lema B.8 iterativamente la distribución de  $\theta_1 | \sigma^2$  que es:

$$\theta_1 | \sigma^2 \sim N(\mathbf{A} \mathbf{A}_{m+1}, \sigma^2 (\mathbf{W} + \mathbf{A} \mathbf{W}_{m+1} \mathbf{A}^t)). \tag{3.76}$$

2.- A partir de (3.67) aplicando propiedades de la distribución normal básicas tenemos que:

$$\mathbf{Y}_s | \theta_1, \sigma^2 \sim N(\mathbf{A}_s \theta_1, \sigma^2 \mathbf{W}_s). \tag{3.77}$$

3.- A partir de (3.77) y (3.76) obtenemos aplicando el resultado (a) del lema B.8 la distribución de  $\mathbf{Y}_s | \sigma^2$  que es:

$$\mathbf{Y}_s | \sigma^2 \sim N_n(\mathbf{A}_s \mathbf{A} \mathbf{A}_{m+1}, \sigma^2 [\mathbf{W}_s + \mathbf{A}_s (\mathbf{W} + \mathbf{A} \mathbf{W}_{m+1} \mathbf{A}^t) \mathbf{A}_s^t]) \tag{3.78}$$

4.- A partir de (3.78) y (3.71) obtenemos aplicando el resultado (b) del lema B.9 la distribución de  $\sigma^2 | (s, \mathbf{y}_s)$  que viene dada por:

$$\sigma^2 | (s, \mathbf{y}_s) \sim Ga^{-1}(a_0 + \frac{n}{2}, b_1). \tag{3.79}$$

El resultado (a) se comprueba aplicando el resultado (a) del lema B.9 sobre la distribución de  $\mathbf{Y}_u|(s, \mathbf{y}_s), \sigma^2$  que se obtiene directamente de (3.58) tomando  $\mathbf{V}_i = \sigma^2 \mathbf{W}_i$ ,  $i = 1, 2, \dots, m+1$  y sobre (3.79), el resultado (b) aplicando<sup>6</sup> el mismo resultado sobre la distribución de  $\bar{Y}_u|(s, \mathbf{y}_s), \sigma^2$  que se obtiene de (3.59) con la transformación anterior y sobre (3.79) y el resultado (c) a partir de (3.73) y de la definición de estimador Bayes de la media poblacional.

C.S.Q.D.

### 3.4.3 Modelos Bayes $m$ -jerárquicos mínimo informativos

En esta subsección presentamos unas versiones mínimo informativas de los dos modelos  $m$ -jerárquicos anteriores, así como los resultados inferenciales habituales asociados a las mismas, que se aplican cuando no deseamos introducir información sobre los parámetros de las distribuciones iniciales.

Estos modelos mínimo informativos generalizan los modelos mínimo informativos definidos en el capítulo 2 y en la sección 3.2, aplicándose los resultados inferenciales que describiremos a continuación para demostrar los correspondientes resultados para estos modelos.

#### Modelo Bayes $m$ -jerárquico normal mínimo informativo

**Definición 3.8** *Llamaremos modelo Bayes en  $m$  jerarquías normal mínimo informativo a un modelo Bayes en  $m$  jerarquías normal en el que la distribución que asumimos sobre  $\theta_m$  es una mínimo informativa de la forma  $\pi(\theta_m) \propto 1$ .*

**Corolario 3.3** *Consideremos el modelo Bayes en  $m$  jerarquías normal mínimo informativo dado en la definición 3.8, entonces se tiene que las distribuciones predictivas para el vector poblacional no observado y la media de la población no observada y estimador de la media poblacional son los obtenidos en (3.58), (3.59) y (3.60), tomando:*

$$\begin{aligned} \mathbf{D}^{-1} &= \mathbf{A}_s^t \mathbf{V}_s^{-1} \mathbf{A}_s + \mathbf{V}^{-1} - \mathbf{V}^{-1} \mathbf{A} (\mathbf{A}^t \mathbf{V}^{-1} \mathbf{A})^{-1} \mathbf{A}^t \mathbf{V}^{-1}, \\ \mathbf{d} &= \mathbf{A}_s^t \mathbf{V}_s^{-1} \mathbf{y}_s, \\ \hat{\theta}_1^B &= E(\theta_1|(s, \mathbf{y}_s)) = \mathbf{D} \mathbf{d}. \end{aligned} \tag{3.80}$$

**Demostración.** La demostración del corolario es inmediata a partir del teorema 3.5 tomando  $\mathbf{V}_{m+1}^{-1} \rightarrow 0$  y aplicando el lema B.1.

C.S.Q.D.

---

<sup>6</sup>Una demostración alternativa consiste en aplicar sobre la distribución predictiva del vector poblacional no observado obtenida en la primera parte de la demostración de la propiedad de que la combinación lineal de  $t$  es  $t$

## Modelo Bayes en $m$ -jerarquías Normal-Gamma mínimo informativo

**Definición 3.9** Llamaremos modelo Bayes en  $m$  jerarquías normal-gamma mínimo informativo a un modelo Bayes en  $m$  jerarquías Normal-Gamma en el que las distribuciones (3.70) y (3.71) son sustituidas por una distribución conjunta para  $(\theta_m, \sigma^2)$  mínimo informativa,<sup>7</sup> de la forma  $\pi(\theta_m, \sigma^2) \propto \frac{1}{\sigma^2}$ .

**Corolario 3.4** Consideremos el modelo Bayes en  $m$  jerarquías normal-gamma mínimo informativo dado en la definición 3.9, entonces se tiene que las distribuciones predictivas para el vector no observado de la población y la media no observada y estimador de la media poblacional son los obtenidos en (3.72), (3.73) y (3.74), tomando, en este caso:

$$\begin{aligned} \mathbf{D}^{-1} &= \mathbf{A}_s^t \mathbf{W}_s^{-1} \mathbf{A}_s + \mathbf{W}^{-1} - \mathbf{W}^{-1} \mathbf{A} (\mathbf{A}^t \mathbf{W}^{-1} \mathbf{A})^{-1} \mathbf{A}^t \mathbf{W}^{-1}, \\ \mathbf{d} &= \mathbf{A}_s^t \mathbf{W}_s^{-1} \mathbf{y}_s, \\ \hat{\theta}_1^B &= E(\theta_1 | (s, \mathbf{y}_s)) = \mathbf{D} \mathbf{d}, \\ \mathbf{E}^{-1} &= \mathbf{W}_s + \mathbf{A}_s \mathbf{W} \mathbf{A}_s^t, \\ a_1 &= \frac{n - r}{2}, \end{aligned} \quad (3.81)$$

<sup>7</sup>Hemos utilizado para describir la situación mínimo informativa cuando la varianza es desconocida a la distribución inicial impropia  $\pi(\theta_m, \sigma^2) \propto \frac{1}{\sigma^2}$ , esta elección no es única y la distribución anterior ha sido elegida porque parece la más conveniente para el conjunto de modelos descritos mínimo informativos con varianza desconocida de los capítulos 2 y 3, al ser común a todos ellos y en la consideración de que los resultados inferenciales para las restantes distribuciones mínimo informativas habituales son similares (si  $n$  es grande en comparación a  $r$ , la dimensión del vector  $\theta_m$ ) y difieren entre sí únicamente en el número de grados de libertad de la  $t$ .

A continuación pasamos a describir brevemente cómo se obtienen las distribuciones mínimo informativas más habituales entre las que se encuentra la aplicada en esta tesis. Considerada la familia normal-gamma invertida de distribuciones conjugadas con el modelo definido de (3.67) a (3.69):

$$\pi(\theta_m, \sigma^2) \propto (\sigma^2)^{-(a_0 + \frac{r}{2} + 1)} \exp \left\{ -\frac{1}{2\sigma^2} [(\theta_m - \mathbf{A}_{m+1})^t \mathbf{W}_{m+1}^{-1} (\theta_m - \mathbf{A}_{m+1}) + 2b_0] \right\},$$

podemos describir tres procedimientos para la obtención de una distribución mínimo informativa para  $(\theta_m, \sigma^2)$ :

- Obtener en primer lugar la distribución mínimo informativa habitual para  $\theta_m$ , hacer  $\mathbf{W}_{m+1}^{-1} \rightarrow 0$ , y en segundo lugar para obtener la distribución marginal impropia para  $\sigma^2$  ( $\pi(\sigma^2) \propto \frac{1}{\sigma^2}$ ), tomar  $a_0 = -\frac{r}{2}$  y  $b_0 = 0$ . El resultado es la distribución  $\pi(\theta_m, \sigma^2) \propto \frac{1}{\sigma^2}$ .
- Fijarse en primer lugar en la distribución marginal sobre  $\sigma^2$  y hacerla informativa, tomar  $a_0 = 0$ ,  $b_0 = 0$ , y después para obtener la distribución impropia para  $\theta_m$ , hacer  $\mathbf{W}_{m+1}^{-1} \rightarrow 0$ . El resultado obtenido es la distribución  $\pi(\theta_m, \sigma^2) \propto (\sigma^2)^{-(\frac{r}{2} + 1)}$ , que es la distribución inicial de Jeffreys.
- Otras distribuciones mínimo informativas razonables pueden obtenerse por los métodos anteriores tomando como distribución marginal impropia para  $\sigma^2$ ,  $\pi(\sigma^2) \propto 1$ .

$$b_1 = \frac{1}{2} \mathbf{y}_s^t [\mathbf{E} - \mathbf{E} \mathbf{A}_s \mathbf{A} (\mathbf{A}^t \mathbf{A}_s^t \mathbf{E} \mathbf{A}_s \mathbf{A})^{-1} \mathbf{A}^t \mathbf{A}_s^t \mathbf{E}] \mathbf{y}_s,$$

donde  $r$  es la dimensión de  $\theta_m$ .

**Demostración.** La demostración de este corolario se basa en los resultados obtenidos en el corolario 3.3 y en la aplicación de los lemas B.8, B.9 y B.10. En primer lugar obtendremos la distribución de  $\sigma^2|(s, \mathbf{y}_s)$  para lo cual realizaremos los siguientes pasos:

1.- A partir de (3.67) aplicando propiedades básicas de la distribución normal, tenemos que:

$$\mathbf{Y}_s | \theta_1, \sigma^2 \sim N(\mathbf{A}_s \theta_1, \sigma^2 \mathbf{W}_s). \quad (3.82)$$

2.- A partir de (3.82) y de (3.68), ..., (3.69) obtenemos, aplicando el resultado (a) del lema B.8 iterativamente la distribución de  $\mathbf{Y}_s | \theta_m, \sigma^2$  que es:

$$\mathbf{Y}_s | \theta_m, \sigma^2 \sim N(\mathbf{A}_s \mathbf{A} \theta_m, \sigma^2 (\mathbf{W}_s + \mathbf{A}_s \mathbf{W} \mathbf{A}_s^t)). \quad (3.83)$$

3.- A partir de (3.83) y la distribución mínimo informativa de  $(\theta_m, \sigma^2)$ , obtenemos, aplicando el resultado (b) del lema B.10 la distribución de  $\sigma^2|(s, \mathbf{y}_s)$  que viene dada por:

$$\sigma^2|(s, \mathbf{y}_s) \sim Ga^{-1}(a_1, b_1). \quad (3.84)$$

La distribución predictiva sobre el vector poblacional no observado se obtiene aplicando el resultado (a) del lema B.9 sobre la distribución de  $\mathbf{Y}_u|(s, \mathbf{y}_s), \sigma^2$  que se obtiene directamente del corolario 3.3 tomando  $\mathbf{V}_i = \sigma^2 \mathbf{W}_i, i = 1, 2, \dots, m$  y sobre (3.84), la distribución predictiva de la media no observada se obtiene <sup>8</sup> aplicando el mismo resultado sobre la distribución de  $\bar{\mathbf{Y}}_u|(s, \mathbf{y}_s), \sigma^2$  que se obtiene de este mismo corolario con la transformación anterior y sobre (3.84) y el estimador Bayes de la media poblacional y la varianza de la media poblacional condicionada a la muestra, a partir de la distribución predictiva de la media poblacional no observada obtenida en este resultado y de la definición de estimador Bayes de la media poblacional.

C.S.Q.D.

<sup>8</sup>Una demostración alternativa consiste en aplicar sobre la distribución predictiva del vector poblacional no observado obtenida en la primera parte de la demostración de la combinación lineal de  $t$  es  $t$

### 3.5 Comentarios finales

Los modelos descritos y analizados en este capítulo y en el capítulo 2 constituyen un sustrato teórico potente para poder modelizar un número bastante amplio de poblaciones finitas y en particular aquellas que resultan de asumir un determinado procedimiento de muestreo no aleatorio (ya comentábamos que una población puede asumirse estructurada por efecto de un muestreo estratificado, cluster, por rutas aleatorias, ...). El objetivo de los próximos dos capítulos es el establecer modelos Bayes (que serán casos particulares de los modelos estudiados en estos capítulos o pequeñas generalizaciones de los mismos) que permitan inferir sobre poblaciones sobre las que se aplican (porque se considera más adecuado, más económico o simplemente porque la población no es censable) procedimientos de muestreo no aleatorios. Concretamente, en el capítulo 4 se modelizará una “misma” población en las situaciones de un muestreo aleatorio, estratificado y cluster comparando, para casos particulares sencillos, las inferencias obtenidas por los tres mecanismos de muestreo en la estimación de la media poblacional. Y en el capítulo 5 se presentarán tres modelos para analizar una población muestreada por rutas aleatorias que incorporan de manera sencilla, la incertidumbre adicional derivada del uso de una ruta aleatoria en lugar de un muestreo aleatorio y permiten diagnosticar la calidad de las inferencias derivadas cuando, como es habitual en la práctica, se hace uso en estas situaciones de muestreo de modelos para poblaciones muestreadas aleatoriamente.

## Capítulo 4

# Comparación Bayes de las inferencias obtenidas a partir de muestras aleatorias, estratificadas y cluster

### 4.1 Introducción

Supondremos para realizar este estudio que la población responde a una estructura monoetápica y analizaremos las inferencias que podemos obtener de estas poblaciones a través de muestras no aleatorias de las misma. Este análisis tiene su punto de partida en la realidad práctica del análisis en poblaciones finitas; así, en algunas ocasiones, no es posible la obtención de una muestra aleatoria porque no disponemos de un censo de la población y en otras ocasiones la práctica ha demostrado que otros tipos de muestreo como el estratificado y el cluster presentan algún tipo de ventaja (el estratificado en precisión de las inferencias y el cluster desde un punto de vista económico al permitir aumentar el número de elementos muestreados a un coste pequeño) frente al muestreo aleatorio de la población. En este capítulo nos centraremos en la segunda situación dejando para el capítulo 5 el análisis de poblaciones sin censo disponible.

Este capítulo tiene como objetivo la descripción de modelos Bayes apropiados para muestreos aleatorios, estratificados y cluster de una población y la comparación desde un punto de vista Bayes (a través del análisis de las distribuciones predictivas del vector no observado de la población dada la muestra) de la precisión de las inferencias obtenidas al estimar la media poblacional según los tres procedimientos de muestreo. En la práctica, suponiendo los mismos tamaños muestrales para los tres procedimientos de muestreo, las inferencias basadas en una muestra estratificada de la población suelen ser más precisas que las basadas en una muestra aleatoria de la misma y las infe-

rencias basadas en un muestreo cluster (al margen de sus ventajas económicas) menos precisas que las apoyadas en un muestreo estratificado o aleatorio de la población. En esta dirección, Cochran(77)[18][capítulo 5] establece, desde un punto de vista de población fija que, si los términos  $\frac{n_i}{N_i}$ ,  $i = 1, 2, \dots, K$  son despreciados (se consideran muy pequeños), un muestreo aleatorio estratificado con  $k = K$  y con asignación proporcional<sup>1</sup> de las  $n_i$ ,  $i = 1, 2, \dots, K$  es más preciso que un muestreo aleatorio simple, y Bayarri y Font(94)[6] demuestran, desde un punto de vista Bayesiano, que en ausencia de covariables y bajo determinadas hipótesis las inferencias sobre la media poblacional basadas en una muestra estratificada son más precisas que las basadas en una muestra aleatoria y las inferencias basadas en un muestreo cluster son menos precisas que las basadas en muestreos aleatorios o estratificados.

En este capítulo se recuperaran los resultados de Bayarri y Font(94)[6], y se realizarán estudios comparativos entre los tres procedimientos de muestreo en presencia de covariables. Para ello, nos apoyaremos en las definiciones sobre muestreo aleatorio (muestreo aleatorio simple), muestreos aleatorios estratificado y cluster sin reemplazamiento establecidas en la sección 1.2 y asumiremos que (habitualmente, los muestreos estratificados y cluster cumplen estas condiciones), el muestreo estratificado se ha apoyado en una división de la población en unidades (estratos) de modo que hay relativamente pocas unidades ( $K$  es relativamente pequeño), de las que se muestrean todas  $k = K$ , con un número relativamente grande de elementos ( $N_i$  grande) por unidad, los elementos de cada unidad son similares entre sí y la heterogeneidad entre unidades es grande. Y la muestra cluster se ha obtenido a partir de una división de la población según la cual, el número de unidades es grande ( $K$  grande) con un número de elementos por unidad pequeño ( $N_i$  pequeño), los elementos dentro de cada unidad son muy similares entre sí, la heterogeneidad entre unidades es muy grande (mayor que en un muestreo estratificado) y se muestrean todos los elementos de las unidades seleccionadas para ser muestreadas, esto es,  $n_i = N_i$ ,  $i \in \{i_1, i_2, \dots, i_k\}$ .

Para la realización de este estudio hemos dividido el capítulo en 5 secciones de las cuales, la sección 4.1 es la presente introducción sobre el tema. En la sección 4.2 se estudia la formulación en términos Bayesianos del problema, presentando dos modelos Bayes de regresión, el primero para modelizar un población sobre la que se aplica un muestreo aleatorio, y el segundo para modelizar a (aproximadamente) la misma población cuando la muestra se obtiene a partir de un muestreo en dos etapas. Este modelo es bastante flexible para representar las situaciones de muestreo estratificado y cluster comentadas arriba, y permite en combinación con el primer modelo, comparar analíticamente, en algunas situaciones sencillas, los tres procedimientos de muestreo. Esta comparación se realizará en la sección 4.3. Finalmente, en la sección 4.4 se analizarán los resultados obtenidos mediante un ejemplo numérico y en la sección 4.5 se realizará una breve discursión sobre las limitaciones del estudio y posibles líneas de trabajo futuras para subsanarlas.

---

<sup>1</sup>Un muestreo estratificado con asignación proporcional es aquel tal que,  $\frac{n_i}{n} = \frac{N_i}{N}$ ,  $\forall i = 1, 2, \dots, K$ .

## 4.2 Formulación del problema y análisis Bayes del modelo para muestras aleatorias

En esta sección introduciremos dos modelos para modelizar una misma población en dos situaciones de muestreo: el muestreo aleatorio y el muestreo en dos etapas. Para ello, aplicaremos en el primer caso un modelo Bayesiano en una etapa de los descritos en la sección 2.2, y en el segundo un modelo Bayesiano en dos etapas con dos jerarquías. En esta sección emplearemos los modelos Bayesianos en dos etapas con dos jerarquías para modelizar poblaciones muestreadas en dos etapas por dos razones, en primer lugar porque los muestreos en dos etapas suponen una estructuración de la población (en el caso de muestreos estratificados, se parte de una división de la población de acuerdo con los valores de alguna característica, habitualmente, socio-económica conocida para todos los elementos de la población que guarda relación con la característica en estudio, y en el muestreo cluster se aprovecha la estructuración de la población en unidades administrativas, familiares, ...) lo que sugiere la formulación de modelos en dos etapas. Y en segundo lugar, porque los modelos en dos jerarquías permiten una modelización flexible y sencilla de la población en estudio de modo que a partir de un mismo modelo podemos estudiar las inferencias obtenidas por un procedimiento estratificado y cluster.

Para la exposición de los modelos y resultados del capítulo se han empleado las notaciones reunidas en el apéndice A. Los mecanismos de muestreo aplicados se ajustan a las definiciones de la sección 1.2 (todos ellos sin reemplazamiento) cumpliendo los muestreos estratificados y cluster las propiedades señaladas en la introducción de este capítulo. La muestra obtenida de la población se representará con la notación habitual  $(s, \mathbf{y}_s)$  y cuando esta notación pueda dar lugar a confusiones, por:

- $(s^1, \mathbf{y}_{s^1})$  cuando se trate de una muestra aleatoria (m.a.s) de la población
- $(s^2, \mathbf{y}_{s^2})$  cuando se trate de una muestra estratificada de la población,
- y  $(s^3, \mathbf{y}_{s^3})$  cuando se trate de una muestra cluster de la población.

Los lemas en los que se apoyan los resultados del capítulo están enunciados y demostrados en el apéndice B.

### 4.2.1 Formulación del problema

Dada una población con vector de características continuo  $\mathbf{Y} = (Y_1, Y_2, \dots, Y_K)^t$ , para el cual asociada a cada variable  $Y_i$  tenemos una información auxiliar representada por un  $p$ -vector de covariables  $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{ip})$  conocido para  $i = 1, 2, \dots, N$ , podemos considerar el siguiente modelo de superpoblación:

$$Y_i | \mathbf{x}_i, \beta_\beta \sim N(\mathbf{x}_i \beta_\beta, \sigma^2 \mathbf{x}_i \mathbf{x}_i^t), \text{ independientes, } i = 1, 2, \dots, N \quad (4.1)$$

con  $\beta_\beta$  un  $p$ -vector desconocido y  $\sigma^2 > 0$  un escalar conocido. El modelo de superpoblación anterior es una generalización de los modelos de superpoblación estudiados

para  $p = 1$  por Royall(70)[91] desde un punto de vista clásico y por Ericson(69)[26] desde un punto de vista Bayesiano y permite obtener tomando  $p = 1$  y  $\mathbf{x}_i = 1$ ,  $i = 1, 2, \dots, N$ , el modelo de superpoblación habitual para describir un muestreo aleatorio. A partir de este modelo de superpoblación podemos describir el siguiente modelo Bayes en una etapa para poblaciones muestreadas aleatoriamente.

**Definición 4.1** Dado el vector poblacional  $\mathbf{Y} = (Y_1, Y_2, \dots, Y_T)^t$ , para el cual asociada a cada variable  $Y_i$  tenemos una información auxiliar representada por un vector  $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})$  conocido llamaremos modelo Bayes de Regresión para muestras aleatorias simples (m.a.s) con varianza conocida al definido por:

$$\begin{aligned} Y_i | \mathbf{x}_i, \beta_\beta &\sim N(\mathbf{x}_i \beta_\beta, \sigma^2 \mathbf{x}_i \mathbf{x}_i^t), \text{ independientes, } i = 1, 2, \dots, N, \\ \pi(\beta_\beta) &\propto 1, \end{aligned} \quad (4.2)$$

donde  $\mathbf{X} = (\mathbf{x}_1^t, \mathbf{x}_2^t, \dots, \mathbf{x}_N^t)^t$ , es una matriz de rango  $p$  y  $\sigma^2$  un escalar positivo conocido.

Este modelo es un caso particular del modelo Bayes de Regresión normal mínimo informativo definido y estudiado en la subsección 2.2.4.

Como ya avanzábamos, al principio de la sección, para modelizar una población muestreada en dos etapas nos apoyaremos en modelos Bayes en dos etapas y dos jerarquías. Supongamos en una primera aproximación el modelo definido por:

$$\begin{aligned} Y_i | \mathbf{X}, \beta &\sim N_{N_i}(\mathbf{X}_i \beta_i, \mathbf{V}_i), \quad i = 1, 2, \dots, K, \text{ independientes,} \\ \beta_i &\sim N_p(\mathbf{b}_\beta, \mathbf{B}_\beta), \quad i = 1, 2, \dots, K, \text{ i.i.d.,} \end{aligned} \quad (4.3)$$

donde  $\beta = (\beta_1^t, \beta_2^t, \dots, \beta_K^t)^t$ ,  $\mathbf{b}_\beta$  es un  $p$ -vector desconocido y  $\mathbf{V}_i$ ,  $i = 1, 2, \dots, K$  y  $\mathbf{B}_\beta$  son matrices definidas positivas conocidas. Y las siguientes tres hipótesis:

$$\mathbf{b}_\beta = \beta_\beta, \quad (4.4)$$

$$\mathbf{V}_i = \lambda \sigma^2 \text{diag}_{j=1}^{N_i} \{\mathbf{x}_{ij} \mathbf{x}_{ij}^t\}, \quad i = 1, 2, \dots, K, \quad (4.5)$$

$$\mathbf{B}_\beta = (1 - \lambda) \sigma^2 \mathbf{I}_p, \quad (4.6)$$

donde  $\mathbf{I}_p$  es la matriz identidad de dimensión  $p$ ,  $\sigma^2$  es un escalar conocido y  $\lambda$  es un escalar entre 0 y 1 ( $0 < \lambda < 1$ ).

El modelo definido en (4.3) permite trasladar la información proporcionada por un muestra en dos etapas a fin de poder hacer inferencias sobre la población pero adolece del defecto de no representar a la misma población que el modelo definido en (4.1) (no da lugar al mismo modelo de superpoblación). Sin embargo, si al modelo (4.3) le

añadimos las hipótesis (4.4), (4.5) y (4.6) la aproximación del modelo de superpoblación resultante respecto al modelo de superpoblación definido en (4.1) es sustancialmente mejor (no pueden representar el mismo modelo exactamente), ya que las distribuciones marginales para cada variable (condicionadas a  $\beta_\beta$  y  $\lambda$ ) coinciden en los dos modelos de superpoblación, tal y como muestra el siguiente resultado:

**Teorema 4.1** *El modelo de superpoblación condicionado a  $\beta_\beta$  y  $\lambda$  para el modelo definido por (4.3), (4.4), (4.5) y (4.6), es tal que los vectores  $\mathbf{Y}_i$ ,  $i = 1, 2, \dots, K$ , son independientes y se distribuyen según una normal:*

$$\mathbf{Y}_i | \mathbf{X}, \beta_\beta, \lambda \sim N_{N_i}(\mathbf{X}_i \beta_\beta, \sigma^2 [\lambda \text{diag}_{j=1}^{N_i} \{\mathbf{x}_{ij} \mathbf{x}_{ij}^t\} + (1 - \lambda) \mathbf{X}_i \mathbf{X}_i^t]), \quad (4.7)$$

con media y varianza marginales, para  $j = 1, 2, \dots, N_i$ ,  $i = 1, 2, \dots, K$ :

$$\begin{aligned} E(Y_{ij} | \mathbf{X}, \beta_\beta, \lambda) &= \mathbf{x}_{ij} \beta_\beta, \\ \text{Var}(Y_{ij} | \mathbf{X}, \beta_\beta, \lambda) &= \mathbf{x}_{ij} \mathbf{x}_{ij}^t \sigma^2 \end{aligned} \quad (4.8)$$

y covarianzas, para  $j, j^* = 1, 2, \dots, N_i$ ,  $i, i^* = 1, 2, \dots, K$ ,

$$\text{Cov}(Y_{ij}, Y_{i^*j^*} | \mathbf{X}, \beta_\beta, \lambda) = \begin{cases} (1 - \lambda) \mathbf{x}_{ij} \mathbf{x}_{i^*j^*}^t \sigma^2 & i = i^* \\ 0 & i \neq i^*, \end{cases} \quad (4.9)$$

**Demostración.** La demostración de este teorema es inmediata a partir del resultado (a) del lema B.8.

C.S.Q.D.

El resultado anterior permite también un análisis sobre el significado de  $\lambda$ , ya que desde (4.8) y (4.9), es inmediato que, para  $j, j^* = 1, 2, \dots, N_i$ ,  $i = 1, 2, \dots, K$ :

$$\text{Cor}(Y_{ij}, Y_{i^*j^*} | \mathbf{X}, \beta_\beta, \lambda) = (1 - \lambda) R_i^{jj^*}, \quad (4.10)$$

donde  $R_i^{jj^*} = \frac{\mathbf{x}_{ij} \mathbf{x}_{i^*j^*}^t}{\sqrt{\mathbf{x}_{ij} \mathbf{x}_{ij}^t} \sqrt{\mathbf{x}_{i^*j^*} \mathbf{x}_{i^*j^*}^t}}$  proporciona una medida de la correlación entre las covariables de los elementos  $(i, j)$  e  $(i, j^*)$  de una misma unidad  $i$ ,  $j, j^* = 1, 2, \dots, N_i$ ,  $i = 1, 2, \dots, K$ . Según (4.10),  $\lambda$  es un factor que mide el cociente entre la correlación dentro de cada unidad respecto a la característica en estudio y la correlación dentro de cada unidad de las covariables. En una estratificación "ideal" los vectores de covariables de los elementos de una misma unidad (estrato) estarán próximos y  $R_i^{jj^*}$  tomará un valor próximo a 1 para todo  $j, j^* = 1, 2, \dots, N_i$  y, por lo tanto un  $\lambda$  próximo a 0 supondrá que los elementos de la unidad son muy homogéneos entre sí (que es la situación habitual en un muestreo estratificado), cuanto más próximo a

0, más homogéneos y un  $\lambda$  grande (más próximo a 1) que los elementos de la unidad son heterogéneos entre sí. En el caso sin covariables, la interpretación en términos del factor  $\lambda$  es mucho más sencilla ya que entonces ( $\mathbf{X} = \mathbf{1}_N$ ), la correlación dentro de cada unidad será igual a  $(1 - \lambda)$ , y de acuerdo con las características que asumíamos al principio del capítulo acerca de los muestreos estratificados  $\lambda$  será pequeño y en el caso de muestreo cluster el valor de  $\lambda$  aún será habitualmente más pequeño.

El factor  $\lambda$  también puede interpretarse a través de (4.3) y (4.6) como un medidor de la varianza introducida en el modelo de superpoblación (4.7) a través de las diferencias entre las unidades, un  $\lambda$  próximo a 0 señala entonces que la variabilidad proviene fundamentalmente de las diferencias entre unidades y un  $\lambda$  próximo a 1 que las unidades son entre sí muy semejantes.

Atendiendo a estas interpretaciones y en el caso de ausencia de covariables podría ser más adecuado para describir la situación de muestreo por cluster el asumir el modelo resultante de sustituir la hipótesis (4.6) por  $\mathbf{B}_\beta = (1 - \lambda_i)\sigma^2\mathbf{I}_p$ ,  $i = 1, 2, \dots, K$  (y consecuentemente, la frase i.i.d por independientes en el modelo definido en (4.3)), con  $\lambda_i$  una variable aleatoria con  $E(\lambda_i) = \lambda$ ,  $i = 1, 2, \dots, K$ , que permitiría modelizar situaciones con distintos grados de homogeneidad de los elementos dentro de un cluster y de heterogeneidad entre clusters y seguiría siendo comparable con el modelo de superpoblación (4.1) en los mismos términos que introduce el teorema 4.1. Otras modelizaciones, que dejamos, junto con la anterior, para posteriores estudios podrían consistir en suponer  $\lambda$  desconocido.

Nuestros estudios en esta tesis se centrarán en el análisis de los procedimientos de muestreo estratificado y cluster a partir del siguiente modelo.

**Definición 4.2** *Dado el vector poblacional  $\mathbf{Y} = (\mathbf{Y}_1^t, \mathbf{Y}_2^t, \dots, \mathbf{Y}_K^t)^t$  con  $\mathbf{Y}_i = (Y_{i1}, Y_{i2}, \dots, Y_{iN_i})^t$ ,  $i = 1, 2, \dots, K$  para el cual asociada a cada variable  $Y_{ij}$  tenemos una información auxiliar representada por un vector  $\mathbf{x}_{ij} = (x_{ij1}, \dots, x_{ijp})$  conocido,  $j = 1, 2, \dots, N_i$ ,  $i = 1, 2, \dots, K$ , llamaremos modelo Bayes de Regresión para muestreos en dos etapas con varianza conocida al definido por el siguiente modelo jerárquico:*

$$\begin{aligned} \mathbf{Y}_i | \mathbf{X}, \beta &\sim N_{N_i}(\mathbf{X}_i \beta_i, \sigma^2 \lambda \text{diag}_{j=1}^{N_i} \{\mathbf{x}_{ij} \mathbf{x}_{ij}^t\}), \text{ independientes, } i = 1, 2, \dots, K, \\ \beta_i | \beta_\beta &\sim N_p(\beta_\beta, \sigma^2(1 - \lambda)\mathbf{I}_p), \text{ i.i.d., } i = 1, 2, \dots, K, \\ \pi(\beta_\beta) &\propto 1, \end{aligned} \quad (4.11)$$

donde  $\mathbf{X}_i = (\mathbf{x}_{i1}^t, \mathbf{x}_{i2}^t, \dots, \mathbf{x}_{iN_i}^t)^t$  es una matriz de rango  $p$ ,  $i = 1, 2, \dots, K$   $\beta = (\beta_1^t, \beta_2^t, \dots, \beta_K^t)^t$ ,  $\sigma^2$  es un escalar positivo conocido,  $\lambda$  es un escalar entre 0 y 1 conocido ( $0 < \lambda < 1$ ) y  $\mathbf{I}_p$  es la matriz identidad de dimensión  $p$ .

Este modelo es un caso particular del modelo Bayes de regresión en dos etapas y dos jerarquías normal mínimo informativo definido en la subsección 3.2.4, y como caso

particular del mismo tenemos la situación para muestreos en dos etapas cuando no tenemos covariables tomando  $p = 1$ ,  $\mathbf{X}_i = \mathbf{1}_{N_i}$ ,  $i = 1, 2, \dots, K$

El análisis Bayes de este modelo y su comparación analítica con el modelo Bayes de Regresión para muestras aleatorias con varianza conocida se realizará en la sección 4.3, presentándose un ejemplo numérico de esta comparación en la sección 4.4.

#### 4.2.2 Análisis Bayes del modelo para muestras aleatorias

Para cerrar esta sección, el siguiente teorema obtiene las distribuciones predictivas del vector poblacional no observado y la media poblacional no observada, así como el estimador Bayes (para función de pérdida cuadrática) de la media poblacional y su varianza condicionada a la muestra.

**Teorema 4.2** Consideremos el modelo Bayes de Regresión para muestras aleatorias con varianza conocida definido en (4.2), entonces se tiene que:

(a)  $\mathbf{Y}_u | \mathbf{X}, (s^1, \mathbf{y}_{s^1})$  se distribuye según una normal  $(N - n)$ -variante con media y varianza dadas respectivamente por:

$$\begin{aligned} E(\mathbf{Y}_u | \mathbf{X}, (s^1, \mathbf{y}_{s^1})) &= \mathbf{X}_u \hat{\boldsymbol{\beta}}, \\ \text{Var}(\mathbf{Y}_u | \mathbf{X}, (s^1, \mathbf{y}_{s^1})) &= \sigma^2 \{ \text{diag}_{i \notin s^1} \{ \mathbf{x}_i \mathbf{x}_i^t \} + \mathbf{X}_u \mathbf{S}^{-1} \mathbf{X}_u^t \}, \end{aligned} \quad (4.12)$$

(b)  $\bar{Y}_u | \mathbf{X}, (s^1, \mathbf{y}_{s^1})$  se distribuye normal con media y varianza dadas por:

$$\begin{aligned} E(\bar{Y}_u | \mathbf{X}, (s^1, \mathbf{y}_{s^1})) &= \frac{1}{N - n} \sum_{i \notin s^1} \mathbf{x}_i \hat{\boldsymbol{\beta}}, \\ \text{Var}(\bar{Y}_u | \mathbf{X}, (s^1, \mathbf{y}_{s^1})) &= \frac{\sigma^2}{(N - n)^2} \left\{ \sum_{i \notin s^1} \mathbf{x}_i \mathbf{x}_i^t + \left( \sum_{i \notin s^1} \mathbf{x}_i \right) \mathbf{S}^{-1} \left( \sum_{i \notin s^1} \mathbf{x}_i^t \right) \right\}, \end{aligned} \quad (4.13)$$

y (c) el estimador Bayes para la media poblacional y su varianza condicionada a la muestra vienen dados por:

$$\begin{aligned} \hat{Y}_1 &= \frac{n}{N} \bar{y}_{s^1} + \frac{1}{N} \sum_{i \notin s^1} \mathbf{x}_i \hat{\boldsymbol{\beta}}, \\ \text{Var}(\bar{Y} | \mathbf{X}, (s^1, \mathbf{y}_{s^1})) &= \frac{\sigma^2}{N^2} \left\{ \sum_{i \notin s^1} \mathbf{x}_i \mathbf{x}_i^t + \left( \sum_{i \notin s^1} \mathbf{x}_i \right) \mathbf{S}^{-1} \left( \sum_{i \notin s^1} \mathbf{x}_i^t \right) \right\}, \end{aligned} \quad (4.14)$$

donde  $\bar{y}_{s^1}$  es la media muestral,  $\mathbf{S} = \sum_{i \in s^1} \mathbf{x}_i^t (\mathbf{x}_i \mathbf{x}_i^t)^{-1} \mathbf{x}_i$  y

$$\hat{\beta} = E(\beta_\rho | \mathbf{X}, (s^1, \mathbf{y}_{s^1})) = S^{-1} \sum_{i \in s^1} \mathbf{x}_i^t (\mathbf{x}_i \mathbf{x}_i^t)^{-1} y_i, \quad (4.15)$$

es el estimador de regresión clásico.

**Demostración.** La demostración de este teorema es inmediata a partir del corolario 2.1 tomando  $\mathbf{V} = \sigma^2 \text{diag}_{i=1}^N \{\mathbf{x}_i \mathbf{x}_i^t\}$ .

C.S.Q.D.

### 4.3 Análisis Bayes del modelo para muestreos en dos etapas y comparación entre los tres procedimientos de muestreo

Dividiremos esta sección para su estudio en dos partes muy diferenciadas, la primera destinada al análisis Bayes del modelo Bayes de Regresión para muestreos en dos etapas con varianza conocida (definición 4.2) y la segunda a la comparación de las inferencias obtenidas para la media poblacional para los tres métodos de muestreo propuestos, estas inferencias se apoyarán en el modelo Bayes para muestras aleatorias con varianza conocida y en los casos particulares del modelo Bayes de Regresión para muestreos en dos etapas con varianza conocida para los casos de muestreo estratificado y cluster.

#### 4.3.1 Análisis Bayes del modelo para muestreos en dos etapas

Las distribuciones predictivas del vector poblacional no observado y la media no observada, así como el estimador Bayes para la media poblacional y su varianza condicionada a la muestra para el modelo Bayes de regresión con varianza conocida vienen dadas por el siguiente teorema:

**Teorema 4.3** *Consideremos el modelo Bayes de Regresión para muestreos en dos etapas con varianzas conocidas dado en (4.11), entonces se tiene que:*

(a)  $\mathbf{Y}_u | \mathbf{X}, (s, \mathbf{y}_s)$  se distribuye normal  $(N - n)$ -variante con media,

$$E(\mathbf{Y}_u | \mathbf{X}, (s, \mathbf{y}_s)) = \begin{cases} \mathbf{X}_{u_i} [\alpha_i \hat{\beta}_i + (\mathbf{I}_p - \alpha_i) \hat{\beta}_w], & i \in \{i_1, \dots, i_k\}, \\ \mathbf{X}_{u_i} \hat{\beta}_w & i \notin \{i_1, i_2, \dots, i_k\}, \end{cases} \quad (4.16)$$

y matriz de varianzas y covarianzas, dada por:

\* para  $i \in \{i_1, \dots, i_k\}$

$$\begin{aligned} \text{Var}(\mathbf{Y}_{u_i} | \mathbf{X}, (s, \mathbf{y}_s)) &= \sigma^2 \lambda \{ \mathbf{X}_{u_i} \boldsymbol{\alpha}_i \mathbf{S}_i^{-1} \mathbf{X}_{u_i}^t + \text{diag}_{j \notin s_i} \{ \mathbf{x}_{ij} \mathbf{x}_{ij}^t \} \} + \\ &+ \sigma^2 \frac{\lambda^2}{1-\lambda} \mathbf{X}_{u_i} \boldsymbol{\alpha}_i \mathbf{S}_i^{-1} \left[ \sum_{i \in \{i_1, \dots, i_k\}} \boldsymbol{\alpha}_i \right]^{-1} \mathbf{S}_i^{-1} \boldsymbol{\alpha}_i \mathbf{X}_{u_i}^t, \end{aligned} \quad (4.17)$$

\* para  $i \notin \{i_1, \dots, i_k\}$ ,

$$\begin{aligned} \text{Var}(\mathbf{Y}_{u_i} | \mathbf{X}, (s, \mathbf{y}_s)) &= \sigma^2 \lambda \left\{ \frac{1-\lambda}{\lambda} \mathbf{X}_{u_i} \mathbf{X}_{u_i}^t + \text{diag}_{j=1}^{N_i} \{ \mathbf{x}_{ij} \mathbf{x}_{ij}^t \} + \right. \\ &\left. + \frac{1-\lambda}{\lambda} \mathbf{X}_{u_i} \left[ \sum_{i \in \{i_1, \dots, i_k\}} \boldsymbol{\alpha}_i \right]^{-1} \mathbf{X}_{u_i}^t \right\} \end{aligned} \quad (4.18)$$

\* para  $i \in \{i_1, \dots, i_k\}$ ,  $j \notin \{i_1, \dots, i_k\}$ ,

$$\text{Cov}(\mathbf{Y}_{u_i}, \mathbf{Y}_{u_j} | \mathbf{X}, (s, \mathbf{y}_s)) = \sigma^2 \lambda \mathbf{X}_{u_i} \boldsymbol{\alpha}_i \mathbf{S}_i^{-1} \left[ \sum_{i \in \{i_1, \dots, i_k\}} \boldsymbol{\alpha}_i \right]^{-1} \mathbf{X}_{u_j}^t, \quad (4.19)$$

\* para  $i \neq j$ ,  $i, j \in \{i_1, \dots, i_k\}$ ,

$$\text{Cov}(\mathbf{Y}_{u_i}, \mathbf{Y}_{u_j} | \mathbf{X}, (s, \mathbf{y}_s)) = \sigma^2 \frac{\lambda^2}{1-\lambda} \mathbf{X}_{u_i} \boldsymbol{\alpha}_i \mathbf{S}_i^{-1} \left[ \sum_{i \in \{i_1, \dots, i_k\}} \boldsymbol{\alpha}_i \right]^{-1} \mathbf{S}_j^{-1} \boldsymbol{\alpha}_j \mathbf{X}_{u_j}^t, \quad (4.20)$$

\* para  $i \neq j$ ,  $i, j \notin \{i_1, \dots, i_k\}$ ,

$$\text{Cov}(\mathbf{Y}_{u_i}, \mathbf{Y}_{u_j} | \mathbf{X}, (s, \mathbf{y}_s)) = \sigma^2 (1-\lambda) \mathbf{X}_{u_i} \left[ \sum_{i \in \{i_1, \dots, i_k\}} \boldsymbol{\alpha}_i \right]^{-1} \mathbf{X}_{u_j}^t, \quad (4.21)$$

(b)  $\bar{\mathbf{Y}}_u | \mathbf{X}, (s, \mathbf{y}_s)$  se distribuye normal con media y varianza dadas por:

$$\begin{aligned} E(\bar{\mathbf{Y}}_u | \mathbf{X}, (s, \mathbf{y}_s)) &= \frac{1}{N-n} \left\{ \sum_{i \in \{i_1, \dots, i_k\}} \mathbf{1}_{N_i-n_i}^t \mathbf{X}_{u_i} [\boldsymbol{\alpha}_i \hat{\boldsymbol{\beta}}_i + (\mathbf{I}_p - \boldsymbol{\alpha}_i) \hat{\boldsymbol{\beta}}_w] + \right. \\ &\left. + \sum_{i \notin \{i_1, \dots, i_k\}} \mathbf{1}_{N_i}^t \hat{\boldsymbol{\beta}}_w \right\}, \end{aligned} \quad (4.22)$$

$$\text{Var}(\bar{\mathbf{Y}}_u | \mathbf{X}, (s, \mathbf{y}_s)) = \frac{\sigma^2}{(N-n)^2} \{ \lambda v_1 + (1-\lambda) v_2 \},$$

y (c) el estimador Bayes para la media poblacional y su varianza condicionada a la muestra vienen dados por:

$$\begin{aligned}\widehat{Y} &= \frac{n}{N}\bar{y}_s + \frac{N-n}{N}E(\bar{Y}_u|\mathbf{X},(s,\mathbf{y}_s)), \\ \text{Var}(\bar{Y}|\mathbf{X},(s,\mathbf{y}_s)) &= \frac{\sigma^2}{N^2}\{\lambda v_1 + (1-\lambda)v_2\},\end{aligned}\quad (4.23)$$

donde<sup>2</sup>  $\bar{y}_s = \frac{\sum_{(i,j)\in s} y_{ij}}{n}$  es la media muestral,  $S_i = \sum_{j\in s_i} \mathbf{x}_{ij}^t (\mathbf{x}_{ij} \mathbf{x}_{ij}^t)^{-1} \mathbf{x}_{ij}$  y

$$\begin{aligned}\alpha_i &= [(1-\lambda)S_i + \lambda\mathbf{I}_p]^{-1}(1-\lambda)S_i, \quad i \in \{i_1, \dots, i_k\}, \\ \hat{\beta}_i &= S_i^{-1} \sum_{j\in s_i} \mathbf{x}_{ij}^t (\mathbf{x}_{ij} \mathbf{x}_{ij}^t)^{-1} y_{ij}, \quad i \in \{i_1, \dots, i_k\}, \\ \hat{\beta}_w &= \left[ \sum_{i\in\{i_1, \dots, i_k\}} \alpha_i \right]^{-1} \sum_{i\in\{i_1, \dots, i_k\}} \alpha_i \hat{\beta}_i, \\ \mathbf{Z}_w &= \sum_{i\in\{i_1, \dots, i_k\}} \mathbf{1}_{N_i-n_i}^t \mathbf{X}_{u_i} (\mathbf{I}_p - \alpha_i) + \sum_{i\notin\{i_1, \dots, i_k\}} \mathbf{1}_{N_i}^t \mathbf{X}_{u_i}, \\ v_1 &= \sum_{i\in\{i_1, \dots, i_k\}} \sum_{j\notin s_i} \mathbf{x}_{ij} \mathbf{x}_{ij}^t + \sum_{i\notin\{i_1, \dots, i_k\}} \sum_{j=1}^{N_i} \mathbf{x}_{ij} \mathbf{x}_{ij}^t + \\ &\quad + \sum_{i\in\{i_1, \dots, i_k\}} \mathbf{1}_{N_i-n_i}^t \mathbf{X}_{u_i} \alpha_i S_i^{-1} \mathbf{X}_{u_i}^t \mathbf{1}_{N_i-n_i} + \\ &\quad + \sum_{i\in\{i_1, \dots, i_k\}} \mathbf{1}_{N_i-n_i}^t \mathbf{X}_{u_i} \alpha_i S_i^{-1} \left[ \sum_{i\in\{i_1, \dots, i_k\}} \alpha_i \right]^{-1} \mathbf{Z}_w^t, \\ v_2 &= \sum_{i\notin\{i_1, \dots, i_k\}} \mathbf{1}_{N_i}^t \mathbf{X}_{u_i} \mathbf{X}_{u_i}^t \mathbf{1}_{N_i} + \\ &\quad + \sum_{i\notin\{i_1, \dots, i_k\}} \mathbf{1}_{N_i}^t \mathbf{X}_{u_i} \left[ \sum_{i\in\{i_1, \dots, i_k\}} \alpha_i \right]^{-1} \mathbf{Z}_w^t.\end{aligned}\quad (4.24)$$

**Demostración.** La demostración de este teorema es inmediata a partir del corolario 3.1 tomando  $\mathbf{V}_i = \sigma^2 \lambda \text{diag}_{j=1}^{N_i} \{\mathbf{x}_{ij} \mathbf{x}_{ij}^t\}$ ,  $i = 1, 2, \dots, K$  y  $\mathbf{B}_\beta = \sigma^2(1-\lambda)\mathbf{I}_p$ .

C.S.Q.D.

La media de la distribución predictiva de  $\mathbf{Y}_{u_i}|\mathbf{X},(s,\mathbf{y}_s)$ ,  $i = 1, 2, \dots, K$ , dada en (4.16) puede ser interpretada en términos de  $\lambda$ . Notemos en primer lugar que, lógicamente

<sup>2</sup>Notemos que el factor  $\alpha_i$ ,  $i \in \{i_1, i_2, \dots, i_k\}$ , es en el caso que nos ocupa una matriz de dimensión  $p$  simétrica

en el caso de unidades no muestreadas, la media es la estimación lineal con pendiente de regresión la resultante de una ponderación de las pendientes de regresión clásicas para cada unidad muestreada<sup>3</sup>,  $\hat{\beta}_w$ , y en el caso de unidades muestreadas, la media es la estimación lineal con pendiente de regresión una “combinación lineal” del estimador de regresión clásico,  $\hat{\beta}_i$ ,  $i \in \{i_1, \dots, i_k\}$ , y de  $\hat{\beta}_w$ , a través de  $\alpha_i$ ,  $i \in \{i_1, \dots, i_k\}$ . El factor de encogimiento (shrinking factor),  $\alpha_i$ ,  $i \in \{i_1, \dots, i_k\}$ , que determina la combinación lineal anterior puede interpretarse como una función decreciente de  $\lambda$  de  $\mathbf{I}_p$  (para  $\lambda \rightarrow 0$ ) a  $\mathbf{0}_p$  (para  $\lambda \rightarrow 1$ ), de modo que, para  $\lambda$  pequeño (si la correlación entre elementos de una unidad es grande) la pendiente Bayes de regresión se aproxima a la pendiente de regresión clásica para cada unidad muestreada  $i$ ,  $i \in \{i_1, \dots, i_k\}$ , y a medida que  $\lambda$  crece (se aproxima a 1), el estimador Bayes de la pendiente de regresión se aproxima a  $\hat{\beta}_w$ . Esta interpretación justifica, desde un punto de vista Bayesiano, el uso del estimador de regresión clásico en las situaciones habituales de muestreo estratificado y cluster, puesto que, según comentábamos en la subsección anterior,  $\lambda$  mide la relación entre los elementos de una misma unidad (indicando un  $\lambda$  pequeño, unidades homogéneas) y habitualmente los estratos se construyen intentando que los elementos dentro de cada estrato sean homogéneos entre sí y la propiedad de homogeneidad dentro de cada unidad suele darse también de manera frecuente en los cluster (aunque en este último caso estamos ante una propiedad dada y no construida). También podemos interpretar  $\alpha_i$ ,  $i \in \{i_1, \dots, i_k\}$  a partir de las siguientes dos expresiones<sup>4</sup>:

$$\begin{aligned}\alpha_i &= \left[ \text{Var}(\beta_i | \beta_\beta) + \text{Var}(\hat{\beta}_i | \mathbf{X}, \beta_i) \right]^{-1} \text{Var}(\beta_i | \beta_\beta) \\ &= \left[ \text{Var}(\beta_i | \beta_\beta) - \text{Var}(\beta_i | \mathbf{X}, (s, \mathbf{y}_s)) \right] \text{Var}^{-1}(\beta_i | \beta_\beta)\end{aligned}\quad (4.25)$$

de la primera se sigue que  $\alpha_i$ , puede interpretarse (en términos del modelo) como la contribución relativa de la varianza de  $\beta_i$  en el total de la variación de  $\hat{\beta}_i$  alrededor de  $\beta_\beta$ , y de la segunda (en términos de distribuciones finales), como la proporción de varianza respecto a  $\beta_i$  no explicada por la muestra.

Las expresiones del estimador Bayes de la media poblacional y de su varianza condicionada a la muestra dadas en (4.23) pueden simplificarse en los casos particulares de muestreos estratificados (con  $k = K$ ) y cluster (con  $n_i = N_i$ ,  $i \in \{i_1, \dots, i_k\}$ ), tal y como se comprueba en el siguiente corolario:

**Corolario 4.1** *El estimador Bayes para la media poblacional y su varianza condicionada a la muestra, dados en (4.23) y obtenidos a partir del modelo Bayes de Regresión para muestras en dos etapas con varianza conocida, tienen la siguiente forma*

<sup>3</sup> $\hat{\beta}_w$  es una ponderación de las pendientes de regresión clásicas de las unidades observadas que en general no coincide con el estimador de regresión clásico para la muestra en su conjunto (descrito en (4.15) para una muestra aleatoria). Este estimador se obtiene en el caso particular  $\mathbf{S}_{i_1} = \mathbf{S}_{i_2} = \dots = \mathbf{S}_{i_k}$ .

<sup>4</sup>La primera expresión se obtiene a partir de la definición de  $\alpha_i$ , notando que  $\text{Var}(\hat{\beta}_i | \mathbf{X}, \beta_i) = \lambda \mathbf{S}_i^{-1}$ , y la segunda se apoya en el estudio del modelo Bayes de regresión en dos etapas con dos jerarquías normal realizado en la sección 3.2.

particular en los casos de muestreos estratificado con  $k = K$  y cluster con  $n_i = N_i$ ,  $i \in \{i_1, \dots, i_k\}$ . Así:

(a) el estimador Bayes de la media poblacional y su varianza condicionada a la muestra para un muestreo estratificado, vienen dados por:

$$\begin{aligned}\widehat{Y}_2 &= \frac{n}{N} \bar{y}_{s^2} + \frac{1}{N} \sum_{i=1}^K \mathbf{1}_{N_i-n_i}^t \mathbf{X}_{u_i} [\alpha_i \widehat{\beta}_i + (\mathbf{I}_p - \alpha_i) \widehat{\beta}_w], \\ \text{Var}(\bar{Y} | \mathbf{X}, (s^2, \mathbf{y}_{s^2})) &= \frac{\sigma^2}{N^2} \left\{ \lambda \sum_{i=1}^K \sum_{j \in s_i^2} \mathbf{x}_{ij} \mathbf{x}_{ij}^t + (1 - \lambda) v_e \right\},\end{aligned}\quad (4.26)$$

donde  $\alpha_i$ ,  $i = 1, 2, \dots, K$ ,  $\widehat{\beta}_i$ ,  $i = 1, 2, \dots, K$  y  $\widehat{\beta}_w$  son las expresiones definidas en (4.24) y,

$$v_e = \sum_{i=1}^K \mathbf{1}_{N_i-n_i}^t \mathbf{X}_{u_i} (\mathbf{I}_p - \alpha_i) \left[ \mathbf{X}_{u_i}^t + \left( \sum_{i=1}^K \alpha_i \right)^{-1} \sum_{i=1}^K (\mathbf{I}_p - \alpha_i) \mathbf{X}_{u_i}^t \mathbf{1}_{N_i-n_i} \right], \quad (4.27)$$

y (b) el estimador Bayes de la media poblacional y su varianza condicionada a la muestra para un muestreo cluster, vienen dados por:

$$\begin{aligned}\widehat{Y}_3 &= \frac{n}{N} \bar{y}_{s^3} + \frac{1}{N} \sum_{i \in \{i_1, \dots, i_k\}} \mathbf{1}_{N_i}^t \mathbf{X}_{u_i} \widehat{\beta}_w, \\ \text{Var}(\bar{Y} | \mathbf{X}, (s^3, \mathbf{y}_{s^3})) &= \frac{\sigma^2}{N^2} \left\{ \lambda \sum_{i \in \{i_1, \dots, i_k\}} \sum_{j=1}^{N_i} \mathbf{x}_{ij} \mathbf{x}_{ij}^t + (1 - \lambda) v_c \right\},\end{aligned}\quad (4.28)$$

donde  $\alpha_i$ ,  $i \in \{i_1, \dots, i_k\}$  y  $\widehat{\beta}_w$  son las expresiones definidas en (4.24) y,

$$v_c = \sum_{i \in \{i_1, \dots, i_k\}} \mathbf{1}_{N_i}^t \mathbf{X}_{u_i} \mathbf{X}_{u_i}^t \mathbf{1}_{N_i} + \mathbf{1}_{N-n}^t \mathbf{X}_u \left( \sum_{i \in \{i_1, \dots, i_k\}} \alpha_i \right)^{-1} \mathbf{X}_u^t \mathbf{1}_{N-n} \quad (4.29)$$

**Demostración.** La demostración de este teorema es inmediata a partir del teorema anterior aplicando que en los muestreos estratificados  $k = K$  y en los muestreos cluster  $n_i = N_i$ ,  $i \in \{i_1, \dots, i_k\}$ .

C.S.Q.D.

### 4.3.2 Comparación entre los tres tipos de muestreo

Dedicaremos esta subsección a la comparación entre los tres procedimientos de muestreo propuestos a través de las expresiones para los estimadores Bayes y sus varianzas condicionadas a la muestra para muestreos estratificados y cluster, dadas en el corolario 4.1, y las expresiones correspondientes para muestreos aleatorios dadas en (4.14).

Para una mejor reflexión sobre los resultados distinguiremos entre dos situaciones: el caso sin covariables y con covariables. En ambos casos, para facilitar la interpretación de los resultados hemos partido de una situación en la que los estimadores Bayes para los tres procedimientos coinciden,  $\widehat{Y} = \frac{n}{N}\bar{y}_s + \frac{1}{N}\sum_{i \in g_s} \mathbf{x}_i \widehat{\beta}$ , con  $\widehat{\beta}$ , el estimador de regresión clásico, y por lo tanto la comparación entre los mecanismos de muestreo puede realizarse en términos de la varianza de la media poblacional dada una muestra aleatoria, estratificada y cluster.

#### Caso sin covariables

El artículo de Bayarri y Font(94)[6] estudia la comparación entre los tres procedimientos de muestreo en el caso sin covariables a partir de los modelos definidos en las definiciones 4.1 (tomando  $p = 1$ ,  $\mathbf{X} = \mathbf{1}_N$ ) y 4.2 (tomando  $p = 1$ ,  $\mathbf{X}_i = \mathbf{1}_{N_i}$ ,  $i = 1, 2, \dots, K$ ). El teorema que describiremos a continuación introduce unas hipótesis para las cuales los estimadores Bayes de la media poblacional para los tres procedimientos de muestreo coinciden y obtiene (bajo esas hipótesis) las expresiones de las varianzas de la media poblacional condicionada a muestras estratificadas y cluster en términos de la varianza de la media poblacional condicionada a una muestra aleatoria.

**Teorema 4.4** *Consideremos el modelo Bayes de Regresión para muestras aleatorias, con  $p = 1$ ,  $\mathbf{x}_i = 1$ ,  $i = 1, 2, \dots, N$ , y una muestra aleatoria de la población en estudio  $(s^1, \mathbf{y}_{s^1})$  y el modelo Bayes de Regresión para muestras en dos etapas, con  $p = 1$ ,  $\mathbf{x}_{ij} = 1$ ,  $j = 1, 2, \dots, N_i$ ,  $i = 1, 2, \dots, K$ , y una muestra estratificada  $(s^2, \mathbf{y}_{s^2})$  y una muestra cluster  $(s^3, \mathbf{y}_{s^3})$  de la misma población. Entonces se tiene que:*

(a) *si la muestra estratificada cumple que  $k = K$  y las siguientes dos hipótesis:*

$$n_1 = n_2 = \dots = n_K = \frac{n}{K}, \quad (4.30)$$

$$N_1 = N_2 = \dots = N_K = \frac{N}{K}, \quad (4.31)$$

*se tiene que:*

$$\begin{aligned} \widehat{Y}_2 &= \bar{y}_{s^2}, \\ \text{Var}(\bar{Y} | (s^2, \mathbf{y}_{s^2})) &= \lambda \text{Var}(\bar{Y} | (s^1, \mathbf{y}_{s^1})), \end{aligned} \quad (4.32)$$

*y (b) si la muestra cluster cumple que  $n_i = N_i$ ,  $i \in \{i_1, i_2, \dots, i_k\}$  y la siguiente hipótesis:*

$$N_1 = N_2 = \dots = N_K = \frac{N}{K}, \quad (4.33)$$

se tiene que:

$$\begin{aligned} \widehat{Y}_3 &= \bar{y}_{s^3}, \\ \text{Var}(\overline{Y}|(s^3, \mathbf{y}_{s^3})) &= [\lambda + \frac{N}{K}(1 - \lambda)]\text{Var}(\overline{Y}|(s^1, \mathbf{y}_{s^1})), \end{aligned} \quad (4.34)$$

donde  $\bar{y}_s$  es la media muestral,  $\text{Var}(\overline{Y}|(s^2, \mathbf{y}_{s^2}))$  y  $\text{Var}(\overline{Y}|(s^3, \mathbf{y}_{s^3}))$  son las varianzas de la distribución predictiva dada la muestra para el modelo de Regresión (con  $p = 1$ ,  $\mathbf{x}_{ij} = 1$ ,  $j = 1, \dots, N_i$ ,  $i = 1, \dots, K$ ) y  $\text{Var}(\overline{Y}|(s^1, \mathbf{y}_{s^1}))$  es la varianza para el modelo de Regresión para muestras aleatorias (con  $p = 1$ ,  $\mathbf{x}_i = 1$ ,  $i = 1, \dots, N$ ).

**Demostración.** La demostración de estos dos resultados se apoya en la realización de operaciones básicas (aplicando las hipótesis descritas anteriormente), sobre las expresiones dadas para los estimadores Bayes de la media poblacional y sus varianzas condicionadas para los muestreos estratificado y cluster en el corolario 4.1 resultados (a) y (b) respectivamente, (tomando  $p = 1$ ,  $\mathbf{x}_{ij} = 1$ ,  $j = 1, \dots, N_i$ ,  $i = 1, \dots, K$ ) y la comparación de estos resultados con los obtenidos para un muestreo aleatorio en el teorema 4.2, (tomando  $\mathbf{x}_i = 1$ ,  $i = 1, \dots, K$ ).

C.S.Q.D.

Las hipótesis (4.30) y (4.31) para el muestreo estratificado y la hipótesis (4.33) para muestreo cluster, que permiten establecer el resultado anterior, son cómodas para establecer la siguiente comparación entre los tres procedimientos (presentada en Bayarri y Font(94)[6]), pero constituyen hipótesis sobre las muestras estratificadas y cluster bastante restrictivas, sobre todo en el caso del muestreo cluster, en el que el tamaño de los cluster es un dato que viene dado y no puede ser controlado por el estadístico.

\* Bajo las hipótesis (4.30) y (4.31), (tomar estratos del mismo tamaño y muestrear en cada uno de ellos el mismo número de elementos), el muestreo estratificado proporciona siempre mejores resultados que el muestreo aleatorio, dado que  $0 < \lambda < 1$ . Además, en los muestreos estratificados, habitualmente, las diferencias entre elementos dentro de cada estrato son pequeñas y por lo tanto  $\lambda$  es pequeño, y la ganancia en precisión del estimador es considerable.

\* Bajo la hipótesis (4.33), (los cluster son del mismo tamaño), el muestreo cluster proporciona peores resultados que el muestreo estratificado y el aleatorio. Efectivamente, en los muestreos cluster las varianzas entre cluster son muy grandes y por lo tanto  $\lambda$  está cerca de cero, con lo que  $\text{Var}(\overline{Y}|(s^3, \mathbf{y}_{s^3}))$  tiende a  $\frac{N}{K}\text{Var}(\overline{Y}|(s^1, \mathbf{y}_{s^1}))$ , con  $\frac{N}{K} > 1$ , si el tamaño de los clusters es mayor que 1. La hipótesis (4.33) puede

debilitarse, manteniendo el resultado de igualdad de estimadores (con lo que la comparación sigue pudiéndose realizar en términos de la varianza condicionada de la media poblacional), tomando  $N_{i_1} = N_{i_2} = \dots = N_{i_k}$ ; en este caso se tiene que:

$$\text{Var}(\bar{Y}|(s^3, \mathbf{y}_{s^3})) = [\lambda + (1 - \lambda)N_w] \text{Var}(\bar{Y}|\mathbf{X}, (s^1, \mathbf{y}_{s^1})), \quad (4.35)$$

con  $N_w = \frac{\sum_{i \notin \{i_1, \dots, i_k\}} N_i^2 + \frac{(N-n)^2}{k}}{\frac{n(N-n)}{k} + \frac{(N-n)^2}{k}}$ . Y el resultado anterior, sobre la precisión de la estimación sigue siendo válido, cuando  $N_i \geq \frac{n}{k}$ ,  $i \notin \{i_1, i_2, \dots, i_k\}$  ( $\frac{n}{k}$  es el tamaño de los estratos muestreados), ya que en este caso, se pueden repetir nuevamente las reflexiones anteriores y  $N_w > 1$ .

### Caso con covariables

En el caso con covariables, la comparación entre los muestreos aleatorios, estratificado y cluster es más complicada y no puede realizarse en términos absolutos (ni siquiera introduciendo hipótesis para que los estimadores Bayes de la media poblacional coincidan), dado que las varianzas de las distribuciones predictivas de la media poblacional dependen de los valores de la covariable de los elementos no muestreados (que no tiene porque coincidir para las tres muestras). El siguiente teorema muestra unas hipótesis bajo las cuales los estimadores de la media poblacional coinciden para los tres tipos de muestreo y compara las varianzas para los muestreos estratificado y cluster con la varianza que se obtendría para estas muestras asumiendo el modelo para muestras aleatorias.

**Teorema 4.5** *Consideremos el modelo Bayes de Regresión para muestras en dos etapas y una muestra estratificada  $(s^2, \mathbf{y}_{s^2})$  y una muestra cluster  $(s^3, \mathbf{y}_{s^3})$  de población en estudio. Entonces se tiene que:*

(a) *si la muestra estratificada cumple que  $k = K$  y las siguientes dos hipótesis:*

$$\mathbf{S}_1 = \mathbf{S}_2 = \dots = \mathbf{S}_K, \quad (4.36)$$

$$\mathbf{1}_{N_1 - n_1}^t \mathbf{X}_{u_1} = \mathbf{1}_{N_2 - n_2}^t \mathbf{X}_{u_2} = \dots = \mathbf{1}_{N_K - n_K}^t \mathbf{X}_{u_K}, \quad (4.37)$$

*se tiene que:*

$$\hat{Y}_2 = \frac{n}{N} \bar{y}_{s^2} + \frac{1}{N} \sum_{(i,j) \notin s^2} \mathbf{x}_{ij} \hat{\beta}, \quad (4.38)$$

$$\text{Var}(\bar{Y}|\mathbf{X}, (s^2, \mathbf{y}_{s^2})) = \lambda \text{Var}_{SRs}(\bar{Y}|\mathbf{X}, (s^2, \mathbf{y}_{s^2})),$$

y (b) *si la muestra cluster cumple que  $n_i = N_i$ ,  $i \in \{i_1, i_2, \dots, i_k\}$  y la siguiente hipótesis:*

$$\mathbf{S}_{i_1} = \mathbf{S}_{i_2} = \cdots = \mathbf{S}_{i_k}, \quad (4.39)$$

se tiene que:

$$\widehat{Y}_3 = \frac{n}{N} \bar{y}_{s^3} + \frac{1}{N} \sum_{(i,j) \in s^3} \mathbf{x}_{ij} \widehat{\beta}, \quad (4.40)$$

$$\text{Var}(\widehat{Y}_3 | \mathbf{X}, (s^3, \mathbf{y}_{s^3})) = [\lambda + (1 - \lambda) \mathbf{R}_w] \text{Var}_{SRS}(\widehat{Y}_3 | \mathbf{X}, (s^3, \mathbf{y}_{s^3})),$$

donde  $\widehat{\beta}$  es el estimador de regresión clásico para la muestra tomada en su globalidad,  $\mathbf{S}_i$ ,  $i \in \{i_1, i_2, \dots, i_k\}$  son las matrices de dimensión  $p$  definidas en (4.24),  $\text{Var}(\widehat{Y}_3 | \mathbf{X}, (s^2, \mathbf{y}_{s^2}))$  y  $\text{Var}(\widehat{Y}_3 | \mathbf{X}, (s^3, \mathbf{y}_{s^3}))$  son las varianzas de la distribución predictiva dadas las dos muestras para el modelo de Regresión,  $\text{Var}_{SRS}(\widehat{Y}_3 | \mathbf{X}, (s^2, \mathbf{y}_{s^2}))$  y  $\text{Var}_{SRS}(\widehat{Y}_3 | \mathbf{X}, (s^3, \mathbf{y}_{s^3}))$  son las varianzas de las distribuciones predictivas dadas las dos muestras asumiendo el modelo de Regresión para muestras aleatorias, y

$$\mathbf{R}_w = \frac{\sigma^2}{N^2} \left\{ \frac{\sum_{i \in \{i_1, \dots, i_k\}} \left( \sum_{j=1}^{N_i} \mathbf{x}_{ij} \right) \left( \sum_{j=1}^{N_i} \mathbf{x}_{ij}^t \right)}{\text{Var}_{SRS}(\widehat{Y}_3 | \mathbf{X}, (s^3, \mathbf{y}_{s^3}))} + \frac{k^{-1} \left( \sum_{i \in \{i_1, \dots, i_k\}} \sum_{j=1}^{N_i} \mathbf{x}_{ij} \right) \left( \sum_{i \in \{i_1, \dots, i_k\}} \sum_{j=1}^{N_i} \mathbf{x}_{ij}^t \right)}{\text{Var}_{SRS}(\widehat{Y}_3 | \mathbf{X}, (s^3, \mathbf{y}_{s^3}))} \right\}. \quad (4.41)$$

**Demstración.** La demostración de estos dos resultados se apoya en la realización de operaciones básicas de matrices (aplicando las hipótesis descritas anteriormente), sobre las expresiones dadas para los estimadores Bayes y sus varianzas condicionadas para los muestreos estratificado y cluster en el corolario 4.1 resultados (a) y (b) respectivamente, y la comparación de estos resultados con los obtenidos asumiendo un muestreo aleatorio y aplicando el estimador Bayes de la media poblacional y su varianza condicionada dados en el teorema 4.2 (expresiones en (4.14)).

C.S.Q.D.

Notemos en primer lugar que las hipótesis (4.36) y (4.37) para la muestra estratificada y (4.39) para la muestra cluster son más restrictivas que las correspondientes hipótesis para el caso sin covariables. Respecto a las hipótesis, señalemos que: (4.36) y (4.39) pueden interpretarse en términos de la varianza de  $\widehat{\beta}_i$ ,  $i \in \{i_1, \dots, i_k\}$ , ( $\text{Var}(\widehat{\beta}_i | \mathbf{X}, \beta_i) = \lambda \mathbf{S}_i^{-1}$ ), y consiste en seleccionar una muestra de modo que la variabilidad de los estimadores de regresión clásicos dentro de cada unidad sea semejante para todas las unidades muestreadas; mientras que, (4.37) hace referencia a que la media de las covariables de elementos no muestreados sea la misma en todas las unidades

muestreadas. Ello no obstante, estas hipótesis pueden darse (al menos aproximadamente) en la práctica, así las hipótesis (4.36) y (4.37) para muestreos estratificados con  $k = K$  se cumplirán, la primera en estratificaciones en las que las variabilidades de los estratos sean similares y la segunda cuando la información “despreciada” en cada estrato (no muestreada) sea similar. Y la hipótesis (4.39) para muestreos cluster con  $n_i = N_i, i \in \{i_1, i_2, \dots, i_k\}$  también puede ser bastante habitual, puesto que los cluster suelen estar formados por elementos muy homogéneos entre sí con variabilidades entre cluster similares.

Respecto a la comparación entre los procedimientos de muestreo aleatorio, estratificado y cluster, para el caso con covariables señalemos que el teorema 4.5 establece que bajo las hipótesis (4.36) y (4.37) para un muestreo estratificado con  $k = K$  y (4.39) para un muestreo cluster con  $n_i = N_i, i \in \{i_1, i_2, \dots, i_k\}$ , los estimadores Bayes de la media poblacional para los tres muestreos coinciden, además se tiene, comparando las varianzas de la media poblacional condicionadas a las tres muestras, que:

\* En un muestreo estratificado “bueno”,  $Var_{SRS}(\bar{Y}|\mathbf{X}, (s^2, \mathbf{y}_{s,2}))$  ha de ser inferior o semejante a  $Var(\bar{Y}|\mathbf{X}, (s^1, \mathbf{y}_{s,1}))$  ya que las diferencias dentro de los estratos son inferiores a las existentes entre los elementos de una población no estratificada y su suma (referida a los elementos no muestreados) tanto en media como en varianza será inferior o al menos del mismo orden que la media y varianza de una población no estratificada y esta reflexión es trasladable a los valores de la covariable, ya que ésta proporciona información acerca de la característica en estudio. Bajo esta hipótesis, se recupera el resultado para el caso de covariables que mostráramos en la subsubsección anterior, según el cual, las inferencias obtenidas por un muestreo estratificado son más precisas que las obtenidas por un muestreo aleatorio.

\* En un muestreo cluster, las varianzas entre clusters suelen ser muy grandes, por lo que  $\lambda$  está cerca de 0 y la varianza tiende a

$$\begin{aligned} \mathbf{R}_w Var_{SRS}(\bar{Y}|\mathbf{X}, (s^3, \mathbf{y}_{s,3})) &= \\ &= \frac{\sigma^2}{N^2} \left[ \sum_{i \notin \{i_1, \dots, i_k\}} \left( \sum_{j=1}^{M_i} \mathbf{x}_{ij} \right) \left( \sum_{j=1}^{M_i} \mathbf{x}_{ij}^t \right) + \right. \\ &\left. + k^{-1} \left( \sum_{i \notin \{i_1, \dots, i_k\}} \sum_{j=1}^{M_i} \mathbf{x}_{ij} \right) \left( \sum_{i \notin \{i_1, \dots, i_k\}} \sum_{j=1}^{M_i} \mathbf{x}_{ij}^t \right) \right]. \end{aligned}$$

Esta cantidad, normalmente, superará a la varianza para un muestreo aleatorio, así, si asumimos que las medias de las covariables de unidades no muestreadas coinciden entonces la expresión anterior se convierte en  $\frac{\sigma^2}{N^2} \left[ \frac{K}{k} \sum_{i \notin \{i_1, \dots, i_k\}} \mathbf{1}_{N_i}^t \mathbf{X}_{u_i} \mathbf{X}_{u_i}^t \mathbf{1}_{N_i} \right]$ , que es un término grande dado que  $K$  es mucho mayor que  $k$  y las diferencias entre clusters suelen ser grandes (y mayores por supuesto que entre elementos no muestreados en un población muestreada aleatoriamente). Nuevamente, en estas situaciones se recupera

el resultado según el cual el muestreo cluster es menos preciso que el muestreo aleatorio (y evidentemente mucho menos que el estratificado).

## 4.4 Ejemplo: comparación numérica de los muestreos aleatorio, estratificado y cluster

En la práctica, las características de la población son las que hacen más adecuado o condicionan el mecanismo de muestreo a aplicar, por lo que no es corriente que una población sea muestreada por los tres procedimientos estudiados en este capítulo. En esta sección hemos recurrido al artificio, para comparar los tres procedimientos numericamente, de simular varias poblaciones y obtener para cada una de ellas tres muestras correspondientes a los tres mecanismos de muestreo estudiados. Concretamente, hemos simulado las poblaciones sin covariables: Pob2Aa, Pob2Ab, Pob2Ba y Pob2Bb, extrayendo para cada una de ellas una muestra aleatoria, una muestra estratificada y una muestra cluster, y finalmente hemos obtenido los estimadores Bayes de la media poblacional y su varianza condicionada a la muestra, usando las expresiones de las mismas proporcionadas en el capítulo, para cada población y cada muestra.

Las características de las poblaciones simuladas, el mecanismo aplicado para la obtención de la muestra y los resultados obtenidos aparecen comentados en las siguientes subsecciones.

### 4.4.1 Descripción de las poblaciones simuladas y obtención de las muestras

Las cuatro poblaciones simuladas: Pob2Aa, Pob2Ab, Pob2Ba y Pob2Bb, responden al mismo esquema de simulación que denominaremos Población 2.

#### Población 2

Para la simulación de la población 2 se aplicó el siguiente procedimiento, que dio lugar dependiendo del valor de  $K$  y  $\lambda$ , a las poblaciones: Pob2Aa, Pob2Ab, Pob2Ba y Pob2Bb, de acuerdo con el esquema presentado en la tabla 4.1.

1. Obtención de una muestra de tamaño  $K$  de una Normal con media 80 y varianza  $15^2(1 - \lambda)$  ( $\rightsquigarrow \{\beta_i\}_{i=1}^K$ ).
2. Obtención para  $i = 1, 2, \dots, K$ , de una muestra de tamaño  $\frac{10000}{K}$  de una Normal con media  $\beta_i$  y varianza  $15^2\lambda$  ( $\rightsquigarrow \{Y_i\}_{i=1}^K$ ).

Notemos que las poblaciones simuladas según este esquema cumplen el modelo Bayes de Regresión para muestreos en dos etapas con varianzas conocida sin covariables definido en la sección 4.2 (con  $p = 1$ ,  $\mathbf{X}_i = \mathbf{1}_{N_i}$ ,  $i = 1, 2, \dots, K$ ), y además cumplen que el tamaño de todas las unidades es el mismo (hipótesis (4.31) y (4.33) del teorema

	$\lambda = 0.3$	$\lambda = 0.7$
$K = 2^k$	Pob2Aa	Pob2Ab
$K = 400$	Pob2Ba	Pob2Bb

Tabla 4.1: Esquema sobre la simulación de la Población 2

4.4). Respecto a las elecciones de  $K$  y  $\lambda$  realizadas, descritas en la tabla 4.1, notemos que una muestra estratificada de Pob2Aa y una muestra cluster de Pob2Ba tendrán respectivamente las características habituales de cada uno de estos tipos de muestreo que comentábamos al principio del capítulo.

Para la obtención de las tres muestras para cada población (para tamaño muestral total  $n = 800$  y  $n = 1600$ ), se ha procedido de la siguiente manera:

- para la obtención de la muestra de la muestra aleatoria, se seleccionaron aleatoriamente y sin reemplazamiento  $n$  elementos de los 10000 de la población.
- para la obtención de la muestra estratificada se seleccionaron aleatoriamente y sin reemplazamiento de cada unidad (estrato),  $\frac{n}{K}$  elementos de los  $\frac{10000}{K}$  existentes.
- para la obtención de la muestra cluster, se seleccionaron aleatoriamente y sin reemplazamiento  $\frac{nK}{10000}$  unidades (clusters) de las  $K$  existentes y se muestrearon totalmente.

Notemos que la muestra estratificada cumple que el número de elementos muestreados en cada estrato es el mismo (hipótesis (4.30)), lo que implica junto a la propiedad de que el número de elementos en todos los estratos coincide, el cumplimiento de las hipótesis (4.30) y (4.31) del teorema 4.4.

#### 4.4.2 Descripción del modelo aplicado para el análisis de las poblaciones simuladas

Para realizar inferencias sobre las poblaciones simuladas anteriores se han aplicados los modelos descritos en las definiciones 4.1 y 4.2, el primero para las muestras aleatorias y el segundo para las muestras estratificadas y cluster, tomando  $\sigma^2 = 15^2$ , esto es el valor auténtico de la varianza de la población, puesto que estamos asumiendo que la varianza poblacional es conocida. En el caso del modelo en dos etapas se tomará también como valor de  $\lambda$ , el auténtico  $\lambda$  de la población simulada por la misma razón.

#### 4.4.3 Resultados obtenidos

El estimador Bayes para la media poblacional y la desviación típica dada la muestra para cada pareja población-muestra, se obtienen por la aplicación de las expresiones

POBLACION	n	MUESTRA ALEATORIA		MUESTRA ESTRATIFICADA		MUESTRA CLUSTER	
		ESTIMADOR	D T	ESTIMADOR	D T	ESTIMADOR	D T
Pob2Aa	800	81.09	0.2587	81.36	0.0776	88.78	72.51
$\bar{Y} = 81.46$	1600	81.48	0.1181	81.35	0.0354	75.65	33.10
Pob2Ab	800	78.71	0.2587	78.66	0.1811	72.78	31.23
$\bar{Y} = 78.45$	1600	78.32	0.1181	78.30	0.0827	75.51	14.25
Pob2Ba	800	79.99	0.2587	79.25	0.0776	80.76	4.605
$\bar{Y} = 79.45$	1600	79.60	0.1181	79.54	0.0354	80.02	2.102
Pob2Bb	800	80.18	0.2587	79.31	0.1811	80.21	2.121
$\bar{Y} = 79.29$	1600	79.38	0.1181	79.40	0.0827	78.37	0.9684

Tabla 4.2: Estimador Bayes de la media poblacional y su desviación típica

(4.14), (4.32) y (4.34) (notemos, nuevamente, que las muestras estratificadas y cluster cumplen respectivamente las hipótesis (4.30)-(4.31) y (4.33)), y vienen reflejados en la tabla 2.2, que contiene además el auténtico valor de la media poblacional de cada una de las cuatro poblaciones.

Notemos del análisis de los resultados que:

\* Las desviaciones típicas de la media poblacional condicionadas a la muestra estratificada son menores que las obtenidas a partir de la muestra aleatoria, especialmente cuando  $\lambda = 0.3$  (en los muestreos estratificados "buenos"  $\lambda$  es pequeño), y las desviaciones típicas de la media poblacional condicionada a la muestra cluster son mayores que las obtenidas para la muestra aleatoria y, obviamente, la muestra estratificada, sobre todo cuando  $K = 25$  (en los muestreos estratificados, usualmente,  $K$  es grande y en el ejemplo para  $K = 400$ , las desviaciones, aunque superiores a las del muestreo aleatorio son mucho menores que las que se producen con  $K = 25$ ). Estos resultados eran esperables y responden a los resultados comparativos que se establecían en la subsección 4.3.2 en base al teorema 4.4.

\* Si comparamos el valor estimado para la media poblacional con el valor auténtico de la media poblacional se confirman las relaciones establecidas en el punto anterior. La mejor estimación es proporcionada por los muestreos estratificados (errores absolutos alrededor de  $\pm 0.15$ ), la segunda estimación es la proporcionada por el muestreo aleatorio (errores absolutos alrededor de  $\pm 0.2$  y  $\pm 0.7$ ) y la peor estimación es la proporcionada con el muestreo cluster (con errores alrededor de las 5 unidades cuando  $K$  es pequeño).

## 4.5 Comentarios finales

En este capítulo hemos comprobado teóricamente y con un ejemplo que bajo algunas hipótesis, la estimación basada en muestreos estratificados es más precisa que la basada

en muestreos aleatorios y que la estimación basada en muestreos cluster es menos precisa que la estimación basada en muestreos aleatorios. Entre las hipótesis introducidas para la comparación y sobre las matizaciones para establecer el resultado comparativo en presencia de covariables, queremos destacar las siguientes observaciones:

1.- El análisis en presencia de covariables requiere para establecer resultados comparativos más concretos de la modelización de la matriz de covariable  $X$  como matriz aleatoria.

2.- La importancia en la descripción del modelo de muestreo en dos etapas del parámetro  $\lambda$  y las dificultades que puede entrañar su valoración a priori (por su relación con la varianza de la población y la varianza de las pendientes de cada unidad), sugiere la formulación de un modelo que asuma  $\lambda$  desconocido.

3.- El objetivo de una comparación analítica de los resultados obtenidos por los tres mecanismos de muestreo (incluso en el caso sin covariables), condiciona el estudio en varios aspectos: empleo de un mismo modelo para los muestreos estratificados y cluster (el modelo sugerido en la subsección 4.2.1 requiere de métodos computacionales para su implementación), la hipótesis sobre el parámetro  $\lambda$  que se asume conocido (nuevamente, modelizaciones sobre  $\lambda$  van a requerir de métodos computacionales) y la introducción de hipótesis sobre los muestreos para poder comparar los resultados inferenciales sobre la media obtenidos a partir de los mismos.

Todo ello, nos conduce a proponer una línea de investigación futura (complementaria a los estudios realizados en este capítulo) consistente en la introducción de modelos algo más complejos, que permitan estudiar más formalmente el caso con covariables y/o modelicen  $\lambda$ , y en su estudio mediante el empleo de procedimientos computacionales.

## Capítulo 5

# Análisis Bayesiano de Rutas Aleatorias

### 5.1 Introducción

En el capítulo anterior hemos estudiado comparativamente los resultados (concretamente el error cometido en la estimación de la media poblacional bajo hipótesis en las que los estimadores Bayes eran coincidentes), de la estimación bajo un muestreo aleatorio, un muestreo estratificado y un muestreo cluster y en qué condiciones un procedimiento daba mejores resultados que otro; sin embargo en la práctica la obtención de una muestra aleatoria, estratificada o incluso cluster no es fácil. Efectivamente, estos muestreos precisan de la existencia de un censo de la población, (para poder elegir así aleatoriamente a los elementos de población que han de ser muestreados) y dicho censo puede ser muy costoso económicamente o muy difícil (a veces imposible) de obtener (por ejemplo, los censos de animales salvajes, de árboles, de peces en un lago,... y los censos de poblaciones temporeras, por ejemplo, censos de turistas, de aves migratorias,...).

Estas dificultades y la necesidad de poder estudiar e inferir en este tipo de poblaciones sin censo ha conducido al empleo de procedimientos que permitan obtener una muestra “aproximadamente” aleatoria de las mismas. En esta línea, Kish(65)[65] es el primero en describir algunos mecanismos y proponer para muestrear a los elementos de una ciudad un mecanismo de muestreo por *rutas aleatorias*. El muestreo por rutas aleatorias (el más común en la práctica cuando la población está formada por las personas de una localidad y también aplicable cuando la población está situada en una zona geográfica que es fácilmente cartografiable), consiste en la división física de la población a través de su localización geográfica particionando la zona en rutas aleatorias disjuntas, que se obtienen definiendo recorridos que pueden responder a la estructura física de la zona (por ejemplo: calles, caminos en el campo,...) o pueden estar definidas de una manera más casual mediante instrucciones muy específicas (por ejemplo: subir la

calle A, a la segunda manzana torcer a la derecha, a la quinta manzana,...) y una vez realizada esta partición, se procede a la elección aleatoria de un número determinado de rutas y dentro de cada ruta de un número determinado de elementos por un procedimiento instrumentalizado de modo que la elección de los elementos en cada ruta sea "aleatoria".

El análisis de una muestra de población obtenida por estos métodos no estrictamente aleatorios como si se tratará de una muestra aleatoria de la población en estudio (que es el análisis usualmente aplicado por los encuestadores) ha sido ampliamente criticado por el desajuste que en la práctica se produce entre la muestra así obtenida y la hipotética muestra aleatoria del modelo teórico, que conduce usualmente a sobreestimar la precisión de los estimadores obtenidos.

El objetivo de este capítulo es la presentación y análisis de modelos Bayesianos para poblaciones no censadas, que permitan incorporar la incertidumbre adicional que se produce cuando la muestra de la población se obtiene por una técnica de muestreo basada en rutas aleatorias en vez de aplicar un muestreo aleatorio, mostrando mediante ejemplos cómo estos modelos representan un sustrato teórico más adecuado para la inferencia para muestras obtenidas por rutas aleatorias que el habitualmente aplicado, puesto que proporcionan mejores estimaciones acerca de la precisión de los estimadores aplicados o al menos nos indican si efectivamente en un caso concreto estamos sobreestimando la precisión de los mismos. El capítulo para su exposición se dividirá en cinco secciones incluyendo a la sección actual introductoria; la sección 5.2 presentará los tres modelos para rutas aleatorias que analizaremos en este capítulo, todos ellos bajo la hipótesis de un muestreo por rutas de una población y dos modelos Bayes adecuados para poblaciones muestreadas según un muestreo aleatorio, el primero de ellos asume que el tamaño de la población es conocido (hipótesis que es cierta cuando realmente se realiza un muestreo aleatorio, pero que en la práctica no se da habitualmente cuando se realiza un muestreo por rutas) y el segundo que éste es desconocido. La sección 5.3 proporcionará las distribuciones condicionales propias del análisis Bayesiano de los modelos propuestos. La sección 5.4 se dedicará al análisis del funcionamiento de los tres modelos presentados, comparando los resultados obtenidos de su aplicación sobre tres poblaciones distintas frente a los obtenidos asumiendo los modelos de muestreo aleatorio definidos y estudiados en las secciones 5.2 y 5.3, y finalmente la sección 5.5 incluirá los comentarios finales.

## **5.2 Cinco modelos Bayes para describir una población**

Como ya indicábamos en la introducción dedicaremos esta sección a la presentación de cinco modelos, los tres primeros (definiciones 5.1, 5.2 y 5.3) constituyen los modelos propuestos en esta tesis para la inferencia en poblaciones muestreadas por rutas aleatorias y el cuarto y quinto (definiciones 5.4 y 5.5) son modelos para realizar inferencia en poblaciones muestreadas según un muestreo aleatorio.

Antes de pasar a la definición de estos modelos, modificaremos la notación que usábamos en los capítulos anteriores, para ajustarla a las características de una población no censada (cuyo tamaño es usualmente desconocido en la práctica), muestreada por rutas aleatorias (cuyo tamaño es asimismo usualmente desconocido) y describiremos las características que ha de cumplir un muestreo por rutas aleatorias para pertenecer al conjunto de los modelizados en este capítulo (a partir de ahora, al referirnos a un muestreo aleatorio por rutas nos referiremos a un muestreo de rutas con las propiedades definidas a continuación). En concreto, asumiremos que:

- el número de elementos de la población es desconocido
- y el número de elementos por ruta aleatoria es desconocido si la ruta no es muestreada y conocido si la ruta es muestreada (en la práctica los tamaños de todas las rutas sean o no muestreadas es usualmente desconocido conociéndose a lo sumo estimadores de los tamaños de rutas muestreadas; los modelos que presentaremos aquí son simplificaciones de la realidad, pero son buenas aproximaciones cuando tenemos buenos estimadores de los tamaños de las rutas muestreadas),

y añadiremos a las notaciones habituales, reunidas en el apéndice A, las siguientes:

- $T$ , una variable aleatoria que mide el número de elementos de la población,
- $T_i$ , una variable aleatoria que mide el número de elementos de la ruta  $i$ ,  $i = 1, 2, \dots, K$ ,
- señalemos, también que en este capítulo cometeremos el abuso de notación de representar a las realizaciones de las variables de los items anteriores también con mayúsculas,

interpretando que las unidades son ahora rutas, con  $K$ , número de rutas (antes número de unidades) conocido. Y consideraremos que el procedimiento seguido para la selección de la muestra ha sido el siguiente:

- selección por de una muestra aleatoria de  $k$  rutas distintas de las  $K$  existentes, (las rutas seleccionadas corresponden a los índices del conjunto  $\{i_1, i_2, \dots, i_k\}$ ),
- estimación o conocimiento de los tamaños  $T_i$  (realización de la variable  $T_i$ ) de las rutas seleccionadas para ser muestreadas y selección para cada ruta  $i \in \{i_1, \dots, i_k\}$  de una muestra aleatoria de  $n_i$  elementos con índices distintos ( $n_i \leq T_i$ , con  $\sum n_i = n$ ) de los  $T_i$  existentes (los elementos muestreados en la ruta  $i$ ,  $i \in \{i_1, i_2, \dots, i_k\}$ , corresponden a los índices  $(i, j)$  con  $j \in s_i$ , midiéndose para estos elementos la característica de interés  $(y_{ij})$ ).

Las limitaciones prácticas de este tipo de muestreo son, por una parte el necesario conocimiento de los tamaños o de una estimación de los mismos (en cierta manera a priori, para seleccionar un tamaño de la muestra para cada ruta inferior al tamaño de la misma) y la hipótesis de aleatoriedad en la elección de los elementos de cada ruta.

A continuación pasaremos a describir tres modelos para rutas aleatorias, con los que se propone, básicamente, incorporar la incertidumbre adicional derivada del uso de un muestreo por rutas aleatorias en lugar de un muestreo aleatorio, mediante un modelo jerárquico Bayes. El primer modelo (definición 5.1), aunque en algunos casos demasiado simplificado para ser realista, es en cualquier caso, un instrumento útil (como veremos más adelante por medio de ejemplos) para estudiar la adecuación de las hipótesis de un muestreo aleatorio. Los otros dos modelos (definiciones 5.2 y 5.3), se introducen también con este fin exploratorio, incorporando cada uno de ellos sólo una parte de la incertidumbre adicional estudiada a través del primero: el modelo de la definición 5.2 en forma de diferencias entre la variabilidad (heterogeneidad) de las rutas y el de la definición 5.3 entre las medias de las rutas. Grandes diferencias entre los resultados inferenciales obtenidos a partir de estos tres modelos, indicarán probablemente que las hipótesis de muestreo aleatorio no son adecuadas para el análisis de la población en estudio a través de la muestra obtenida por rutas y que es precisa una cuidadosa modelización de las rutas, incluso si el modelo jerárquico de la definición 5.1 no se juzga adecuado. Siendo, además las varianzas finales de la función de características estudiada (en nuestro caso media de cada ruta y media poblacional) derivadas del modelo de la definición 5.1 una medida de la fiabilidad del estimador mucho más adecuada que la derivada del muestreo aleatorio.

**Definición 5.1** Dado el vector poblacional  $\mathbf{Y} = (\mathbf{Y}_1^t, \mathbf{Y}_2^t, \dots, \mathbf{Y}_K^t)^t$  con  $\mathbf{Y}_i = (Y_{i1}, Y_{i2}, \dots, Y_{iT_i})^t$ ,  $i = 1, 2, \dots, K$  y el vector aleatorio  $\mathbf{T} = (T_1, T_2, \dots, T_K)^t$ , que representa el número de elementos de cada una de  $K$  rutas, llamaremos modelo Bayes de rutas aleatorias al definido por el siguiente modelo jerárquico:

$$\begin{aligned}
\mathbf{Y}_i | \mathbf{T}, \theta, \sigma^2 &\sim N_{T_i}(\mathbf{1}_{T_i} \theta_i, \mathbf{I}_{T_i} \sigma_i^2), \quad i = 1, 2, \dots, K, \text{ independientes,} \\
\theta_i | \mu, \sigma^2 &\sim N(\mu, c_i \sigma_i^2), \quad i = 1, 2, \dots, K, \text{ independientes,} \\
\mu &\sim N(\mu_0, \sigma_0^2), \\
\sigma_i^2 &\sim Ga^{-1}(a_0, b_0), \quad i = 1, 2, \dots, K, \text{ i.i.d.,} \\
T_i | \lambda &\sim Po(\lambda), \quad i = 1, 2, \dots, K, \text{ i.i.d.,} \\
\lambda &\sim Ga(\alpha, \beta),
\end{aligned} \tag{5.1}$$

donde  $\theta = (\theta_1, \theta_2, \dots, \theta_K)^t$ ,  $\sigma^2 = (\sigma_1^2, \sigma_2^2, \dots, \sigma_K^2)^t$ ,  $\mu_0$  es un escalar positivo,  $\sigma_0^2$ ,  $\{c_i\}_{i=1}^K$ ,  $a_0$ ,  $b_0$ ,  $\alpha$  y  $\beta$  son escalares positivos asimismo conocidos,  $\mathbf{1}_{T_i}$  e  $\mathbf{I}_{T_i}$  son respectivamente el vector de unos y la matriz identidad de dimensión  $T_i$  e i.i.d. es la abreviatura de independientes e igualmente distribuidos.

**Definición 5.2** Dado el vector poblacional  $\mathbf{Y} = (\mathbf{Y}_1^t, \mathbf{Y}_2^t, \dots, \mathbf{Y}_K^t)^t$  con  $\mathbf{Y}_i = (Y_{i1}, Y_{i2}, \dots, Y_{iT_i})^t$ ,  $i = 1, 2, \dots, K$  y el vector aleatorio  $\mathbf{T} = (T_1, T_2, \dots, T_K)^t$ , que representa el número de elementos de cada una de  $K$  rutas, llamaremos modelo Bayes de rutas aleatorias con medias iguales al definido por el siguiente modelo jerárquico:

$$\begin{aligned}
\mathbf{Y}_i | \mathbf{T}, \mu, \sigma^2 &\sim N_{T_i}(\mathbf{1}_{T_i} \mu, \mathbf{I}_{T_i} \sigma_i^2), \quad i = 1, 2, \dots, K, \text{ independientes,} \\
\mu &\sim N(\mu_0, \sigma_0^2), \\
\sigma_i^2 &\sim Ga^{-1}(a_0, b_0), \quad i = 1, 2, \dots, K, \text{ i.i.d.,} \\
T_i | \lambda &\sim Po(\lambda), \quad i = 1, 2, \dots, K, \text{ i.i.d.,} \\
\lambda &\sim Ga(\alpha, \beta),
\end{aligned} \tag{5.2}$$

donde  $\sigma^2 = (\sigma_1^2, \sigma_2^2, \dots, \sigma_K^2)^t$ ,  $\mu_0$  es un escalar positivo,  $\sigma_0^2$ ,  $a_0$ ,  $b_0$ ,  $\alpha$  y  $\beta$  son escalares positivos asimismo conocidos,  $\mathbf{1}_{T_i}$  e  $\mathbf{I}_{T_i}$  son respectivamente el vector de unos y la matriz identidad de dimensión  $T_i$  e i.i.d. es la abreviatura de independientes e igualmente distribuidos.

**Definición 5.3** Dado el vector poblacional  $\mathbf{Y} = (\mathbf{Y}_1^t, \mathbf{Y}_2^t, \dots, \mathbf{Y}_K^t)^t$  con  $\mathbf{Y}_i = (Y_{i1}, Y_{i2}, \dots, Y_{iT_i})^t$ ,  $i = 1, 2, \dots, K$  y el vector aleatorio  $\mathbf{T} = (T_1, T_2, \dots, T_K)^t$ , que representa el número de elementos de cada una de  $K$  rutas, llamaremos **modelo Bayes de rutas aleatorias homocedásticas** al definido por el siguiente modelo jerárquico:

$$\begin{aligned}
\mathbf{Y}_i | \mathbf{T}, \theta, \sigma^2 &\sim N_{T_i}(\mathbf{1}_{T_i} \theta_i, \mathbf{I}_{T_i} \sigma^2), \quad i = 1, 2, \dots, K, \text{ independientes,} \\
\theta_i | \mu, \sigma^2 &\sim N(\mu, c\sigma^2), \quad i = 1, 2, \dots, K \text{ i.i.d.,} \\
\mu &\sim N(\mu_0, \sigma_0^2), \\
\sigma^2 &\sim Ga^{-1}(a_0, b_0), \\
T_i | \lambda &\sim Po(\lambda), \quad i = 1, 2, \dots, K, \text{ i.i.d.,} \\
\lambda &\sim Ga(\alpha, \beta),
\end{aligned} \tag{5.3}$$

donde  $\theta = (\theta_1, \theta_2, \dots, \theta_K)^t$ ,  $\mu_0$  es un escalar positivo,  $\sigma_0^2$ ,  $c$ ,  $a_0$ ,  $b_0$ ,  $\alpha$  y  $\beta$  son escalares positivos asimismo conocidos,  $\mathbf{1}_{T_i}$  e  $\mathbf{I}_{T_i}$  son respectivamente el vector de unos y la matriz identidad de dimensión  $T_i$  e i.i.d. es la abreviatura de independientes e igualmente distribuidos.

Los modelos de superpoblación condicionados a  $\mu$  y  $\mathbf{T}$  para los modelos anteriores, que servirán de base para un comentario más amplio de los tres modelos que será realizado en la próxima sección, vienen dados por el siguiente teorema:

**Teorema 5.1** Los modelos de superpoblación condicionados a  $\mu$  y  $\mathbf{T}$ , son:

(a) para el modelo Bayes para rutas aleatorias definido en (5.1):

$$\mathbf{Y}_i | \mu, T_i \sim t_{T_i}(\mathbf{1}_{T_i} \mu, \frac{b_0}{a_0} (\mathbf{I}_{T_i} + \mathbf{1}_{T_i} \mathbf{1}_{T_i}^t c_i), 2a_0), \tag{5.4}$$

independientes,  $i = 1, 2, \dots, K$ , con media y varianza marginales, para  $j = 1, 2, \dots, T_i$ ,  $i = 1, 2, \dots, K$ :

$$\begin{aligned} E(Y_{ij}|\mu, T_i) &= \mu, \\ \text{Var}(Y_{ij}|\mu, T_i) &= \frac{b_0}{a_0 - 1} (1 + c_i), \end{aligned} \quad (5.5)$$

y covarianzas, para  $j, j^* = 1, 2, \dots, T_i$ ,  $i, i^* = 1, 2, \dots, K$ ,

$$\text{Cov}(Y_{ij}, Y_{i^*j^*}|\mu, T_i, T_{i^*}) = \begin{cases} \frac{b_0}{a_0 - 1} c_i & i = i^* \\ 0 & i \neq i^*, \end{cases} \quad (5.6)$$

(b) para el modelo Bayes para rutas aleatorias con medias iguales definido en (5.2):

$$\mathbf{Y}_i|\mu, T_i \sim t_{T_i}(\mathbf{1}_{T_i}\mu, \frac{b_0}{a_0}\mathbf{I}_{T_i}, 2a_0), \quad (5.7)$$

independientes,  $i = 1, 2, \dots, K$ , con media y varianza marginales, para  $j = 1, 2, \dots, T_i$ ,  $i = 1, 2, \dots, K$ :

$$\begin{aligned} E(Y_{ij}|\mu, T_i) &= \mu, \\ \text{Var}(Y_{ij}|\mu, T_i) &= \frac{b_0}{a_0 - 1}, \end{aligned} \quad (5.8)$$

y covarianzas entre elementos 0,

(c) para el modelo Bayes para rutas aleatorias homocedásticas definido en (5.3):

$$\mathbf{Y}|\mu, T_i, i = 1, 2, \dots, K \sim t_T(\mathbf{1}_T\mu, \frac{b_0}{a_0}(\mathbf{I}_T + c \cdot \text{diag}_{i=1}^K\{\mathbf{1}_{T_i}\mathbf{1}_{T_i}^t\}), 2a_0), \quad (5.9)$$

con media y varianza marginales, para  $j = 1, 2, \dots, T_i$ ,  $i = 1, 2, \dots, K$ :

$$\begin{aligned} E(Y_{ij}|\mu, T_i) &= \mu, \\ \text{Var}(Y_{ij}|\mu, T_i) &= \frac{b_0}{a_0 - 1} (1 + c), \end{aligned} \quad (5.10)$$

y covarianzas, para  $j, j^* = 1, 2, \dots, T_i$ ,  $i, i^* = 1, 2, \dots, K$ ,

$$\text{Cov}(Y_{ij}, Y_{i^*j^*}|\mu, T_i, T_{i^*}) = \begin{cases} \frac{b_0}{a_0 - 1} c & i = i^* \\ 0 & i \neq i^*, \end{cases} \quad (5.11)$$

donde  $T = \sum_{i=1}^K T_i$ .

**Demostración.** La demostración se apoya en la obtención de la distribución de  $\mathbf{Y}|\mu, \mathbf{T}$  para los tres modelos, aplicando los resultados (a) de los lemas B.8, B.9 y B.11, y las propiedades sobre los dos primeros momentos de una distribución  $t$ .

Las distribuciones del vector  $\mathbf{Y}|\mu, \mathbf{T}$  dadas en (5.4), (5.7) y (5.9), se obtienen de la siguiente manera:

- A partir de las distribuciones de  $\mathbf{Y}|\theta, \sigma^2, \mathbf{T}$  y  $\theta|\mu, \sigma^2$  definidas en (5.1), se obtiene, aplicando el resultado (a) del lema B.8, la distribución de  $\mathbf{Y}|\mu, \sigma^2, \mathbf{T}$ , que viene dada por:

$$\mathbf{Y}|\mu, \sigma^2, \mathbf{T} \sim N_T(\mathbf{1}_T\mu, \text{diag}_{i=1}^K\{(\mathbf{I}_{T_i} + \mathbf{1}_{T_i}\mathbf{1}_{T_i}^t, c_i)\sigma_i^2\}).$$

Y a partir de la distribución de  $\mathbf{Y}|\mu, \sigma^2, \mathbf{T}$  y la distribución inicial de  $\sigma^2$  dada en (5.1), se obtiene aplicando el resultado (a) del lema B.11 la distribución de  $\mathbf{Y}|\mu, \mathbf{T}$  dada en (5.4).

- A partir de la distribución de  $\mathbf{Y}|\mu, \sigma^2, \mathbf{T}$  y la distribución inicial de  $\sigma^2$  dadas en (5.2), se obtiene aplicando el resultado (a) del lema B.11 la distribución de  $\mathbf{Y}|\mu, \mathbf{T}$  dada en (5.7).
- A partir de las distribuciones de  $\mathbf{Y}|\theta, \sigma^2, \mathbf{T}$  y  $\theta|\mu, \sigma^2$  definidas en (5.3), se obtiene, aplicando el resultado (a) del lema B.8, la distribución de  $\mathbf{Y}|\mu, \sigma^2, \mathbf{T}$ , que viene dada por:

$$\mathbf{Y}|\mu, \sigma^2, \mathbf{T} \sim N_T(\mathbf{1}_T\mu, \sigma^2 \text{diag}_{i=1}^K\{\mathbf{I}_{T_i} + \mathbf{1}_{T_i}\mathbf{1}_{T_i}^t, c\}).$$

Y a partir de la distribución de  $\mathbf{Y}|\mu, \sigma^2, \mathbf{T}$  y la distribución inicial de  $\sigma^2$  dada en (5.3), se obtiene aplicando el resultado (a) del lema B.9 la distribución de  $\mathbf{Y}|\mu, \mathbf{T}$  dada en (5.9).

Los resultados sobre los primeros momentos dados en (5.5) y (5.6), (5.8) y (5.10) y (5.11), se obtienen a partir de los primeros momentos de las distribuciones (5.4), (5.7) y (5.9) respectivamente.

C.S.Q.D.

### 5.2.1 Comentarios de los modelos Bayes para rutas aleatorias

Notemos que el modelo Bayes para rutas aleatorias definido en (5.1) y los modelos Bayes para rutas aleatorias con medias iguales y homocedásticas definidos en (5.2) y (5.3) son tres modelos de población que se ajustan a la estructura que adquiere una población al ser dividida en rutas y a las características de muestreo por rutas que definíamos al principio de la sección, en el sentido de que en ellos, los elementos de la población se clasifican y modelizan según la ruta a la que pertenecen (cuyo tamaño se considera como es usual en la práctica desconocido, salvo en el caso de rutas muestreadas) con media según ruta distinta y varianza distinta en el modelo Bayes para rutas aleatorias

y con medias iguales o varianzas iguales por ruta en los modelos Bayes para rutas aleatorias con medias iguales y homocedásticas respectivamente.

Efectivamente, los tres modelos asumen que la población está organizada según rutas aleatorias con vectores  $\mathbf{Y}_i|\theta, \sigma^2, \mathbf{T}$ ,  $\mathbf{Y}_i|\mu, \sigma^2, \mathbf{T}$  y  $\mathbf{Y}_i|\theta, \sigma^2, \mathbf{T}$ , respectivamente independientes y normales (situación ésta que sugiere un muestreo en dos etapas por un muestreo aleatorio, en primer lugar de las rutas y en segundo lugar de los elementos de las rutas muestreadas), con:

- rutas con medias y varianzas distintas entre sí en el caso del modelo Bayes para rutas aleatorias,
- rutas con medias iguales y varianzas distintas entre sí en el caso del modelo Bayes para rutas aleatorias con medias iguales y,
- rutas con medias distintas y varianzas iguales entre sí en el caso del modelo Bayes para rutas aleatorias homocedásticas,

Estas hipótesis conducen a los tres modelos de superpoblación, condicionados a  $\mu$  y  $\mathbf{T}$ , descritos en el teorema 5.1, según los cuales, el vector poblacional se distribuye según una  $t$  de Student con la particularidad de que:

- en el modelo Bayes de rutas aleatorias, las rutas son independientes entre sí y los elementos de cada ruta están correlados entre sí,
- en el modelo Bayes de rutas aleatorias con medias iguales, las rutas son independientes entre sí y los elementos de cada ruta están incorrelados entre sí y,
- en el modelo Bayes de rutas aleatorias homocedásticas las rutas son incorreladas entre sí y los elementos de cada ruta están correlados entre sí,

frente al modelo habitual de superpoblación (para varianza desconocidas), condicionado a  $\mu$  y  $T$ , para una población sobre la que se aplica un muestreo aleatorio, según el cual el vector poblacional se distribuiría según una  $t$  con elementos de la población incorrelados entre sí<sup>1</sup>

Los modelos por rutas proporcionan, por tanto, un instrumento “simple” para juzgar las desviaciones inferenciales producidas al analizar una población por una muestra por rutas como si ésta fuera muestra aleatoria. Ya que constituyen variantes “sencillas” del modelo habitual para poblaciones muestreadas con muestras aleatorias que incorporan en mayor o menor medida parte de la incertidumbre derivada del uso de una muestra por rutas aleatorias en lugar de una muestra aleatoria.

### 5.2.2 Modelos Bayes para muestreos aleatorios con varianzas desconocidas

Para comparar el funcionamiento de los modelos para rutas aleatorias definidos en las definiciones 5.1, 5.2 y 5.3 entre sí en relación con los resultados habitualmente obtenidos

<sup>1</sup>si asumimos el modelo definido por:  $\mathbf{Y}|\mu, \sigma^2 \sim N_T(\mathbf{1}_T\mu, \mathbf{I}_T\sigma^2)$  y  $\sigma^2 \sim Ga^{-1}(a_0, b_0)$ , entonces  $\mathbf{Y}|\mu, T \sim t_T(\mathbf{1}_T\mu, \mathbf{I}_T \frac{b_0}{a_0}, 2a_0)$ .

bajo la hipótesis de que un muestreo por rutas es aproximadamente un muestreo aleatorio simple de la población, describiremos en esta subsección dos modelos Bayes para muestreos aleatorios. El primero de ellos constituye el modelo (mínimo informativo) habitualmente aplicado cuando se aplica un muestreo aleatorio y se supone la varianza de la población desconocida y el segundo constituye una variante del primero que supone adicionalmente que el tamaño de la población (que se describe con la variable aleatoria  $T$ ) es desconocido.

**Definición 5.4** *Dado el vector poblacional  $\mathbf{Y} = (Y_1, Y_2, \dots, Y_T)^t$  y el tamaño de la población  $N$  conocido, llamaremos modelo Bayes para muestras aleatorias (m.a.s) con varianza desconocida al definido por:*

$$\begin{aligned} \mathbf{Y}|\mu, \sigma^2 &\sim N_N(\mathbf{1}_N\mu, \mathbf{I}_N\sigma^2), \\ \pi(\mu, \sigma^2) &\propto \frac{1}{\sigma^2}, \end{aligned} \quad (5.12)$$

donde  $\alpha$  y  $\beta$  son escalares positivos conocidos, y  $\mathbf{1}_N$  e  $\mathbf{I}_N$  son respectivamente el vector de unos y la matriz identidad de dimensión  $N$ .

El modelo Bayes para m.a.s con varianza desconocida es un caso particular del modelo Bayes de Regresión Normal-Gamma mínimo informativo descrito en la subsección 2.2.4 del capítulo 2. Este modelo parte de la hipótesis de tamaño poblacional conocido; esta hipótesis, en las condiciones descritas en esta sección para el análisis por rutas no se cumple, lo que nos conduce a la siguiente definición de un modelo Bayes para m.a.s con varianza y tamaño desconocidos.

**Definición 5.5** *Dado el vector poblacional  $\mathbf{Y} = (Y_1, Y_2, \dots, Y_T)^t$  y  $T = T_u + N$ , con  $N$ , un entero positivo conocido que representa el número de elementos que según nuestros conocimientos tiene al menos la población, llamaremos modelo Bayes para muestras aleatorias (m.a.s) con varianza y tamaño poblacional desconocidos al definido por:*

$$\begin{aligned} \mathbf{Y}|T, \mu, \sigma^2 &\sim N_T(\mathbf{1}_T\mu, \mathbf{I}_T\sigma^2), \\ \pi(\mu, \sigma^2) &\propto \frac{1}{\sigma^2}, \\ T_u &\sim Po(t\lambda), \\ \lambda &\sim Ga(\alpha, \beta), \end{aligned} \quad (5.13)$$

donde  $\alpha$ ,  $\beta$  y  $t$  son escalares positivos conocidos, y  $\mathbf{1}_T$  e  $\mathbf{I}_T$  son respectivamente el vector de unos y la matriz identidad de dimensión  $T$ .

El modelo Bayes para m.a.s. con varianza y tamaño poblacional desconocidos, se apoya en una notación que pretende facilitar la comparación del modelo con los modelos Bayes para rutas descritos con anterioridad a través de una introducción fácil de la información que disponemos sobre el tamaño de la población en un procedimiento de muestreo por rutas del tipo explicado en la sección anterior, pero permite además modelizar situaciones en las que el tamaño es desconocido y no se ha realizado ningún experimento para aprender sobre el mismo. Concretamente:

- Para introducir la información que un muestreo por rutas del tipo descrito proporciona sobre el tamaño poblacional, basta con suponer que la población se puede dividir en  $K$  segmentos y en cada uno de ellos se puede asumir que  $T_i$ ,  $i = 1, 2, \dots, K$ , el tamaño de cada segmento se distribuye Poisson independientes con parámetro  $\lambda$ . Y considerar que de estos  $K$  segmentos, se muestrean  $k$  (correspondientes a los índices:  $(i_1, i_2, \dots, i_k)$  y se observa  $N_s = \sum_{i \in \{i_1, \dots, i_k\}} T_i$ . Bajo estas hipótesis se cumple obviamente que:  $T_u | \lambda, N_s \sim Po((K - k)\lambda)$  (por propiedades conocidas de la distribución Poisson) y  $\lambda | N_s \sim Ga(\alpha + N_s, \beta + k)$  (aplicando el resultado (c) del lema B.12); con lo que la información del experimento realizado es fácilmente trasladable al modelo de la definición 5.5.
- Cuando no se tenga información acerca del tamaño poblacional, obviamente  $T = T_u + n$ , con  $n$  el número de elementos muestreados de la población con lo que nuevamente, si asumimos que  $T_u \sim Po(\lambda^*)$  con  $\lambda^* \sim Ga(\alpha^*, \beta^*)$ , tenemos una situación de incertidumbre sobre el tamaño poblacional fácilmente trasladable al modelo anterior.

En la sección 5.4 emplearemos estos modelos (el segundo de ellos en la situación analizada en el primer punto de los comentarios del párrafo anterior, con  $N_s \equiv \sum_{i \in \{i_1, \dots, i_k\}} T_i$ ,  $t \equiv K - k$ ,  $\alpha \equiv \alpha + N_s$  y  $\beta \equiv \beta + k$ ), para comparar los resultados obtenidos mediante los modelos para rutas aleatorias descritos en (5.1), (5.2) y (5.3), frente a los que se obtendrían asumiendo que el muestreo por rutas realizado constituye una muestra aleatoria de la población (en las situaciones de tamaño poblacional conocido y desconocido). Esta comparación se realizará a través de los estimadores de la media poblacional y la media no observada poblacional y su varianza condicionada a la muestra obtenida y los tamaños observados de las rutas muestreadas, tomándose en el caso particular del modelo de la definición 5.4 para obtener estimador y varianza condicionada a la muestra como valor de  $N$ , el auténtico tamaño de la población.

### 5.3 Análisis Bayesiano de los modelos propuestos

En esta sección, se emprenderá el análisis Bayes de los modelos presentados en la sección anterior a fin de obtener las distribuciones predictivas del vector poblacional  $\mathbf{Y}$  dada una muestra  $((s, \mathbf{y}_s))$  obtenida por un muestreo por rutas aleatorias del tipo descrito en la sección anterior y los tamaños observados  $\{T_i\}_{i \in \{i_1, \dots, i_k\}}$ . Estos modelos, tal y como veremos a continuación no son modelos conjugados y no dan lugar a distribuciones predictivas estándar ni tampoco a una estructura de condicionales

estándar de modo que los primeros momentos (que existen, tal y como veremos a continuación) de las distribuciones predictivas de la media poblacional no observada y la media poblacional (y por lo tanto los estimadores Bayes de la media poblacional no observada y la media poblacional y sus varianzas condicionadas a la muestra y a los tamaños de las rutas muestreadas) sean fáciles de calcular, lo que nos conduce a proponer un muestreo de dichas distribuciones por Gibbs Sampling (ver C.1). El algoritmo del Gibbs es fácil de implementar para estos modelos, dado que las distribuciones condicionales (que obtendremos en esta sección para los tres modelos por rutas) son fáciles de simular puesto que son distribuciones estándares o bien distribuciones con función de probabilidad log-concava.

La exposición de los resultados se realizará por modelos y constará de dos partes en el caso de los modelos Bayes para rutas aleatorias: una primera parte en la que se realizará todo el análisis condicional necesario para poder garantizar la existencia de distribuciones predictivas propias y las condiciones en las que estas distribuciones tienen primeros dos momentos y una segunda parte en la que vista la complejidad del análisis anterior se obtendrán las distribuciones condicionales para poder aplicar el Gibbs y obtener a partir de este método muestras de las distribuciones predictivas aproximadas para la media no observada de cada ruta, la media no observada de la población y la media poblacional<sup>2</sup>. La exposición de los resultados para el modelo Bayes para m.a.s constará en cambio de una única parte puesto que los modelos estudiados son más sencillos que contendrá el análisis condicional necesario y el estimador Bayes y varianzas condicionadas a la muestra para la media poblacional.

Las demostraciones de los resultados presentados en esta sección se apoyan en unos resultados previos enunciados y demostrados en forma de lemas en el apéndice B.

### 5.3.1 Modelo Bayes de rutas aleatorias

El análisis de este modelo comenzará por la obtención de las distribuciones condicionales necesarias para la inferencia según este modelo y de las condiciones suficientes para garantizar que las distribuciones predictivas de interés (dado que éstas no son distribuciones estándar), son propias y tienen primeros dos momentos.

**Teorema 5.2** *Consideremos el modelo Bayes para rutas aleatorias definido en (5.1), entonces se tiene que:*

(a)  $\mathbf{Y}_{u_i} | (s, \mathbf{y}_s), \mu, \mathbf{T}$  son independientes para  $i = 1, 2, \dots, K$  con distribuciones predictivas condicionales dadas por:

$$\begin{aligned} f(\mathbf{y}_{u_i} | (s, \mathbf{y}_s), \mu, \mathbf{T}) &= t_{T_i - n_i}(\mathbf{y}_{u_i} | 1_{T_i - n_i}, \tilde{\theta}_i, \mathbf{A}_i \frac{b_i}{a_i}, 2a_i), \quad i \in \{i_1, \dots, i_k\} \\ &= t_{T_i}(\mathbf{y}_{u_i} | 1_{T_i}, \mu, \mathbf{B}_i \frac{b_0}{a_0}, 2a_0), \quad i \notin \{i_1, \dots, i_k\}, \end{aligned} \quad (5.14)$$

<sup>2</sup>la restricción al cálculo de estas distribuciones predictivas no afecta a los resultados inferenciales que buscamos obtener y en cambio facilita los cálculos y la aplicación del programa BUGS para implementar el método Gibbs (ver C.2)

(b)  $\bar{Y}_{u_i}|(s, \mathbf{y}_s), \mu, \mathbf{T}$  son independientes para  $i = 1, 2, \dots, K$  con distribuciones predictivas condicionales dadas por:

$$\begin{aligned} f(\bar{y}_{u_i}|(s, \mathbf{y}_s), \mu, \mathbf{T}) &= t(\bar{y}_{u_i}|\tilde{\theta}_i, [(T_i - n_i)^{-1} + \frac{c_i}{n_i c_i + 1}] \frac{b_i}{a_i}, 2a_i), \quad i \in \{i_1, \dots, i_k\}, \\ &= t(\bar{y}_{u_i}|\mu, (T_i^{-1} + c_i) \frac{b_0}{a_0}, 2a_0), \quad i \notin \{i_1, \dots, i_k\}, \end{aligned} \quad (5.15)$$

(c)  $\bar{Y}_u|(s, \mathbf{y}_s), \sigma^2, \mu, \mathbf{T}$  se distribuye según una normal con media y varianza dadas respectivamente por:

$$\begin{aligned} E(\bar{Y}_u|(s, \mathbf{y}_s), \sigma^2, \mu, \mathbf{T}) &= \frac{1}{T_u + T_s - n} \left[ \sum_{i \in \{i_1, \dots, i_k\}} (T_i - n_i) \tilde{\theta}_i + \mu T_u \right], \quad (5.16) \\ \text{Var}(\bar{Y}_u|(s, \mathbf{y}_s), \sigma^2, \mu, \mathbf{T}) &= \frac{1}{(T_u + T_s - n)^2} \left[ \sum_{i \notin \{i_1, \dots, i_k\}} T_i (1 + T_i c_i) \sigma_i^2 + \right. \\ &\quad \left. + \sum_{i \in \{i_1, \dots, i_k\}} (T_i - n_i) \left( 1 + \frac{(T_i - n_i) c_i}{n_i c_i + 1} \right) \sigma_i^2 \right], \end{aligned}$$

(d)  $\sigma_i^2|(s, \mathbf{y}_s), \mu, \mathbf{T}$  son independientes para  $i = 1, 2, \dots, K$  con distribuciones condicionales  $\pi(\sigma_i^2|(s, \mathbf{y}_s), \mu, \mathbf{T}) = \pi(\sigma_i^2|(s, \mathbf{y}_s), \mu)$  dadas por:

$$\begin{aligned} \pi(\sigma_i^2|(s, \mathbf{y}_s), \mu) &= Ga^{-1}(\sigma_i^2|a_i, b_i), \quad i \in \{i_1, \dots, i_k\}, \\ &= Ga^{-1}(\sigma_i^2|a_0, b_0), \quad i \notin \{i_1, \dots, i_k\}, \end{aligned} \quad (5.17)$$

(e)  $\mu|(s, \mathbf{y}_s), \mathbf{T}$  se distribuye con función de densidad propia  $\pi(\mu|(s, \mathbf{y}_s), \mathbf{T}) = \pi(\mu|(s, \mathbf{y}_s))$  dada por:

$$\pi(\mu|(s, \mathbf{y}_s)) \propto \prod_{i \in \{i_1, \dots, i_k\}} \left[ 1 + \frac{1}{2b_0 Z_i} (\mu - \bar{y}_{s_i})^2 \right]^{-a_i} \times \exp \left\{ -\frac{1}{2\sigma_0^2} (\mu - \mu_0)^2 \right\}, \quad (5.18)$$

(f)  $\mathbf{T}_u|(s, \mathbf{y}_s), \{T_i\}_{i \in \{i_1, \dots, i_k\}}$  se distribuye con función de densidad  $f(\mathbf{T}_u|(s, \mathbf{y}_s), \{T_i\}_{i \in \{i_1, \dots, i_k\}}) = f(\mathbf{T}_u|\{T_i\}_{i \in \{i_1, \dots, i_k\}})$  dada por:

$$f(\mathbf{T}_u|\{T_i\}_{i \in \{i_1, \dots, i_k\}}) = \frac{(\beta + k)^{T_s + \alpha}}{\Gamma(T_s + \alpha)} \frac{\Gamma(T_u + T_s + \alpha)}{(\beta + K)^{T_u + T_s + \alpha}} \prod_{i \notin \{i_1, \dots, i_k\}} \frac{1}{T_i!}, \quad (5.19)$$

(g)  $T_u|(s, \mathbf{y}_s), \{T_i\}_{i \in \{i_1, \dots, i_k\}}$  se distribuye según una función de densidad  $f(T_u|(s, \mathbf{y}_s), \{T_i\}_{i \in \{i_1, \dots, i_k\}}) = f(T_u|\{T_i\}_{i \in \{i_1, \dots, i_k\}})$  gamma Poisson, dada por <sup>3</sup>:

$$f(T_u|\{T_i\}_{i \in \{i_1, \dots, i_k\}}) = Gp(T_u|\alpha + T_s, \beta + k, K - k), \quad (5.20)$$

donde  $\mathbf{T}_u = (T_i : i \notin \{i_1, \dots, i_k\})^t$ ,  $T_s = \sum_{i \in \{i_1, \dots, i_k\}} T_i$ ,  $T_u = \sum_{i \notin \{i_1, \dots, i_k\}} T_i$ , y

$$\begin{aligned} \tilde{\theta}_i &= E(\theta_i|(s, \mathbf{y}_s), \mu) = \frac{c_i n_i \bar{y}_{s_i} + \mu}{c_i n_i + 1}, \quad i \in \{i_1, \dots, i_k\}, \\ \mathbf{A}_i &= \mathbf{I}_{T_i - n_i} + \frac{c_i \mathbf{1}_{T_i - n_i} \mathbf{1}_{T_i - n_i}^t}{c_i n_i + 1}, \quad i \in \{i_1, \dots, i_k\}, \\ \mathbf{B}_i &= \mathbf{I}_{T_i} + c_i \mathbf{1}_{T_i} \mathbf{1}_{T_i}^t, \quad i \notin \{i_1, \dots, i_k\}, \\ \mathbf{Z}_i &= \left[ 1 + \frac{1}{2b_0} \sum_{j \in s_i} (y_{ij} - \bar{y}_{s_i})^2 \right] \frac{n_i c_i + 1}{n_i}, \quad i \in \{i_1, \dots, i_k\}, \\ a_i &= a_0 + \frac{n_i}{2}, \quad i \in \{i_1, \dots, i_k\}, \\ b_i &= b_0 + \frac{1}{2} (\mathbf{y}_{s_i} - \mathbf{1}_{n_i} \mu)^t (\mathbf{I}_{n_i} + \mathbf{1}_{n_i} \mathbf{1}_{n_i}^t c_i)^{-1} (\mathbf{y}_{s_i} - \mathbf{1}_{n_i} \mu), \quad i \in \{i_1, \dots, i_k\}. \end{aligned} \quad (5.21)$$

**Demostración.** La demostración de este teorema se apoya en la aplicación de los lemas B.5, B.8, B.11 y B.12, y será dividida en varios pasos señalando a lo largo de su desarrollo cuando quedan demostrados los apartados en los que se divide el presente teorema:

1.- A partir de la distribución de  $\mathbf{Y}|\theta, \sigma^2, \mathbf{T}$ , aplicando propiedades de la distribución normal básicas tenemos que:

$$\mathbf{Y}_{s_i}|\theta, \sigma^2, \mathbf{T} \sim N_{n_i}(\mathbf{1}_{n_i} \theta_i, \mathbf{I}_{n_i} \sigma_i^2), \text{ independientes, } i \in \{i_1, \dots, i_k\}. \quad (5.22)$$

2.- A partir de (5.22) y la distribución de  $\theta|\mu, \sigma^2$ , dada en (5.1) se obtiene aplicando el resultado (a) del lema B.8, la distribución de  $\mathbf{Y}_s|\mu, \sigma^2, \mathbf{T}$ , que es:

$$\mathbf{Y}_s|\mu, \sigma^2, \mathbf{T} \sim N_n(\mathbf{1}_n \mu, \text{diag}_{i \in \{i_1, \dots, i_k\}} \{(\mathbf{I}_{n_i} + \mathbf{1}_{n_i} \mathbf{1}_{n_i}^t c_i) \sigma_i^2\}), \quad (5.23)$$

y aplicando el resultado (b) del lema B.8, la distribución de  $\theta|(s, \mathbf{y}_s), \mu, \sigma^2$ , según la cual  $\theta_i|(s, \mathbf{y}_s), \mu, \sigma^2$ , son independientes para  $i = 1, 2, \dots, K$  con distribución dada por:

<sup>3</sup>notemos que esta distribución se obtiene a partir de:  $T_u|\lambda, \{T_i\}_{i \in \{i_1, \dots, i_k\}} \sim Po((K - k)\lambda)$  y  $\lambda|\{T_i\}_{i \in \{i_1, \dots, i_k\}} \sim Ga(\alpha + T_s, \beta + k)$

$$\begin{aligned}\theta_i|(s, \mathbf{y}_s), \mu, \sigma^2 &\sim N(\tilde{\theta}_i, \frac{c_i}{n_i c_i + 1} \sigma_i^2), \quad i \in \{i_1, \dots, i_k\}, \\ &\sim N(\mu, c_i \sigma_i^2), \quad i \notin \{i_1, \dots, i_k\}.\end{aligned}\quad (5.24)$$

3.- A partir de (5.23) y la distribución inicial para  $\sigma^2$  dada en (5.1), se obtiene, aplicando el resultado (b) del lema B.11, la distribución de  $\sigma^2|(s, \mathbf{y}_s), \mu$  dada en (5.17) ( $\rightsquigarrow$  comprobación del resultado (d)), y aplicando el resultado (a) del lema B.11 la distribución de  $\mathbf{Y}_s|\mu, \mathbf{T}$ , dada por:

$$\mathbf{Y}_{s_i}|\mu, \mathbf{T} \sim t_{n_i}(\mathbf{1}_{n_i} \mu, \frac{b_0}{a_0} (\mathbf{I}_{n_i} + \mathbf{1}_{n_i} \mathbf{1}_{n_i}^t c_i), 2a_0), \text{ independientes, } i \in \{i_1, \dots, i_k\}.\quad (5.25)$$

4.- A partir de (5.25) y la distribución inicial de  $\mu$  dada en (5.1) se obtiene, aplicando el teorema de Bayes, la distribución de  $\mu|(s, \mathbf{y}_s), \mathbf{T}$  dada en (5.18), esta distribución es propia por el lema B.5 ( $\rightsquigarrow$  comprobación del resultado (e)).

5.- A partir de la distribución de  $\mathbf{Y}|\theta, \sigma^2, \mathbf{T}$ , aplicando propiedades de la distribución normal básicas tenemos que:

$$\mathbf{Y}_{u_i}|(s, \mathbf{y}_s), \theta, \sigma^2, \mathbf{T} \sim N_{T_i - n_i}(\mathbf{1}_{T_i - n_i} \theta_i, \mathbf{I}_{N_i - n_i} \sigma_i^2), \text{ independientes, } i = 1, 2, \dots, K.\quad (5.26)$$

6.- A partir de (5.26) y (5.24) se obtiene, aplicando el resultado (a) del lema B.8, la distribución de  $\mathbf{Y}_{u_i}|(s, \mathbf{y}_s), \mu, \sigma^2, \mathbf{T}$ , según la cual  $\mathbf{Y}_{u_i}|(s, \mathbf{y}_s), \mu, \sigma^2, \mathbf{T}$ , son independientes para  $i = 1, 2, \dots, K$ , con distribución dada por:

$$\begin{aligned}\mathbf{Y}_{u_i}|(s, \mathbf{y}_s), \mu, \sigma^2, \mathbf{T} &\sim N_{T_i - n_i}(\mathbf{1}_{T_i - n_i} \tilde{\theta}_i, (\mathbf{I}_{T_i - n_i} + \frac{\mathbf{1}_{T_i - n_i} \mathbf{1}_{T_i - n_i}^t c_i}{n_i c_i + 1}) \sigma_i^2), \quad i \in \{i_1, \dots, i_k\}, \\ &\sim N_{T_i}(\mathbf{1}_{T_i} \mu, (\mathbf{I}_{T_i} + \mathbf{1}_{T_i} \mathbf{1}_{T_i}^t c_i) \sigma_i^2), \quad i \notin \{i_1, \dots, i_k\},\end{aligned}\quad (5.27)$$

y aplicando las definiciones de media no observada por ruta y media no observada y que la combinación lineal de normales es normal, se obtienen las distribuciones de  $\bar{\mathbf{Y}}_{u_i}|(s, \mathbf{y}_s), \mu, \sigma^2, \mathbf{T}$ ,  $i = 1, 2, \dots, K$ , que son independientes y vienen dadas por:

$$\begin{aligned}\bar{\mathbf{Y}}_{u_i}|(s, \mathbf{y}_s), \mu, \sigma^2, \mathbf{T} &\sim N(\tilde{\theta}_i, [(T_i - n_i)^{-1} + \frac{c_i}{n_i c_i + 1}] \sigma_i^2), \quad i \in \{i_1, \dots, i_k\}, \\ &\sim N(\mu, (T_i^{-1} + c_i) \sigma_i^2), \quad i \notin \{i_1, \dots, i_k\},\end{aligned}\quad (5.28)$$

y la distribución de  $\bar{\mathbf{Y}}_{u_i}|(s, \mathbf{y}_s), \mu, \sigma^2, \mathbf{T}$ , dada en (5.16) ( $\rightsquigarrow$  comprobación del resultado (c)).

7.- A partir de (5.27) y el resultado (d) de este teorema (ya comprobado en el punto 3 de esta demostración), se obtiene la distribución de  $\mathbf{Y}_u|(s, \mathbf{y}_s), \mu, \mathbf{T}$  dada en (5.14) ( $\leadsto$  comprobación del resultado (a)) y a partir de (5.28) y el resultado (d) de este teorema, la distribución de  $\bar{Y}_{u_i}|(s, \mathbf{y}_s), \mu, \mathbf{T}, i = 1, 2, \dots, K$ , dada en (5.15) ( $\leadsto$  comprobación del resultado (b)). En ambos casos se aplica el resultado (a) del lema B.11.

8.- Para demostrar los resultados (f) y (g) del teorema, notemos en primer lugar, que el vector  $\mathbf{y}_s$  y el vector  $\mathbf{T}$  son independientes según el modelo definido en (5.1), por lo que  $f(\mathbf{T}_u|(s, \mathbf{y}_s), \{T_i\}_{i \in \{i_1, \dots, i_k\}}) = f(\mathbf{T}_u, \{T_i\}_{i \in \{i_1, \dots, i_k\}})$  y  $f(T_u|(s, \mathbf{y}_s), \{T_i\}_{i \in \{i_1, \dots, i_k\}}) = f(T_u|\{T_i\}_{i \in \{i_1, \dots, i_k\}})$ .

9.- A partir de la distribución de las variables  $T_i|\lambda, i \in \{i_1, \dots, i_k\}$  y la distribución inicial de  $\lambda$  dadas en (5.1) se obtiene, aplicando el resultado (c) del lema B.12, la distribución de  $\lambda|\{T_i\}_{i \in \{i_1, \dots, i_k\}}$ , que viene dada por:

$$\lambda|\{T_i\}_{i \in \{i_1, \dots, i_k\}} \sim Ga(\alpha + T_s, \beta + k). \quad (5.29)$$

10.- A partir de la distribución de  $T_j|\lambda, \{T_i\}_{i \in \{i_1, \dots, i_k\}}, j \notin \{i_1, \dots, i_k\}$  y de (5.29) se obtiene, aplicando el resultado (b) del lema B.12, la distribución de  $\mathbf{T}_u|(s, \mathbf{y}_s), \{T_i\}_{i \in \{i_1, \dots, i_k\}}$  dada en (5.19) ( $\leadsto$  comprobación del resultado (f)).

11.- A partir de la distribución de  $T_j|\lambda, \{T_i\}_{i \in \{i_1, \dots, i_k\}}, j \notin \{i_1, \dots, i_k\}$  y aplicando la propiedad de que la suma de Poisson independientes es Poisson, se obtiene que  $T_u|\lambda, \{T_i\}_{i \in \{i_1, \dots, i_k\}} \sim Po((K - k)\lambda)$ . Y aplicando sobre  $T_u|\lambda, \{T_i\}_{i \in \{i_1, \dots, i_k\}}$  y (5.29) el resultado (a) del lema B.12 se obtiene la distribución de  $T_u|(s, \mathbf{y}_s), \{T_i\}_{i \in \{i_1, \dots, i_k\}}$  dada en (5.20) ( $\leadsto$  comprobación del resultado (g)).

C.S.Q.D.

Señalemos que la integración en cualquiera de las distribuciones predictivas anteriores de alguno de sus parámetros da lugar a distribuciones no estándar, tal y como anunciábamos al principio de la sección. Ello no obstante, a partir de las distribuciones proporcionadas en el teorema anterior podemos establecer el siguiente resultado acerca de la existencia de la media y varianza de las distribuciones predictivas y finales de interés, (y por tanto de los estimadores Bayes y sus varianzas condicionadas a la muestra y los tamaños de las rutas muestreadas), que proporciona también una expresión para el estimador Bayes de la media poblacional y su varianza condicionada a la muestra y los tamaños de las rutas muestreadas.

**Teorema 5.3** Consideremos el modelo Bayes para rutas aleatorias definido en (5.1), entonces se tiene que:

(a) las distribuciones predictivas de  $\mathbf{Y}_u|(s, \mathbf{y}_s), \{T_i\}_{i \in \{i_1, \dots, i_k\}}, \bar{Y}_{u_i}|(s, \mathbf{y}_s), \{T_i\}_{i \in \{i_1, \dots, i_k\}}, i = 1, 2, \dots, K, \bar{Y}_u|(s, \mathbf{y}_s), \{T_i\}_{i \in \{i_1, \dots, i_k\}}$  y  $\bar{Y}|(s, \mathbf{y}_s), \{T_i\}_{i \in \{i_1, \dots, i_k\}}$ , son propias y bajo

la hipótesis<sup>4</sup>  $a_0 > 1$  tienen primeros dos momentos,

(b) las distribuciones finales de  $\sigma^2$ ,  $\theta$ ,  $\mu$  y  $\mathbf{T}_u$  condicionadas a  $(s, \mathbf{y}_s)$  y  $\{T_i\}_{i \in \{i_1, \dots, i_k\}}$ , son propias y tienen media y varianza bajo la hipótesis anterior ( $a_0 > 1$ ),

y (c) el estimador Bayes para la media poblacional y su varianza vienen dados, si  $a_0 > 1$ , por:

$$\begin{aligned} \widehat{Y} &= d_2 E \left[ \frac{1}{T_u + T_s} \right] + E(\mu|(s, \mathbf{y}_s)) E \left[ \frac{T_u}{T_u + T_s} \right], \\ \text{Var}(\overline{Y}|(s, \mathbf{y}_s), \{T_i\}_{j=1}^k) &= (d_1^2 \text{Var}(\mu|(s, \mathbf{y}_s)) + d_3) E \left[ \frac{1}{(T_u + T_s)^2} \right] + \\ &+ 2[d_1 \text{Var}(\mu|(s, \mathbf{y}_s)) + d_2 E(\mu|(s, \mathbf{y}_s))] E \left[ \frac{T_u}{(T_u + T_s)^2} \right] + \\ &+ E(\mu^2|(s, \mathbf{y}_s)) E \left[ \frac{T_u^2}{(T_u + T_s)^2} \right] + \\ &+ d_4 \sum_{i \in \{i_1, \dots, i_k\}} \left\{ E \left[ \frac{T_i}{(T_u + T_s)^2} \right] + c_i E \left[ \frac{T_i^2}{(T_u + T_s)^2} \right] \right\} - \\ &- \left\{ d_2 E \left[ \frac{1}{T_u + T_s} \right] + E(\mu|(s, \mathbf{y}_s)) E \left[ \frac{T_u}{T_u + T_s} \right] \right\}^2 \quad (5.30) \end{aligned}$$

donde  $E(\cdot)$  es el operador esperanza respecto a  $\mathbf{T}_u | \{T_i\}_{i \in s}$ , cuya distribución está definida por (5.19),

$$\begin{aligned} d_1 &= \sum_{i \in \{i_1, \dots, i_k\}} \frac{T_i - n_i}{c_i n_i + 1}, \\ d_2 &= n \bar{y}_s + \sum_{i \in \{i_1, \dots, i_k\}} (T_i - n_i) \frac{c_i n_i \bar{y}_{s_i}}{c_i n_i + 1} + d_1 E(\mu|(s, \mathbf{y}_s)), \\ d_3 &= \sum_{i \in \{i_1, \dots, i_k\}} \frac{T_i - n_i}{a_i - 1} \left( 1 + \frac{(N_i - n_i) c_i}{c_i n_i + 1} \right) E^{\mu|(s, \mathbf{y}_s)}(b_i) + d_2^2, \\ d_4 &= \frac{b_0}{a_0 - 1}, \quad (5.31) \end{aligned}$$

y  $\mathbf{T}_u$ ,  $T_u$ ,  $T_s$ ,  $a_i$ ,  $b_i$ ,  $i \in \{i_1, \dots, i_k\}$  son el vector y escalares definidos en (5.21).

**Demostración.** La demostración de estos resultados se apoya en el teorema 5.2 y los lemas B.5 y B.6.

<sup>4</sup>en el caso de la distribución de  $\overline{Y}_{u_i}|(s, \mathbf{y}_s), \{T_i\}_{i \in \{i_1, \dots, i_k\}}$ ,  $i \in \{i_1, \dots, i_k\}$  la hipótesis  $a_0 > 1$  para existencia de los dos primeros momentos puede sustituirse por la condición  $n_i \geq 3$



\* Para comprobar el resultado (a) notemos que las distribuciones  $\mathbf{Y}_u|(s, \mathbf{y}_s), \{T_i\}_{i \in \{i_1, \dots, i_k\}}, \bar{Y}_{u_i}|(s, \mathbf{y}_s), \{T_i\}_{i \in \{i_1, \dots, i_k\}}, i = 1, 2, \dots, K$  y  $\bar{Y}_u|(s, \mathbf{y}_s), \{T_i\}_{i \in \{i_1, \dots, i_k\}}$ , son propias y tienen primeros dos momentos para  $a_0 > 1$  por ser las condicionales necesarias para su obtención (para  $\mathbf{Y}_u$  las distribuciones (5.14), (5.18) y (5.19), para  $\bar{Y}_{u_i}$  las distribuciones (5.15), (5.18) y (5.19), y para  $\bar{Y}_u$  las distribuciones (5.16), (5.17), (5.18) y (5.19)) propias y tener primeros dos momentos bajo la hipótesis<sup>5</sup>  $a_0 > 1$ . Esta última afirmación se apoya en las propiedades conocidas de las distribuciones normales, t y gamma, en el estudio de las distribuciones (5.18) y (5.19) a partir de los lemas B.5 y B.6 respectivamente y en las definiciones de  $a_i$  y  $b_i$  dadas en (5.21).

Respecto a la distribución y primeros dos momentos de  $\bar{Y}|(s, \mathbf{y}_s), \{T_i\}_{i \in \{i_1, \dots, i_k\}}$ , notemos que por propiedades básicas de las distribuciones normales y teniendo en cuenta que  $\bar{Y} = \frac{n}{T}\bar{y}_s + \frac{T-n}{T}\bar{Y}_u$ , se tiene que la distribución de  $\bar{Y}|(s, \mathbf{y}_s), \sigma^2, \mu, \mathbf{T}$  es normal con media y varianzas dadas respectivamente por:

$$\begin{aligned} E(\bar{Y}|(s, \mathbf{y}_s), \sigma^2, \mu, \mathbf{T}) &= \frac{n}{T}\bar{y}_s + \frac{T-n}{T}E(\bar{Y}_u|(s, \mathbf{y}_s), \sigma^2, \mu, \mathbf{T}), \\ \text{Var}(\bar{Y}|(s, \mathbf{y}_s), \sigma^2, \mu, \mathbf{T}) &= \left(\frac{T-n}{T}\right)^2 \text{Var}(\bar{Y}_u|(s, \mathbf{y}_s), \sigma^2, \mu, \mathbf{T}) \end{aligned}$$

con  $E(\bar{Y}_u|(s, \mathbf{y}_s), \sigma^2, \mu, \mathbf{T})$  y  $\text{Var}(\bar{Y}_u|(s, \mathbf{y}_s), \sigma^2, \mu, \mathbf{T})$ , definidas en (5.16). A partir de esta distribución predictiva condicional y las distribuciones dadas por (5.17), (5.18) y (5.19), y repitiendo las reflexiones anteriores se comprueba que la distribución de  $\bar{Y}|(s, \mathbf{y}_s), \{T_i\}_{i \in \{i_1, \dots, i_k\}}$ , es propia y tiene primeros dos momentos bajo la hipótesis  $a_0 > 1$ .

\* El resultado (b) se deriva de los resultados del teorema 5.2. Así:

- la distribución final para  $\sigma^2$  es propia y tiene primeros dos momentos bajo la hipótesis anterior por ser las distribuciones dadas en (5.17) y (5.18) propias y tener primeros dos momentos bajo esa hipótesis (lo que se deduce de las propiedades de la distribución gamma y de la aplicación del lema B.5),
- la distribución final de  $\theta$  es propia y tiene primeros dos momentos bajo la hipótesis anterior por ser las distribuciones dadas en (5.24), (5.17) y (5.18) propias y tener primeros dos momentos bajo esa hipótesis,
- la distribución predictiva para  $\mu$ , viene dada en (5.18) y es propia y tiene primeros dos momentos por aplicación inmediata del lema B.5, y
- la distribución predictiva para  $\mathbf{T}_u$ , viene dada en (5.19) y es propia y tiene primeros dos momentos por aplicación inmediata del lema B.6.

\* Y el resultado (c) se obtiene a partir de las distribuciones dadas en (5.16), (5.17), (5.18) y (5.19) y la definición de estimador Bayes, aplicando que<sup>6</sup>:

<sup>5</sup> esta hipótesis puede debilitarse en el caso de la distribución predictiva de la media no observada en ruta muestradas por la hipótesis  $n_i \geq 3$

<sup>6</sup> la existencia del estimador Bayes de la media poblacional y de su varianza condicionada a la muestra y los tamaños de las rutas muestradas se deduce de la definición de estimador Bayes para función de pérdida cuadrática y el resultado (a) de este teorema

$$\begin{aligned}\widehat{Y} = E(\bar{Y}|(s, \mathbf{y}_s), \{T_{i_j}\}_{j=1}^k) &= E^{\mathbf{T}_u|\{T_{i_j}\}_{j=1}^k} E(\bar{Y}|(s, \mathbf{y}_s), \mathbf{T}) = \\ &= n\bar{y}_s E^{\mathbf{T}_u|\{T_{i_j}\}_{j=1}^k} \left[ \frac{1}{T_u+T_s} \right] + \\ &+ E^{\mathbf{T}_u|\{T_{i_j}\}_{j=1}^k} \left[ \frac{T_u+T_s-n}{T_u+T_s} E(\bar{Y}_u|(s, \mathbf{y}_s), \mathbf{T}) \right]; \\ \\ \text{Var}(\bar{Y}|(s, \mathbf{y}_s), \{T_{i_j}\}_{j=1}^k) &= E^{\mathbf{T}_u|\{T_{i_j}\}_{j=1}^k} \text{Var}(\bar{Y}|(s, \mathbf{y}_s), \mathbf{T}) + \\ &+ \text{Var}^{\mathbf{T}_u|\{T_{i_j}\}_{j=1}^k} E(\bar{Y}|(s, \mathbf{y}_s), \mathbf{T}) = \\ &= E^{\mathbf{T}_u|\{T_{i_j}\}_{j=1}^k} \left[ \left( \frac{T_u+T_s-n}{T_u+T_s} \right)^2 \text{Var}(\bar{Y}_u|(s, \mathbf{y}_s), \mathbf{T}) \right] + \\ &+ \text{Var}^{\mathbf{T}_u|\{T_{i_j}\}_{j=1}^k} \left[ \frac{n\bar{y}_s}{T_u+T_s} + \frac{T_u+T_s-n}{T_u+T_s} E(\bar{Y}_u|(s, \mathbf{y}_s), \mathbf{T}) \right],\end{aligned}$$

donde:

$$\begin{aligned}E(\bar{Y}_u|(s, \mathbf{y}_s), \mathbf{T}) &= E^{\mu|(s, \mathbf{y}_s)} E^{\sigma^2|(s, \mathbf{y}_s), \mu} E(\bar{Y}_u|(s, \mathbf{y}_s), \sigma^2, \mu, \mathbf{T}) = \\ &= \frac{d_2 - n\bar{y}_s + T_u E(\mu|(s, \mathbf{y}_s))}{T_u + T_s - n}, \\ \\ \text{Var}(\bar{Y}_u|(s, \mathbf{y}_s), \mathbf{T}) &= E^{\mu|(s, \mathbf{y}_s)} \left[ E^{\sigma^2|(s, \mathbf{y}_s), \mu} \text{Var}(\bar{Y}_u|(s, \mathbf{y}_s), \sigma^2, \mu, \mathbf{T}) + \right. \\ &\quad \left. + \text{Var}^{\sigma^2|(s, \mathbf{y}_s), \mu} E(\bar{Y}_u|(s, \mathbf{y}_s), \sigma^2, \mu, \mathbf{T}) \right] + \\ &\quad + \text{Var}^{\mu|(s, \mathbf{y}_s)} E^{\sigma^2|(s, \mathbf{y}_s), \mu} E(\bar{Y}_u|(s, \mathbf{y}_s), \sigma^2, \mu, \mathbf{T}) = \\ &= \frac{d_3 - d_2^2 + d_4 \sum_{i \in \{i_1, \dots, i_k\}} T_i(1+T_i c_i) + (d_1 + T_u)^2 \text{Var}(\mu|(s, \mathbf{y}_s))}{(T_u + T_s - n)^2}.\end{aligned}$$

**C.S.Q.D.**

Este resultado proporciona las hipótesis suficientes para garantizar que las distribuciones predictivas del vector poblacional no observado, de la media no observada de cada ruta, de la media poblacional no observada y de la media poblacional sean propias y posean primeros dos momentos (lo que permite asegurar la existencia de los estimadores Bayes y varianza condicionada a la muestra para cualquier función lineal del vector poblacional) y una expresión del estimador Bayes y varianza condicionada a la muestra de la media poblacional (expresiones similares pueden obtenerse para los estimadores Bayes y varianza asociada para la media poblacional por ruta), que requiere del cálculo por métodos numéricos (dado que las distribuciones correspondientes no son estándar) de los dos primeros momentos de la distribución de  $\mu|(s, \mathbf{y}_s)$  y los valores esperados para funciones de los tamaños de rutas no observados a partir de la distribución de  $\mathbf{T}_u|\{T_{i_j}\}_{i \in \{i_1, \dots, i_k\}}$  (ambas distribuciones obtenidas en el teorema 5.2).

El cálculo de estos momentos no es fácil, respecto a la obtención de los dos primeros momentos de la distribución de  $\mu|(s, \mathbf{y}_s)$ , el trabajo de Bayarri y Font(94)[5] proporciona para el caso  $a_0 = 1$  y  $p(\mu) \propto 1$  un mecanismo para el cálculo exacto del primer

momento, pero los cálculos necesarios son muy laboriosos y en este caso particular el segundo momento no existe. Y respecto al cálculo de los momentos relativos a funciones de los tamaños de rutas no muestreadas, son particularmente largos y complicados los cálculos de  $E\left[\frac{T_i}{(T_u+T_i)^2}\right]$ ,  $E\left[\frac{T_i^2}{(T_u+T_i)^2}\right]$ , para  $i \notin \{i_1, \dots, i_k\}$  que requieren la distribución multivariante del vector  $\mathbf{T}_u | \{T_i\}_{i \in \{i_1, \dots, i_k\}}$ . Estas dificultades de cálculo para la obtención del estimador de la media poblacional y de su varianza condicionada a la muestra aconsejan el uso del método de muestreo Gibbs para el cálculo de las distribuciones predictivas y estimadores Bayes para las medias no observadas en cada ruta y la media poblacional así como de sus respectivas varianzas condicionadas a la muestra.

A continuación estudiaremos las distribuciones condicionales necesarias para un análisis del modelo apoyado en el muestreo por Gibbs.

**Teorema 5.4** *El modelo Bayes para rutas aleatorias definido en (5.1) es un modelo DAG (figura 5.1), con distribuciones condicionales dadas por:*

$$f((\bar{y}_{u_1}, \dots, \bar{y}_{u_K})^t | \mathbf{y}_s, \boldsymbol{\theta}, \boldsymbol{\sigma}^2, \mu, \{T_i\}_{i=1}^K, \lambda) = f((\bar{y}_{u_1}, \dots, \bar{y}_{u_K})^t | \boldsymbol{\theta}, \boldsymbol{\sigma}^2, \{T_i\}_{i=1}^K) = N_K \left( (\bar{y}_{u_1}, \dots, \bar{y}_{u_K})^t | \boldsymbol{\theta}, \begin{pmatrix} \text{diag}_{i=1}^k \{(T_i - n_i)^{-1} \sigma_i^2\} & \mathbf{0}_{k \times (K-k)} \\ \mathbf{0}_{(K-k) \times k} & \text{diag}_{i=k+1}^K \{T_i^{-1} \sigma_i^2\} \end{pmatrix} \right), \quad (5.32)$$

$$\pi(\boldsymbol{\theta} | \{\bar{y}_{u_i}\}_{i=1}^K, \mathbf{y}_s, \boldsymbol{\sigma}^2, \mu, \{T_i\}_{i=1}^K, \lambda) = \pi(\boldsymbol{\theta} | \{\bar{y}_{u_i}\}_{i=1}^K, \mathbf{y}_s, \boldsymbol{\sigma}^2, \mu, \{T_i\}_{i=1}^K) = N_K(\boldsymbol{\theta} | \bar{\boldsymbol{\mu}}, \text{diag}_{i=1}^K \{c_i(T_i c_i + 1)^{-1} \sigma_i^2\}), \quad (5.33)$$

$$\pi(\mu | \{\bar{y}_{u_i}\}_{i=1}^K, \mathbf{y}_s, \boldsymbol{\theta}, \boldsymbol{\sigma}^2, \{T_i\}_{i=1}^K, \lambda) = \pi(\mu | \boldsymbol{\theta}, \boldsymbol{\sigma}^2) = N(\mu | (\sum_{i=1}^K c_i^{-1} \sigma_i^{-2} + \sigma_0^{-2})^{-1} (\sum_{i=1}^K c_i^{-1} \sigma_i^{-2} \theta_i + \sigma_0^{-2} \mu_0), (\sum_{i=1}^K c_i^{-1} \sigma_i^{-2} + \sigma_0^{-2})^{-1}), \quad (5.34)$$

$$\pi(\boldsymbol{\sigma}^2 | \{\bar{y}_{u_i}\}_{i=1}^K, \mathbf{y}_s, \boldsymbol{\theta}, \mu, \{T_i\}_{i=1}^K, \lambda) = \pi(\boldsymbol{\sigma}^2 | \{\bar{y}_{u_i}\}_{i=1}^K, \mathbf{y}_s, \boldsymbol{\theta}, \mu, \{T_i\}_{i=1}^K) = \prod_{i=1}^K \text{Ga}^{-1}(\sigma_i^2 | a_i^*, b_i^*), \quad (5.35)$$

$$f(\mathbf{T}_u | \{\bar{y}_{u_i}\}_{i=1}^K, \mathbf{y}_s, \boldsymbol{\theta}, \mu, \boldsymbol{\sigma}^2, \{T_i\}_{i=1}^k, \lambda) = f(\mathbf{T}_u | \{\bar{y}_{u_i}\}_{i \notin \{i_1, \dots, i_k\}}, \{\theta_i\}_{i \notin \{i_1, \dots, i_k\}}, \{\sigma_i^2\}_{i \notin \{i_1, \dots, i_k\}}, \lambda) \propto \prod_{i \notin \{i_1, \dots, i_k\}} \left[ T_i^{1/2} \exp \left\{ -\frac{1}{2\sigma_i^2} T_i (\bar{y}_{u_i} - \theta_i)^2 \right\} \frac{\lambda^{T_i}}{T_i!} \right], \quad (5.36)$$

$$\pi(\lambda | \{\bar{y}_{u_i}\}_{i=1}^K, \mathbf{y}_s, \boldsymbol{\theta}, \mu, \boldsymbol{\sigma}^2, \{T_i\}_{i=1}^K) = \pi(\lambda | \{T_i\}_{i=1}^K) = \text{Ga}(\lambda | \sum_{i=1}^K T_i + \alpha, K + \beta), \quad (5.37)$$

donde  $\bar{\boldsymbol{\mu}} = (\bar{\mu}_1, \bar{\mu}_2, \dots, \bar{\mu}_K)^t$ ,  $\mathbf{T}_u$  aparece definido en (5.21) y,

$$\begin{aligned}
\tilde{\mu}_i &= \frac{(T_i - n_i)c_i\bar{y}_{u_i} + n_i c_i \bar{y}_{s_i} + \mu}{T_i c_i + 1}, \quad i \in \{i_1, \dots, i_k\}, \\
&= \frac{T_i c_i \bar{y}_{u_i} + \mu}{T_i c_i + 1}, \quad i \notin \{i_1, \dots, i_k\}, \\
a_i^* &= a_0 + \frac{n_i}{2} + 1, \quad i \in \{i_1, \dots, i_k\}, \\
&= a_0 + 1, \quad i \notin \{i_1, \dots, i_k\}, \\
b_i^* &= b_0 + \frac{1}{2c_i} \left\{ (T_i - n_i)c_i(\bar{y}_{u_i} - \theta_i)^2 + c_i \sum_{j \in s_i} (y_{ij} - \theta_i)^2 + \right. \\
&\quad \left. + (\theta_i - \mu)^2 \right\}, \quad i \in \{i_1, \dots, i_k\}, \\
&= b_0 + \frac{1}{2c_i} \left\{ T_i c_i (\bar{y}_{u_i} - \theta_i)^2 + (\theta_i - \mu)^2 \right\}, \quad i \notin \{i_1, \dots, i_k\}.
\end{aligned} \tag{5.38}$$

Además, se cumple que, las distribuciones  $T_i | \{\bar{y}_{u_i}\}_1^K, \mathbf{y}_s^t, \boldsymbol{\theta}^t, \mu, \boldsymbol{\sigma}^2, \{T_i\}_{i=1}^k, \lambda, i \notin \{i_1, \dots, i_k\}$  son independientes, con funciones de probabilidad marginales log-concavas en los enteros positivos incluyendo el cero.

**Demostración.** De la definición 5.1 del modelo en estudio se deduce directamente la representación gráfica de la figura 5.1, que responde a un modelo DAG<sup>7</sup>. Las distribuciones condicionales se obtienen a partir de la distribución conjunta de  $(\bar{Y}_{u_1}, \dots, \bar{Y}_{u_K}, \mathbf{Y}_s^t, \boldsymbol{\theta}^t, (\boldsymbol{\sigma}^2)^t, \mu, \mathbf{T}^t)^t$ , que viene dada por:

$$\begin{aligned}
f((\bar{y}_{u_1}, \dots, \bar{y}_{u_K}, \mathbf{y}_s^t, \boldsymbol{\theta}^t, (\boldsymbol{\sigma}^2)^t, \mu, \mathbf{T}^t)^t) = \\
\prod_{i=1}^K N(\bar{y}_{u_i} | \theta_i, (T_i - n_i)^{-1} \sigma_i^2) \times \prod_{i \in \{i_1, \dots, i_k\}} \prod_{j \in s_i} N(y_{ij} | \theta_i, \sigma_i^2) \times \\
\times \prod_{i=1}^K N(\theta_i | \mu, c_i \sigma_i^2) \times N(\mu | \mu_0, \sigma_0^2) \times \prod_{i=1}^K Ga^{-1}(\sigma_i^2 | a_0, b_0) \times \\
\times \prod_{i=1}^K Po(T_i | \lambda) \times Ga(\lambda | \alpha, \beta),
\end{aligned}$$

separando los términos correspondientes a cada condicional (que contienen el vector o variable cuya condicional estamos hallando). Concretamente,

- $f((\bar{y}_{u_1}, \dots, \bar{y}_{u_K})^t | \mathbf{y}_s, \boldsymbol{\theta}, \boldsymbol{\sigma}^2, \mu, \{T_i\}_1^K, \lambda) = \prod_{i=1}^K N(\bar{y}_{u_i} | \theta_i, (T_i - n_i)^{-1} \sigma_i^2)$ , esto es el resultado dado en (5.32).
- $\pi(\boldsymbol{\theta} | \{\bar{y}_{u_i}\}_1^K, \mathbf{y}_s, \boldsymbol{\sigma}^2, \mu, \{T_i\}_1^K, \lambda) \propto \prod_{i=1}^K N(\theta_i | \mu, c_i \sigma_i^2) \times \prod_{i=1}^K N(\bar{y}_{u_i} | \theta_i, (T_i - n_i)^{-1} \sigma_i^2) \times \prod_{i \in \{i_1, \dots, i_k\}} \prod_{j \in s_i} N(y_{ij} | \theta_i, \sigma_i^2)$ , de donde se obtiene, completando cuadrados respecto a las  $\theta_i, i = 1, 2, \dots, K$ , la distribución dada en (5.33).
- $\pi(\mu | \{\bar{y}_{u_i}\}_1^K, \mathbf{y}_s, \boldsymbol{\theta}, \boldsymbol{\sigma}^2, \{T_i\}_1^K, \lambda) \propto \prod_{i=1}^K N(\theta_i | \mu, c_i \sigma_i^2) \times N(\mu | \mu_0, \sigma_0^2)$ , de donde se obtiene, completando cuadrados respecto a  $\mu$ , la distribución dada en (5.34).

<sup>7</sup>ver definición y propiedades de estos modelos en C.2.1 y C.2.2

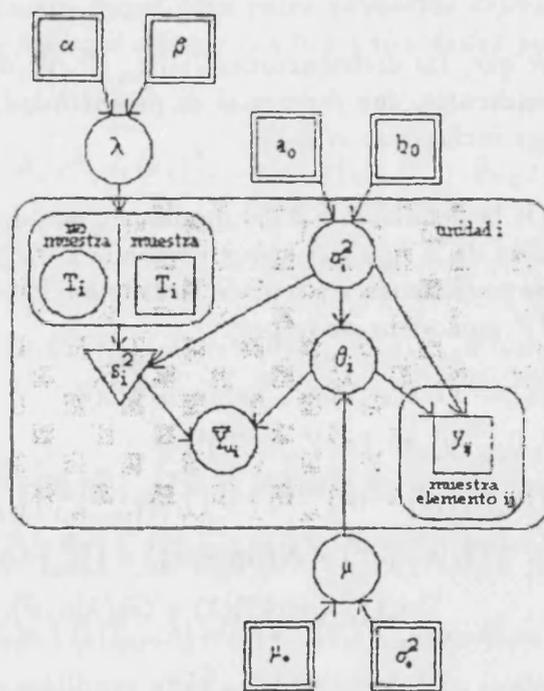


Figura 5.1: Representación gráfica del modelo Bayes para rutas aleatorias

- $\pi(\sigma^2 | \{\bar{y}_{u_i}\}_1^K, \mathbf{y}_s, \boldsymbol{\theta}, \mu, \{T_i\}_1^K, \lambda) \propto \prod_{i=1}^K N(\theta_i | \mu, c_i \sigma_i^2) \times \prod_{i=1}^K Ga^{-1}(\sigma_i^2 | a_0, b_0) \times \prod_{i=1}^K N(\bar{y}_{u_i} | \theta_i, (T_i - n_i)^{-1} \sigma_i^2) \times \prod_{i \in \{i_1, \dots, i_k\}} \prod_{j \in s_i} N(y_{ij} | \theta_i, \sigma_i^2)$ , de donde se obtiene, juntando los términos correspondientes a potencias y a funciones exponenciales de  $\sigma_i^2$ ,  $i = 1, 2, \dots, K$ , la distribución dada en (5.35).
- $f(\mathbf{T}_u | \{\bar{y}_{u_i}\}_1^K, \mathbf{y}_s, \boldsymbol{\theta}, \mu, \sigma^2, \{T_i\}_1^K, \lambda) \propto \prod_{i \notin \{i_1, \dots, i_k\}} N(\bar{y}_{u_i} | \theta_i, (T_i - n_i)^{-1} \sigma_i^2) \times \prod_{i \notin \{i_1, \dots, i_k\}} Po(T_i | \lambda)$ , esto es, la distribución dada en (5.36). De esta distribución se deduce que las variables condicionadas  $T_i$ ,  $i \notin \{i_1, \dots, i_k\}$  son independientes con distribuciones con funciones de probabilidad log-concavas por aplicación inmediata del lema B.2.
  - $\pi(\lambda | \{\bar{y}_{u_i}\}_1^K, \mathbf{y}_s, \boldsymbol{\theta}, \mu, \sigma^2, \{T_i\}_1^K) \propto \prod_{i=1}^K Po(T_i | \lambda) \times Ga(\lambda | \alpha, \beta)$ , de donde se deduce, juntando los términos correspondientes a potencia y a funciones exponenciales de  $\lambda$ , la distribución dada en (5.37).

C.S.Q.D.

El resultado anterior cierra el análisis necesario para este modelo al proporcionar las herramientas necesarias para la realización de inferencias a partir del método de muestreo por Gibbs; notemos que las distribuciones condicionales necesarias para su aplicación son fáciles de simular dado que todas ellas tienen forma estándar, salvo la condicional de  $\mathbf{T}_u$  que es producto de distribuciones independientes con funciones de probabilidad log-concavas (las distribuciones estándar se pueden simular aplicando algoritmos específicos, ver por ejemplo Devroye(86)[25] y Ripley(87)[84], y la distribución condicional de  $\mathbf{T}_u$  al ser producto de distribuciones con funciones de probabilidad log-concavas, simulando cada una de sus componentes por un algoritmo de aceptación/rechazo como el estudiado por Gilks y Wild(92)[47] y Wild y Gilks(93)[119]).

El algoritmo Gibbs Sampling para este modelo puede implementarse mediante el programa BUGS (ver C.2.1), puesto que se cumplen los requisitos para modelos implementables por este programa por la definición (5.1) del modelo y el teorema anterior.

### 5.3.2 Modelo Bayes de rutas aleatorias con medias iguales

Dedicaremos esta subsección al estudio del modelo Bayes de rutas aleatorias con medias iguales, para ello siguiendo un esquema de resultados similar al de la subsección anterior realizaremos en primer lugar (primeros dos teoremas) el estudio de las distribuciones condicionales necesarias para realizar inferencia en poblaciones finitas a partir de este modelo y las hipótesis que garantizan que las distribuciones predictivas del vector poblacional no observado, la media no observada en cada ruta, la media poblacional no observada y la media poblacional tengan media y varianza. Y en segundo lugar el estudio de las distribuciones condicionales necesarias para aplicar el método de muestreo por Gibbs Sampling.

**Teorema 5.5** *Consideremos el modelo Bayes para rutas aleatorias con medias iguales definido en (5.2), entonces se tiene que:*

(a)  $\mathbf{Y}_{u_i}|(s, \mathbf{y}_s), \mu, \mathbf{T}$  son independientes para  $i = 1, 2, \dots, K$  con distribuciones predictivas condicionales dadas por:

$$\begin{aligned} f(\mathbf{y}_{u_i}|(s, \mathbf{y}_s), \mu, \{T_i\}_1^K) &= t_{T_i - n_i}(\mathbf{y}_{u_i} | \mathbf{1}_{T_i - n_i}, \mu, \frac{b_i}{a_i} \mathbf{I}_{T_i - n_i}, 2a_i), \quad i \in \{i_1, \dots, i_k\} \\ &= t_{T_i}(\mathbf{y}_{u_i} | \mathbf{1}_{T_i}, \mu, \frac{b_0}{a_0} \mathbf{I}_{T_i}, 2a_0), \quad i \notin \{i_1, \dots, i_k\}, \end{aligned} \quad (5.39)$$

(b)  $\bar{Y}_{u_i}|(s, \mathbf{y}_s), \mu, \mathbf{T}$  son independientes para  $i = 1, 2, \dots, K$  con distribuciones predictivas condicionales dadas por:

$$\begin{aligned} f(\bar{y}_{u_i}|(s, \mathbf{y}_s), \mu, \{T_i\}_1^K) &= t(\bar{y}_{u_i} | \mu, (T_i - n_i)^{-1} \frac{b_i}{a_i}, 2a_i), \quad i \in \{i_1, \dots, i_k\}, \\ &= t(\bar{y}_{u_i} | \mu, T_i^{-1} \frac{b_0}{a_0}, 2a_0), \quad i \notin \{i_1, \dots, i_k\}, \end{aligned} \quad (5.40)$$

(c)  $\bar{Y}_{u_i}|(s, \mathbf{y}_s), \sigma^2, \mu, \mathbf{T}$  se distribuye según una normal con media y varianza dadas respectivamente por:

$$\begin{aligned} E(\bar{Y}_{u_i}|(s, \mathbf{y}_s), \sigma^2, \mu, \mathbf{T}) &= \mu, \quad (5.41) \\ \text{Var}(\bar{Y}_{u_i}|(s, \mathbf{y}_s), \sigma^2, \mu, \mathbf{T}) &= \frac{1}{(T_u + T_s - n)^2} \left[ \sum_{i \in \{i_1, \dots, i_k\}} (T_i - n_i) \sigma_i^2 + \sum_{i \notin \{i_1, \dots, i_k\}} T_i \sigma_i^2 \right] \end{aligned}$$

(d)  $\sigma_i^2|(s, \mathbf{y}_s), \mu, \mathbf{T}$  son independientes para  $i = 1, 2, \dots, K$  con distribuciones condicionales  $\pi(\sigma_i^2|(s, \mathbf{y}_s), \mu, \mathbf{T}) = \pi(\sigma_i^2|(s, \mathbf{y}_s), \mu)$  dadas por:

$$\begin{aligned} \pi(\sigma_i^2|(s, \mathbf{y}_s), \mu) &= Ga^{-1}(\sigma_i^2 | a_i, b_i), \quad i \in \{i_1, \dots, i_k\}, \\ &= Ga^{-1}(\sigma_i^2 | a_0, b_0), \quad i \notin \{i_1, \dots, i_k\}, \end{aligned} \quad (5.42)$$

(e)  $\mu|(s, \mathbf{y}_s), \{T_i\}_1^K$  se distribuye con función de densidad  $\pi(\mu|(s, \mathbf{y}_s), \mathbf{T}) = \pi(\mu|(s, \mathbf{y}_s))$  dada por:

$$\pi(\mu|(s, \mathbf{y}_s)) \propto \prod_{i \in \{i_1, \dots, i_k\}} \left[ 1 + \frac{1}{2b_0 Z_i} (\mu - \bar{y}_{s_i})^2 \right]^{-a_i} \times \exp \left\{ -\frac{1}{2\sigma_0^2} (\mu - \mu_0)^2 \right\}, \quad (5.43)$$

(f)  $\mathbf{T}_u|(s, \mathbf{y}_s), \{T_i\}_{i \in \{i_1, \dots, i_k\}}$  se distribuye con función de densidad  $f(\mathbf{T}_u|(s, \mathbf{y}_s), \{T_i\}_{i \in \{i_1, \dots, i_k\}}) = f(\mathbf{T}_u|\{T_i\}_{i \in \{i_1, \dots, i_k\}})$  dada por:

$$f(\mathbf{T}_u | \{T_i\}_{i \in \{i_1, \dots, i_k\}}) = \frac{(\beta + k)^{T_s + \alpha}}{\Gamma(T_s + \alpha)} \frac{\Gamma(T_u + T_s + \alpha)}{(\beta + K)^{T_u + T_s + \alpha}} \prod_{i \notin \{i_1, \dots, i_k\}} \frac{1}{T_i!}, \quad (5.44)$$

(g)  $T_u | (s, \mathbf{y}_s), \{T_i\}_{i \in \{i_1, \dots, i_k\}}$  se distribuye según una función de densidad,  $f(T_u | (s, \mathbf{y}_s), \{T_i\}_{i \in \{i_1, \dots, i_k\}}) = f(T_u | \{T_i\}_{i \in \{i_1, \dots, i_k\}})$ , gamma Poisson,

$$f(T_u | \{T_i\}_{i \in \{i_1, \dots, i_k\}}) = Gp(T_u | \alpha + T_s, \beta + k, K - k) \quad (5.45)$$

donde  $\mathbf{T}_u = (T_i : i \notin \{i_1, \dots, i_k\})^t$ ,  $T_s = \sum_{i \in s} N_i$ ,  $T_u = \sum_{i \notin \{i_1, \dots, i_k\}} T_i$ , y

$$\begin{aligned} Z_i &= \left[ 1 + \frac{1}{2b_0} \sum_{j \in s_i} (y_{ij} - \bar{y}_{s_i})^2 \right] \frac{1}{n_i}, \quad i \in \{i_1, \dots, i_k\}, \\ a_i &= a_0 + \frac{n_i}{2}, \quad i \in \{i_1, \dots, i_k\}, \\ b_i &= b_0 + \frac{1}{2} \sum_{j \in s_i} (y_{ij} - \mu)^2, \quad i \in \{i_1, \dots, i_k\}. \end{aligned} \quad (5.46)$$

**Demostración.** La demostración de este teorema se apoya en la aplicación de los lemas B.8, B.5 y B.11, y será dividida en varios pasos señalando a lo largo de su desarrollo cuando quedan demostrados los apartados en los que se divide el presente teorema:

1.- A partir de la distribución de  $\mathbf{Y} | \mu, \sigma^2, \mathbf{T}$ , aplicando propiedades de la distribución normal básicas tenemos que:

$$\mathbf{Y}_{s_i} | \theta, \sigma^2, \mathbf{T} \sim N_{n_i}(\mathbf{1}_{n_i} \mu, \mathbf{I}_{n_i} \sigma_i^2), \text{ independientes, } i \in \{i_1, \dots, i_k\}. \quad (5.47)$$

2.- A partir de (5.47) y la distribución inicial para  $\sigma^2$  dada en (5.2), se obtiene, aplicando el resultado (b) del lema B.11, la distribución de  $\sigma^2 | (s, \mathbf{y}_s), \mu$  dada en (5.42) ( $\leadsto$  comprobación del resultado (d)), y aplicando el resultado (a) del lema B.11 la distribución de  $\mathbf{Y}_s | \mu, \mathbf{T}$ , dada por:

$$\mathbf{Y}_{s_i} | \mu, \mathbf{T} \sim t_{n_i}(\mathbf{1}_{n_i} \mu, \frac{b_0}{a_0} \mathbf{I}_{n_i}, 2a_0), \text{ independientes, } i \in \{i_1, \dots, i_k\}. \quad (5.48)$$

3.- A partir de (5.48) y la distribución inicial de  $\mu$  dada en (5.2) se obtiene, aplicando el teorema de Bayes, la distribución de  $\mu | (s, \mathbf{y}_s), \mathbf{T}$  dada en (5.43), esta distribución es propia por el lema B.5 ( $\leadsto$  comprobación del resultado (e)).

4.- A partir de la distribución de  $\mathbf{Y} | \mu, \sigma^2, \mathbf{T}$ , aplicando propiedades de la distribución normal básicas, tenemos que:

$$\mathbf{Y}_{u_i}|(s, \mathbf{y}_s), \mu, \sigma^2, \mathbf{T} \sim N_{T_i - n_i}(\mathbf{1}_{T_i - n_i} \mu, \mathbf{I}_{N_i - n_i} \sigma_i^2), \text{ independientes, } i = 1, 2, \dots, K, \quad (5.49)$$

y aplicando las definiciones de media no observada por ruta y media no observada y que la combinación lineal de normales es normal, se obtienen las distribuciones de  $\bar{Y}_{u_i}|(s, \mathbf{y}_s), \mu, \sigma^2, \mathbf{T}$ ,  $i = 1, 2, \dots, K$ , que son independientes y vienen dadas por:

$$\begin{aligned} \bar{Y}_{u_i}|(s, \mathbf{y}_s), \mu, \sigma^2, \mathbf{T} &\sim N(\mu, (T_i - n_i)^{-1} \sigma_i^2), \quad i \in \{i_1, \dots, i_k\}, \\ &\sim N(\mu, T_i^{-1} \sigma_i^2), \quad i \notin \{i_1, \dots, i_k\}, \end{aligned} \quad (5.50)$$

y la distribución de  $\bar{Y}_u|(s, \mathbf{y}_s), \mu, \sigma^2, \mathbf{T}$ , dada en (5.41) ( $\leadsto$  comprobación del resultado (c)).

5.- A partir de (5.49) y el resultado (d) de este teorema (ya comprobado en el punto 2 de esta demostración), se obtiene la distribución de  $\mathbf{Y}_u|(s, \mathbf{y}_s), \mu, \mathbf{T}$  dada en (5.39) ( $\leadsto$  comprobación del resultado (a)) y a partir de (5.50) y el resultado (d) de este teorema, la distribución de  $\bar{Y}_{u_i}|(s, \mathbf{y}_s), \mu, \mathbf{T}$ ,  $i = 1, 2, \dots, K$ , dada en (5.40) ( $\leadsto$  comprobación del resultado (b)). En ambos casos se aplica el resultado (a) del lema B.11.

6.- La demostración de los resultados (f) y (g) del teorema es idéntica a la demostración de los apartados (f) y (g) del teorema 5.2 (pasos 8 a 11).

C.S.Q.D.

**Teorema 5.6** Consideremos el modelo Bayes para rutas aleatorias con medias iguales definido en (5.2), entonces se tiene que:

(a) las distribuciones predictivas de  $\mathbf{Y}_u|(s, \mathbf{y}_s), \{T_i\}_{i \in \{i_1, \dots, i_k\}}$ ,  $\bar{Y}_{u_i}|(s, \mathbf{y}_s), \{T_i\}_{i \in \{i_1, \dots, i_k\}}$ ,  $i = 1, 2, \dots, K$ ,  $\bar{Y}_u|(s, \mathbf{y}_s), \{T_i\}_{i \in \{i_1, \dots, i_k\}}$  y  $\bar{Y}|(s, \mathbf{y}_s), \{T_i\}_{i \in \{i_1, \dots, i_k\}}$ , son propias y bajo la hipótesis<sup>8</sup>  $a_0 > 1$  tienen primeros dos momentos,

(b) las distribuciones finales de  $\sigma^2$ ,  $\mu$  y  $\mathbf{T}_u$  condicionadas a  $(s, \mathbf{y}_s)$  y  $\{T_i\}_{i \in \{i_1, \dots, i_k\}}$ , son propias y tienen primeros media y varianza bajo la hipótesis anterior ( $a_0 > 1$ ),

y (c) el estimador Bayes para la media poblacional y su varianza vienen dados, si  $a_0 > 1$ , por:

$$\begin{aligned} \hat{Y} &= d_2 E \left[ \frac{1}{T_u + T_s} \right] + E(\mu|(s, \mathbf{y}_s)) E \left[ \frac{T_u}{T_u + T_s} \right], \\ \text{Var}(\bar{Y}|(s, \mathbf{y}_s), \{T_i\}_{i \in \{i_1, \dots, i_k\}}) &= (d_1^2 \text{Var}(\mu|(s, \mathbf{y}_s)) + d_3) E \left[ \frac{1}{(T_u + T_s)^2} \right] + \end{aligned}$$

<sup>8</sup>en el caso de la distribución de  $\bar{Y}_{u_i}|(s, \mathbf{y}_s), \{T_i\}_{i \in \{i_1, \dots, i_k\}}$ ,  $i \in \{i_1, \dots, i_k\}$  la hipótesis  $a_0 > 1$  para existencia de los dos primeros momentos puede sustituirse por la condición  $n_i \geq 3$

$$\begin{aligned}
& +2[d_1 \text{Var}(\mu|(s, \mathbf{y}_s)) + d_2 E(\mu|(s, \mathbf{y}_s))] E \left[ \frac{T_u}{(T_u + T_s)^2} \right] + \\
& + d_4 E \left[ \frac{T_u}{(T_u + T_s)^2} \right] + E(\mu^2|(s, \mathbf{y}_s)) E \left[ \frac{T_u^2}{(T_u + T_s)^2} \right] - \\
& - \left\{ d_2 E \left[ \frac{1}{T_u + T_s} \right] + E(\mu|(s, \mathbf{y}_s)) E \left[ \frac{T_u}{T_u + T_s} \right] \right\}^2 \quad (5.51)
\end{aligned}$$

donde  $E(\cdot)$  es el operador esperanza respecto a  $T_u | \{T_i\}_{i \in \{i_1, \dots, i_k\}}$ , cuya distribución está definida por (5.45),

$$\begin{aligned}
d_1 &= T_s - n, \\
d_2 &= n\bar{y}_s + d_1 E(\mu|(s, \mathbf{y}_s)), \\
d_3 &= \sum_{i \in \{i_1, \dots, i_k\}} (T_i - n_i) \frac{E^{\mu|(s, \mathbf{y}_s)}(b_i)}{a_i - 1} + d_2^2, \\
d_4 &= \frac{b_0}{a_0 - 1}, \quad (5.52)
\end{aligned}$$

y  $T_u$ ,  $T_s$  y  $a_i$ ,  $b_i$ ,  $i \in \{i_1, \dots, i_k\}$  son escalares definidos en (5.46).

**Demostración.** La demostración de estos resultados se apoya en el teorema 5.5 y los lemas B.5 y B.6.

\* Los resultados (a) y (b) se demuestran de forma totalmente análoga (realizando las correcciones obvias) a los resultados (a) y (b) del teorema 5.3.

\* Y el resultado (c) se obtiene a partir de las distribuciones dadas en (5.41), (5.42), (5.43) y (5.44) por un proceso de demostración totalmente idéntico al del resultado (c) del teorema 5.3 teniendo en este caso que:

$$\begin{aligned}
E(\bar{Y}_u|(s, \mathbf{y}_s), \mathbf{T}) &= E(\mu|(s, \mathbf{y}_s)), \\
\text{Var}(\bar{Y}_u|(s, \mathbf{y}_s), \mathbf{T}) &= \frac{d_3 - d_2^2 + d_4 T_u}{(T_u + T_s - n)^2} + \text{Var}(\mu|(s, \mathbf{y}_s)).
\end{aligned}$$

**C.S.Q.D.**

Notemos que, la obtención del estimador Bayes de la media poblacional y su varianza (y de los dos primeros momentos de cualquiera de las distribuciones predictivas de interés señaladas en la letra (a) del teorema anterior) requiere del cálculo numérico de los dos primeros momentos de la distribución de  $\mu|(s, \mathbf{y}_s)$  dada en (5.43) y del cálculo de la esperanza de funciones de la variable aleatoria  $T_u | \{T_i\}_{i \in \{i_1, \dots, i_k\}}$  cuya función de

probabilidad viene dada por (5.45). Todos estos momentos se pueden obtener por técnicas de cálculo univariantes.

Otro mecanismo para obtener las distribuciones predictivas y sus primeros dos momentos (en este caso aproximadas) es el método de muestreo Gibbs (que ya hemos propuesto para el modelo Bayes de rutas aleatorias). El siguiente teorema describe las distribuciones condicionales necesarias para su aplicación.

**Teorema 5.7** *El modelo Bayes para rutas aleatorias con medias iguales definido en (5.2) es un modelo DAG (figura 5.2), con distribuciones condicionales dadas por:*

$$f((\bar{y}_{u_1}, \dots, \bar{y}_{u_K})^t | \mathbf{y}_s, \sigma^2, \mu, \{T_i\}_{i=1}^K, \lambda) = f((\bar{y}_{u_1}, \dots, \bar{y}_{u_K})^t | \mu, \sigma^2, \{T_i\}_1^K) = N_K \left( (\bar{y}_{u_1}, \dots, \bar{y}_{u_K})^t | 1_K \mu, \begin{pmatrix} \text{diag}_1^k \{ (T_i - n_i)^{-1} \sigma_i^2 \} & \mathbf{0}_{k \times (K-k)} \\ \mathbf{0}_{(K-k) \times k} & \text{diag}_{k+1}^K \{ T_i^{-1} \sigma_i^2 \} \end{pmatrix} \right), \quad (5.53)$$

$$\pi(\mu | \{\bar{y}_{u_i}\}_1^K, \mathbf{y}_s, \sigma^2, \{T_i\}_1^K, \lambda) = \pi(\mu | \{\bar{y}_{u_i}\}_1^K, \mathbf{y}_s, \sigma^2) = N(\mu | (\sum_{i=1}^K T_i \sigma_i^{-2} + \sigma_0^{-2})^{-1} (\sum_{i=1}^K \sigma_i^{-2} \sum_{j=1}^{T_i} y_{ij} + \sigma_0^{-2} \mu_0), (\sum_{i=1}^K T_i \sigma_i^{-2} + \sigma_0^{-2})^{-1}), \quad (5.54)$$

$$\pi(\sigma^2 | \{\bar{y}_{u_i}\}_1^K, \mathbf{y}_s, \mu, \{T_i\}_1^K, \lambda) = \pi(\sigma^2 | \{\bar{y}_{u_i}\}_1^K, \mathbf{y}_s, \mu, \{T_i\}_1^K) = \prod_{i=1}^K Ga^{-1}(\sigma_i^2 | a_i^*, b_i^*), \quad (5.55)$$

$$f(\mathbf{T}_u | \{\bar{y}_{u_i}\}_1^K, \mathbf{y}_s, \mu, \sigma^2, \{T_i\}_1^K, \lambda) = f(\mathbf{T}_u | \{\bar{y}_{u_i}\}_{k+1}^K, \mu, \{\sigma_i^2\}_{k+1}^K, \lambda) \propto \prod_{i \notin \{i_1, \dots, i_k\}} \left[ T_i^{1/2} \exp \left\{ -\frac{1}{2\sigma_i^2} T_i (\bar{y}_{u_i} - \mu)^2 \right\} \frac{\lambda^{T_i}}{T_i!} \right], \quad (5.56)$$

$$\pi(\lambda | \{\bar{y}_{u_i}\}_1^K, \mathbf{y}_s, \mu, \sigma^2, \{T_i\}_1^K) = \pi(\lambda | \{T_i\}_1^K) = Ga(\lambda | \sum_{i=1}^K T_i + \alpha, K + \beta), \quad (5.57)$$

donde  $\mathbf{T}_u$  aparece definido en (5.46) y,

$$\begin{aligned} a_i^* &= a_0 + \frac{n_i + 1}{2}, \quad i \in \{i_1, \dots, i_k\}, \\ &= a_0 + \frac{1}{2}, \quad i \notin \{i_1, \dots, i_k\}, \\ b_i^* &= b_0 + \frac{1}{2} \left\{ (T_i - n_i) (\bar{y}_{u_i} - \mu)^2 + \sum_{j \in s_i} (y_{ij} - \mu)^2 \right\}, \quad i \in \{i_1, \dots, i_k\}, \\ &= b_0 + \frac{1}{2} T_i (\bar{y}_{u_i} - \mu)^2, \quad i \notin \{i_1, \dots, i_k\}. \end{aligned} \quad (5.58)$$

Además, se cumple que, las distribuciones  $T_i | \{\bar{y}_{u_i}\}_1^K, \mathbf{y}_s, \theta, \mu, \sigma^2, \{T_i\}_1^K, \lambda, i \notin \{i_1, \dots, i_k\}$  son independientes con funciones de probabilidad log-concavas en los enteros positivos incluyendo el cero.

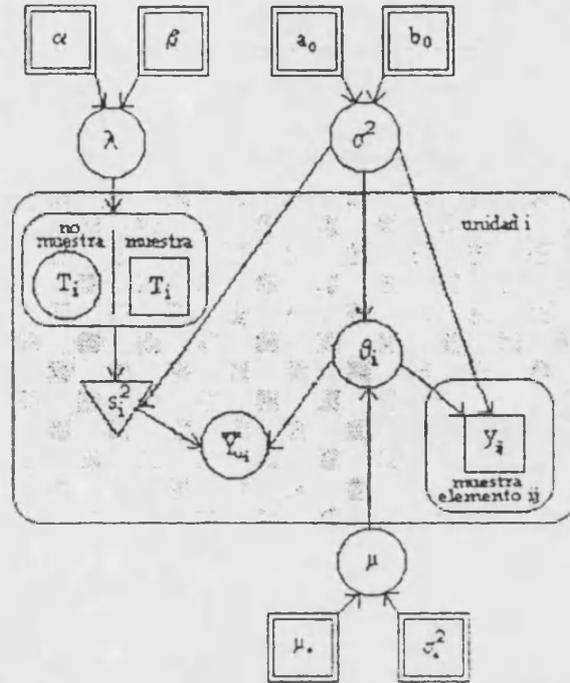


Figura 5.2: Representación gráfica del modelo Bayes para rutas aleatorias con medias iguales

**Demostración.** De la definición 5.2 del modelo en estudio se deduce directamente la representación gráfica de la figura 5.2, que responde a un modelo DAG. Las distribuciones condicionales se obtienen a partir de la distribución conjunta de  $(\bar{Y}_{u_1}, \dots, \bar{Y}_{u_K}, \mathbf{Y}_s^t, \mu, (\sigma^2)^t, \mathbf{T}^t)^t$ , que viene dada por:

$$\begin{aligned}
 f((\bar{y}_{u_1}, \dots, \bar{y}_{u_K}, \mathbf{y}_s^t, \mu, (\sigma^2)^t, \mathbf{T}^t)^t) = & \\
 & \prod_{i=1}^K N(\bar{y}_{u_i} | \mu, (T_i - n_i)^{-1} \sigma_i^2) \times \prod_{i \in \{i_1, \dots, i_k\}} \prod_{j \in s_i} N(y_{ij} | \mu, \sigma_i^2) \times \\
 & \times N(\mu | \mu_0, \sigma_0^2) \times \prod_{i=1}^K Ga^{-1}(\sigma_i^2 | a_0, b_0) \times \prod_{i=1}^K Po(T_i | \lambda) \times Ga(\lambda | \alpha, \beta),
 \end{aligned}$$

separando los términos correspondientes a cada condicional (que contienen el vector o variable cuya condicional estamos hallando) de manera análoga a como lo hacíamos en el teorema 5.4. La independencia de las variables condicionadas  $T_i$ ,  $i \notin \{i_1, \dots, i_k\}$  y la log-concavidad de sus funciones de probabilidad se deducen de la distribución (5.56) (obtenida en el proceso anterior) y la aplicación del lema B.2.

C.S.Q.D.

De la definición del modelo en (5.2) y del teorema se deducen los requisitos (ver C.2.1) para poder implementar el algoritmo Gibbs mediante el programa BUGS.

### 5.3.3 Modelo Bayes para rutas aleatorias homocedásticas

En esta subsección estudiaremos el modelo Bayes de rutas aleatorias homocedásticas, para ello siguiendo un esquema de resultados similar al de las dos subsecciones anteriores realizaremos en primer lugar (primeros dos teoremas) el estudio de las distribuciones condicionales para realizar inferencia sobre el vector de características poblacional a partir de este modelo y las hipótesis suficientes que garantizan que las distribuciones predictivas del vector poblacional no observado, la media no observada en cada ruta, la media poblacional no observada y la media poblacional tengan media y varianza. Y en segundo lugar el estudio de las distribuciones condicionales necesarias para aplicar el método de muestreo por Gibbs Sampling.

**Teorema 5.8** *Consideremos el modelo Bayes para rutas aleatorias homocedástico definido en (5.3), entonces se tiene que:*

(a)  $\mathbf{Y}_{u_i} | (s, \mathbf{y}_s), \mu, \mathbf{T}$  se distribuye  $t(N - n)$ -variante con  $2a_1$  grados de libertad con para  $a_1 > 1$ , media:

$$\begin{aligned} E(\mathbf{Y}_{u_i} | (s, \mathbf{y}_s), \mu, \mathbf{T}) &= \mathbf{1}_{T_i - n} \tilde{\theta}_i, \quad i \in \{i_1, \dots, i_k\}, \\ &= \mathbf{1}_{T_i} \mu, \quad i \notin \{i_1, \dots, i_k\} \end{aligned} \quad (5.59)$$

y matriz de varianzas y covarianzas,

$$\begin{aligned} Cov(\mathbf{Y}_{u_i}, \mathbf{Y}_{u_{i^*}} | (s, \mathbf{y}_s), \mu, \mathbf{T}) &= \mathbf{A}_i \frac{b_1}{a_1 - 1}, \quad i = i^*, \quad i \in \{i_1, \dots, i_k\} \\ &= \mathbf{B}_i \frac{b_1}{a_1 - 1}, \quad i = i^*, \quad i \notin \{i_1, \dots, i_k\} \\ &= 0, \quad i \neq i^* \end{aligned} \quad (5.60)$$

(b)  $(\bar{Y}_{u_1}, \bar{Y}_{u_2}, \dots, \bar{Y}_{u_K})^t | (s, \mathbf{y}_s), \mu, \mathbf{T}$  se distribuye  $t$   $K$ -variante con  $2a_1$  grados de libertad con para  $a_1 > 1$  media:

$$\begin{aligned} E(\bar{Y}_{u_i} | (s, \mathbf{y}_s), \mu, \mathbf{T}) &= \tilde{\theta}_i, \quad i \in \{i_1, \dots, i_k\}, \\ &= \mu, \quad i \notin \{i_1, \dots, i_k\} \end{aligned} \quad (5.61)$$

y matriz de varianzas y covarianzas,

$$\begin{aligned} Cov(\bar{Y}_{u_i}, \bar{Y}_{u_{i^*}} | (s, \mathbf{y}_s), \mu, \mathbf{T}) &= [(T_i - n_i)^{-1} + \frac{c}{n_i c + 1}] \frac{b_1}{a_1 - 1}, \quad i = i^*, \quad i \in \{i_1, \dots, i_k\} \\ &= (T_i^{-1} + c) \frac{b_1}{a_1 - 1}, \quad i = i^*, \quad i \notin \{i_1, \dots, i_k\} \\ &= 0 \quad i \neq i^* \end{aligned} \quad (5.62)$$

(c)  $\bar{Y}_u|(s, \mathbf{y}_s), \mu, \mathbf{T}$  se distribuye  $t$  con  $2a_1$  grados de libertad con para  $a_1 > 1$ , media y varianza dadas respectivamente por:

$$E(\bar{Y}_u|(s, \mathbf{y}_s), \mu, \mathbf{T}) = \frac{1}{T_u + T_s - n} \left[ \sum_{i \in \{i_1, \dots, i_k\}} (T_i - n_i) \tilde{\theta}_i + \mu T_u \right],$$

$$\text{Var}(\bar{Y}_u|(s, \mathbf{y}_s), \mu, \mathbf{T}) = \frac{1}{(T_u + T_s - n)^2} \frac{b_1}{a_1 - 1} \left[ \sum_{i \notin \{i_1, \dots, i_k\}} T_i (1 + T_i c) + \sum_{i \in \{i_1, \dots, i_k\}} (T_i - n_i) \left( 1 + \frac{(T_i - n_i) c}{n_i c + 1} \right) \right], \quad (5.63)$$

(d)  $\sigma^2|(s, \mathbf{y}_s), \mu, \mathbf{T}$  se distribuye  $\pi(\sigma^2|(s, \mathbf{y}_s), \mu, \mathbf{T}) = \pi(\sigma^2|(s, \mathbf{y}_s), \mu)$  según una gamma inversa de parámetros  $a_1$  y  $b_1$ ,

(e)  $\mu|(s, \mathbf{y}_s), \{T_i\}_1^K$  se distribuye con función de densidad  $\pi(\mu|(s, \mathbf{y}_s), \mathbf{T}) = \pi(\mu|(s, \mathbf{y}_s))$  dada por:

$$\pi(\mu|(s, \mathbf{y}_s)) \propto \left[ 1 + \frac{1}{2b_0 Z_1} (\mu - Z_2)^2 \right]^{-a_1} \times \exp \left\{ -\frac{1}{2\sigma_0^2} (\mu - \mu_0)^2 \right\}, \quad (5.64)$$

(f)  $\mathbf{T}_u|(s, \mathbf{y}_s), \{T_i\}_{i \in \{i_1, \dots, i_k\}}$  se distribuye con función de densidad  $f(\mathbf{T}_u|(s, \mathbf{y}_s), \{T_i\}_{i \in \{i_1, \dots, i_k\}}) = f(\mathbf{T}_u|\{T_i\}_{i \in \{i_1, \dots, i_k\}})$  dada por:

$$f(\mathbf{T}_u|\{T_i\}_{i \in \{i_1, \dots, i_k\}}) = \frac{(\beta + k)^{T_u + \alpha}}{\Gamma(T_s + \alpha)} \frac{\Gamma(T_u + T_s + \alpha)}{(\beta + K)^{T_u + T_s + \alpha}} \prod_{i \notin \{i_1, \dots, i_k\}} \frac{1}{T_i!}, \quad (5.65)$$

(g)  $T_u|(s, \mathbf{y}_s), \{T_i\}_{i \in \{i_1, \dots, i_k\}}$  se distribuye con función de densidad,  $f(T_u|(s, \mathbf{y}_s), \{T_i\}_{i \in \{i_1, \dots, i_k\}}) = f(T_u|\{T_i\}_{i \in \{i_1, \dots, i_k\}})$  gamma Poisson,

$$f(T_u|\{T_i\}_{i \in \{i_1, \dots, i_k\}}) = Gp(T_u|\alpha + T_s, \beta + k, K - k) \quad (5.66)$$

donde  $\mathbf{T}_u = (T_i : i \notin \{i_1, \dots, i_k\})^t$ ,  $T_s = \sum_{i \in \{i_1, \dots, i_k\}} T_i$ ,  $T_u = \sum_{i \notin \{i_1, \dots, i_k\}} T_i$ , y

$$\begin{aligned} \tilde{\theta}_i &= E(\theta_i|(s, \mathbf{y}_s), \mu) = \frac{cn_i \bar{y}_{s_i} + \mu}{cn_i + 1}, \quad i \in \{i_1, \dots, i_k\}, \\ \mathbf{A}_i &= \mathbf{I}_{T_i - n_i} + \frac{c \mathbf{1}_{T_i - n_i} \mathbf{1}_{T_i - n_i}^t}{cn_i + 1}, \quad i \in \{i_1, \dots, i_k\}, \\ \mathbf{B}_i &= \mathbf{I}_{T_i} + c \mathbf{1}_{T_i} \mathbf{1}_{T_i}^t, \quad i \notin \{i_1, \dots, i_k\}, \end{aligned} \quad (5.67)$$

$$\begin{aligned}
Z_1 &= \left[ 1 + \frac{1}{2b_0} \sum_{i \in \{i_1, \dots, i_k\}} \left( \sum_{j \in s_i} (y_{ij} - \bar{y}_{s_i})^2 + \frac{n_i \bar{y}_{s_i}}{cn_i + 1} \right) \right] \left[ \sum_{i \in \{i_1, \dots, i_k\}} \frac{n_i}{cn_i + 1} \right]^{-1}, \\
Z_2 &= \left[ \sum_{i \in \{i_1, \dots, i_k\}} \frac{n_i \bar{y}_{s_i}}{cn_i + 1} \right] \left[ \sum_{i \in \{i_1, \dots, i_k\}} \frac{n_i}{cn_i + 1} \right]^{-1}, \\
a_1 &= a_0 + \frac{n}{2}, \\
b_1 &= b_0 + \frac{1}{2} \sum_{i \in \{i_1, \dots, i_k\}} (\mathbf{y}_{s_i} - \mathbf{1}_{n_i} \mu)^t (\mathbf{I}_{n_i} + \mathbf{1}_{n_i} \mathbf{1}_{n_i}^t c)^{-1} (\mathbf{y}_{s_i} - \mathbf{1}_{n_i} \mu).
\end{aligned}$$

**Demostración.** La demostración de este teorema se apoya en la aplicación de los lemas B.5, B.8, B.9, y será dividida en varios pasos señalando a lo largo de su desarrollo cuando quedan demostrados los apartados en los que se divide el presente teorema:

1.- A partir de la distribución de  $\mathbf{Y}|\theta, \sigma^2, \mathbf{T}$ , aplicando propiedades de la distribución normal básicas tenemos que:

$$\mathbf{Y}_{s_i}|\theta, \sigma^2, \mathbf{T} \sim N_{n_i}(\mathbf{1}_{n_i} \theta_i, \mathbf{I}_{n_i} \sigma^2), \text{ independientes, } i \in \{i_1, \dots, i_k\}. \quad (5.68)$$

2.- A partir de (5.68) y la distribución de  $\theta|\mu, \sigma^2$ , dada en (5.3) se obtiene aplicando el resultado (a) del lema B.8, la distribución de  $\mathbf{Y}_s|\mu, \sigma^2, \mathbf{T}$ , que es:

$$\mathbf{Y}_s|\mu, \sigma^2, \mathbf{T} \sim N_n(\mathbf{1}_n \mu, \text{diag}_{i \in \{i_1, \dots, i_k\}} \{(\mathbf{I}_{n_i} + \mathbf{1}_{n_i} \mathbf{1}_{n_i}^t c) \sigma^2\}), \quad (5.69)$$

y aplicando el resultado (b) del lema B.8, la distribución de  $\theta|(s, \mathbf{y}_s), \mu, \sigma^2$ , según la cual  $\theta_i|(s, \mathbf{y}_s), \mu, \sigma^2$ , son independientes para  $i = 1, 2, \dots, K$  con distribución dada por:

$$\begin{aligned}
\theta_i|(s, \mathbf{y}_s), \mu, \sigma^2 &\sim N(\bar{\theta}_i, \frac{c}{n_i c + 1} \sigma^2), \quad i \in \{i_1, \dots, i_k\}, \\
&\sim N(\mu, c \sigma^2), \quad i \notin \{i_1, \dots, i_k\}.
\end{aligned} \quad (5.70)$$

3.- A partir de (5.69) y la distribución inicial para  $\sigma^2$  dada en (5.3), se obtiene, aplicando el resultado (b) del lema B.9, la distribución de  $\sigma^2|(s, \mathbf{y}_s), \mu$  ( $\rightsquigarrow$  comprobación del resultado (d)), y aplicando el resultado (a) del lema B.9 la distribución de  $\mathbf{Y}_s|\mu, \mathbf{T}$ , dada por:

$$\mathbf{Y}_s|\mu, \mathbf{T} \sim t_n(\mathbf{1}_n \mu, \frac{b_0}{a_0} \text{diag}_{i \in \{i_1, \dots, i_k\}} \{\mathbf{I}_{n_i} + \mathbf{1}_{n_i} \mathbf{1}_{n_i}^t c\}, 2a_0). \quad (5.71)$$

4.- A partir de (5.71) y la distribución inicial de  $\mu$  dada en (5.1) se obtiene, aplicando el teorema de Bayes, la distribución de  $\mu|(s, \mathbf{y}_s), \mathbf{T}$  dada en (5.64), esta distribución es propia por el lema B.5 ( $\rightsquigarrow$  comprobación del resultado (e)).

5.- A partir de la distribución de  $\mathbf{Y}|\theta, \sigma^2, \mathbf{T}$ , aplicando propiedades de la distribución normal básicas tenemos que:

$$\mathbf{Y}_{u_i}|(s, \mathbf{y}_s), \theta, \sigma^2, \mathbf{T} \sim N_{T_i-n_i}(\mathbf{1}_{T_i-n_i}\theta_i, \mathbf{I}_{N_i-n_i}\sigma^2), \text{ independientes, } i = 1, 2, \dots, K. \quad (5.72)$$

6.- A partir de (5.72) y (5.70) se obtiene, aplicando el resultado (a) del lema B.8, la distribución de  $\mathbf{Y}_u|(s, \mathbf{y}_s), \mu, \sigma^2, \mathbf{T}$ , según la cual  $\mathbf{Y}_{u_i}|(s, \mathbf{y}_s), \mu, \sigma^2, \mathbf{T}$ , son independientes para  $i = 1, 2, \dots, K$ , con distribución dada por:

$$\begin{aligned} \mathbf{Y}_{u_i}|(s, \mathbf{y}_s), \mu, \sigma^2, \mathbf{T} &\sim \\ &\sim N_{T_i-n_i}(\mathbf{1}_{T_i-n_i}\tilde{\theta}_i, (\mathbf{I}_{T_i-n_i} + \frac{c}{n_i c + 1} \mathbf{1}_{T_i-n_i} \mathbf{1}_{T_i-n_i}^t) \sigma_i^2), \quad i \in \{i_1, \dots, i_k\}, \\ &\sim N_{T_i}(\mathbf{1}_{T_i}\mu, (\mathbf{I}_{T_i} + \mathbf{1}_{T_i} \mathbf{1}_{T_i}^t) \sigma_i^2), \quad i \notin \{i_1, \dots, i_k\}, \end{aligned} \quad (5.73)$$

y aplicando las definiciones de media no observada por ruta y media no observada y que la combinación lineal de normales es normal, se obtienen las distribuciones de  $\bar{\mathbf{Y}}_{u_i}|(s, \mathbf{y}_s), \mu, \sigma^2, \mathbf{T}$ ,  $i = 1, 2, \dots, K$ , que son independientes y vienen dadas por:

$$\begin{aligned} \bar{\mathbf{Y}}_{u_i}|(s, \mathbf{y}_s), \mu, \sigma^2, \mathbf{T} &\sim N(\tilde{\theta}_i, [(T_i - n_i)^{-1} + \frac{c_i}{n_i c_i + 1}] \sigma_i^2), \quad i \in \{i_1, \dots, i_k\}, \\ &\sim N(\mu, (T_i^{-1} + c_i) \sigma_i^2), \quad i \notin \{i_1, \dots, i_k\}, \end{aligned} \quad (5.74)$$

y la distribución de  $\bar{\mathbf{Y}}_u|(s, \mathbf{y}_s), \mu, \sigma^2, \mathbf{T}$ , que es normal con media y varianza dadas por:

$$\begin{aligned} E(\bar{\mathbf{Y}}_u|(s, \mathbf{y}_s), \mu, \sigma^2, \mathbf{T}) &= \frac{1}{T_u + T_s - n} \left[ \sum_{i \in \{i_1, \dots, i_k\}} (T_i - n_i) \tilde{\theta}_i + \mu T_u \right], \\ \text{Var}(\bar{\mathbf{Y}}_u|(s, \mathbf{y}_s), \mu, \sigma^2, \mathbf{T}) &= \frac{\sigma^2}{(T_u + T_s - n)^2} \left[ \sum_{i \notin \{i_1, \dots, i_k\}} T_i (1 + T_i c) + \right. \\ &\quad \left. + \sum_{i \in \{i_1, \dots, i_k\}} (T_i - n_i) \left( 1 + \frac{(T_i - n_i) c}{n_i c + 1} \right) \right], \end{aligned} \quad (5.75)$$

7.- A partir de (5.73) y el resultado (d) de este teorema (ya comprobado en el punto 3 de esta demostración), se obtiene la distribución de  $\mathbf{Y}_u|(s, \mathbf{y}_s), \mu, \mathbf{T}$  dada en (5.59) y (5.60) ( $\leadsto$  comprobación del resultado (a)), a partir de (5.74) y el resultado (d) de este teorema, la distribución de  $\bar{\mathbf{Y}}_{u_i}|(s, \mathbf{y}_s), \mu, \mathbf{T}$ ,  $i = 1, 2, \dots, K$ , dada en (5.61) y (5.62) ( $\leadsto$  comprobación del resultado (b)) y a partir de (5.75) y el resultado (d) de

este teorema, la distribución de  $\bar{Y}_u|(s, \mathbf{y}_s), \mu, \mathbf{T}$ , dada en (5.63) ( $\leadsto$  comprobación del resultado (c)). En los tres casos se aplica el resultado (a) del lema B.9.

8.- La demostración de los resultados (f) y (g) del teorema es idéntica a la demostración de los apartados (f) y (g) del teorema 5.2 (pasos 8 a 11).

C.S.Q.D.

**Teorema 5.9** Consideremos el modelo Bayes para rutas aleatorias homocedásticas definido en (5.3), entonces se tiene que:

(a) las distribuciones predictivas de  $\mathbf{Y}_u|(s, \mathbf{y}_s), \{T_i\}_{i \in \{i_1, \dots, i_k\}}, \bar{Y}_{u_i}|(s, \mathbf{y}_s), \{T_i\}_{i \in \{i_1, \dots, i_k\}}, i = 1, 2, \dots, K, \bar{Y}_u|(s, \mathbf{y}_s), \{T_i\}_{i \in \{i_1, \dots, i_k\}}$  y  $\bar{Y}|(s, \mathbf{y}_s), \{T_i\}_{i \in \{i_1, \dots, i_k\}}$ , son propias y bajo la hipótesis  $a_0 > 1$  (ó  $n \geq 3$ ) tienen primeros dos momentos,

(b) las distribuciones finales de  $\sigma^2, \theta^2, \mu$  y  $\mathbf{T}_u$  condicionadas a  $(s, \mathbf{y}_s)$  y  $\{T_i\}_{i \in \{i_1, \dots, i_k\}}$ , son propias y tienen media y varianza bajo la hipótesis anterior  $a_0 > 1$  (ó  $n \geq 3$ ),

y (c) el estimador Bayes para la media poblacional y su varianza vienen dados, si  $a_0 > 1$  (ó  $n \geq 3$ ), por:

$$\begin{aligned} \hat{\bar{Y}} &= d_2 E \left[ \frac{1}{T_u + T_s} \right] + E(\mu|(s, \mathbf{y}_s)) E \left[ \frac{T_u}{T_u + T_s} \right], \\ \text{Var}(\bar{Y}|(s, \mathbf{y}_s), \{T_i\}_{i \in \{i_1, \dots, i_k\}}) &= (d_1^2 \text{Var}(\mu|(s, \mathbf{y}_s)) + d_3) E \left[ \frac{1}{(T_u + T_s)^2} \right] + \\ &+ 2[d_1 \text{Var}(\mu|(s, \mathbf{y}_s)) + d_2 E(\mu|(s, \mathbf{y}_s))] E \left[ \frac{T_u}{(T_u + T_s)^2} \right] + \\ &+ E(\mu^2|(s, \mathbf{y}_s)) E \left[ \frac{T_u^2}{(T_u + T_s)^2} \right] + \\ &+ d_4 \sum_{i \notin \{i_1, \dots, i_k\}} \left\{ E \left[ \frac{T_i}{(T_u + T_s)^2} \right] + c E \left[ \frac{T_i^2}{(T_u + T_s)^2} \right] \right\} - \\ &- \left\{ d_2 E \left[ \frac{1}{T_u + T_s} \right] + E(\mu|(s, \mathbf{y}_s)) E \left[ \frac{T_u}{T_u + T_s} \right] \right\}^2 \quad (5.76) \end{aligned}$$

donde  $E(\cdot)$  es el operador esperanza respecto a  $\mathbf{T}_u|\{T_i\}_{i \in \{i_1, \dots, i_k\}}$ , cuya distribución está definida por (5.65),

$$\begin{aligned} d_1 &= \sum_{i \in \{i_1, \dots, i_k\}} \frac{T_i - n_i}{cn_i + 1}, \\ d_2 &= n\bar{y}_s + \sum_{i \in \{i_1, \dots, i_k\}} (T_i - n_i) \frac{cn_i \bar{y}_{s_i}}{cn_i + 1} + d_1 E(\mu|(s, \mathbf{y}_s)), \end{aligned}$$

$$\begin{aligned}
d_3 &= \frac{E^{\mu(s, \mathbf{y}_s)}(b_1)}{a_1 - 1} \sum_{i \in \{i_1, \dots, i_k\}} (T_i - n_i) \left( 1 + \frac{(T_i - n_i)c}{cn_i + 1} \right) + d_2^2, \\
d_4 &= \frac{E^{\mu(s, \mathbf{y}_s)}(b_1)}{a_1 - 1}, \tag{5.77}
\end{aligned}$$

y  $\mathbf{T}_u, T_u, T_s, a_1$  y  $b_1$ , vector y escalares definidos en (5.67).

**Demostración.** La demostración de estos resultados se apoya en el teorema 5.8 y los lemas B.5 y B.6.

\* Para comprobar el resultado (a) notemos que las distribuciones  $\mathbf{Y}_u|(s, \mathbf{y}_s)$ ,  $\{T_i\}_{i \in \{i_1, \dots, i_k\}}$ ,  $\bar{Y}_{u_i}|(s, \mathbf{y}_s)$ ,  $\{T_i\}_{i \in \{i_1, \dots, i_k\}}$ ,  $i = 1, 2, \dots, K$ ,  $\bar{Y}_u|(s, \mathbf{y}_s)$ ,  $\{T_i\}_{i \in \{i_1, \dots, i_k\}}$  y  $\bar{Y}|(s, \mathbf{y}_s)$ ,  $\{T_i\}_{i \in \{i_1, \dots, i_k\}}$  son propias y tienen primeros dos momentos para  $a_0 > 1$  por ser las condicionales necesarias para su obtención (para  $\mathbf{Y}_u$  las distribuciones del resultado (a), (5.64) y (5.65), para  $\bar{Y}_{u_i}$  las distribuciones del resultado (b), (5.64) y (5.65), para  $\bar{Y}_u$  las distribuciones del resultado (c), (5.64) y (5.65) y para  $\bar{Y}$  las distribuciones (5.63) (aplicando que la combinación lineal de normales es normal) del resultado (d), (5.64) y (5.65) ) propias y tener primeros dos momentos bajo la hipótesis<sup>9</sup>  $a_0 > 1$ . Esta última afirmación se apoya en los lemas B.5 y B.6 y en las propiedades de las distribuciones normales,  $t$  y  $\gamma$ .

\* El resultado (b) se demuestra de forma análoga (realizando las correcciones obvias) al resultado (b) del teorema 5.3.

\* Y el resultado (c) se obtiene a partir de las distribuciones dadas en (5.63), (5.64) y (5.65) por un proceso de demostración totalmente idéntico al del resultado (c) del teorema 5.3 teniendo en este caso que:

$$\begin{aligned}
E(\bar{Y}_u|(s, \mathbf{y}_s), \mathbf{T}) &= E^{\mu(s, \mathbf{y}_s)} E(\bar{Y}_u|(s, \mathbf{y}_s), \mu, \mathbf{T}) = \\
&= \frac{d_2 - n\bar{y}_s + T_u E(\mu|(s, \mathbf{y}_s))}{T_u + T_s - n}, \\
\text{Var}(\bar{Y}_u|(s, \mathbf{y}_s), \mathbf{T}) &= E^{\mu(s, \mathbf{y}_s)} \text{Var}(\bar{Y}_u|(s, \mathbf{y}_s), \mu, \mathbf{T}) + \\
&+ \text{Var}^{\mu(s, \mathbf{y}_s)} E(\bar{Y}_u|(s, \mathbf{y}_s), \mu, \mathbf{T}) = \\
&= \frac{d_3 - d_2^2 + d_4 \sum_{i \in \{i_1, \dots, i_k\}} T_i(1 + T_i c_i) + (d_1 + T_u)^2 \text{Var}(\mu|(s, \mathbf{y}_s))}{(T_u + T_s - n)^2}.
\end{aligned}$$

C.S.Q.D.

Nuevamente, la obtención del estimador Bayes de la media poblacional y su varianza condicionada a la muestra y los tamaños de las rutas muestreadas (y de los dos primeros

<sup>9</sup> esta hipótesis puede debilitarse por la hipótesis  $n \geq 3$

momentos de cualquiera de las distribuciones predictivas de interés señaladas en la letra (a) del teorema anterior) requiere del cálculo numérico de los dos primeros momentos de la distribución de  $\mu|(s, \mathbf{y}_s)$  dada en (5.64) y del cálculo de la esperanza de funciones del vector aleatorio  $\mathbf{T}_u|\{T_i\}_{i \in \{i_1, \dots, i_k\}}$  cuya función de probabilidad viene dada por (5.65). En el caso del cálculo de los dos primeros momentos de la distribución de  $\mu|(s, \mathbf{y}_s)$  se requieren técnicas de integración numérica de una variable (salvo cuando el modelo Bayes de rutas aleatorias homocedásticas es mínimo informativo con distribución inicial para  $\mu$ ,  $\pi(\mu) \propto 1$ , y consecuentemente la distribución condicional de la variable  $\mu|(s, \mathbf{y}_s)$  es una  $t$  univariante, distribución estándar cuyos dos primeros momentos son conocidos), y en el caso de los momentos de funciones del vector  $\mathbf{T}_u|\{T_i\}_{i \in \{i_1, \dots, i_k\}}$  se necesita recurrir a procedimientos numéricos  $(K - k)$ -variables (de hecho se necesitan calcular los mismos momentos que en el modelo Bayes de rutas aleatorias definido en (5.1)).

Otro mecanismo para obtener las distribuciones predictivas y sus primeros dos momentos (en este caso aproximadas) es el método de muestreo Gibbs (que ya hemos propuesto para el modelo Bayes de rutas aleatorias y el modelo Bayes con medias iguales). El siguiente teorema describe las distribuciones condicionales necesarias para su aplicación.

**Teorema 5.10** *El modelo Bayes para rutas aleatorias homocedástico definido en (5.9) es un modelo DAG (figura 5.9), con distribuciones condicionales dadas por:*

$$f((\bar{y}_{u_1}, \dots, \bar{y}_{u_K})^t | \mathbf{y}_s, \theta, \sigma^2, \mu, \{T_i\}_1^K, \lambda) = f((\bar{y}_{u_1}, \dots, \bar{y}_{u_K})^t | \theta, \sigma^2, \{T_i\}_1^K) = N_K \left( (\bar{y}_{u_1}, \dots, \bar{y}_{u_K})^t | \theta, \sigma^2 \begin{pmatrix} \text{diag}_1^k \{(T_i - n_i)^{-1}\} & \mathbf{0}_{k \times (K-k)} \\ \mathbf{0}_{(K-k) \times k} & \text{diag}_{k+1}^K \{T_i^{-1}\} \end{pmatrix} \right), \quad (5.78)$$

$$\pi(\theta | \{\bar{y}_{u_i}\}_1^K, \mathbf{y}_s, \sigma^2, \mu, \{T_i\}_1^K, \lambda) = \pi(\theta | \{\bar{y}_{u_i}\}_1^K, \mathbf{y}_s, \sigma^2, \mu, \{T_i\}_1^K) = N_K(\theta | \tilde{\mu}, \sigma^2 \text{diag}_1^K \{c(T_i c + 1)^{-1}\}), \quad (5.79)$$

$$\pi(\mu | \{\bar{y}_{u_i}\}_1^K, \mathbf{y}_s, \theta, \sigma^2, \{T_i\}_1^K, \lambda) = \pi(\mu | \theta, \sigma^2) = N(\mu | (Kc^{-1}\sigma^{-2} + \sigma_0^{-2})^{-1}(c^{-1}\sigma^{-2} \sum_{i=1}^K \theta_i + \sigma_0^{-2}\mu_0), (Kc^{-1}\sigma^{-2} + \sigma_0^{-2})^{-1}), \quad (5.80)$$

$$\pi(\sigma^2 | \{\bar{y}_{u_i}\}_1^K, \mathbf{y}_s, \theta, \mu, \{T_i\}_1^K, \lambda) = \pi(\sigma^2 | \{\bar{y}_{u_i}\}_1^K, \mathbf{y}_s, \theta, \mu, \{T_i\}_1^K) = Ga^{-1}(\sigma^2 | a_1^*, b_1^*), \quad (5.81)$$

$$f(\mathbf{T}_u | \{\bar{y}_{u_i}\}_1^K, \mathbf{y}_s, \theta, \mu, \sigma^2, \{T_i\}_1^K, \lambda) = f(\mathbf{T}_u | \{\bar{y}_{u_i}\}_{k+1}^K, \{\theta_i\}_{k+1}^K, \sigma^2, \lambda) \propto \prod_{i \notin \{i_1, \dots, i_k\}} \left[ T_i^{1/2} \exp \left\{ -\frac{1}{2\sigma^2} T_i (\bar{y}_{u_i} - \theta_i)^2 \right\} \frac{\lambda^{T_i}}{T_i!} \right], \quad (5.82)$$

$$\pi(\lambda | \{\bar{y}_{u_i}\}_1^K, \mathbf{y}_s, \theta, \mu, \sigma^2, \{T_i\}_1^K) = \pi(\lambda | \{T_i\}_1^K) = Ga(\lambda | \sum_{i=1}^K T_i + \alpha, K + \beta), \quad (5.83)$$

donde  $\tilde{\mu} = (\tilde{\mu}_1, \tilde{\mu}_2, \dots, \tilde{\mu}_K)^t$ ,  $\mathbf{T}_u$  aparece definido en (5.67) y,

$$\begin{aligned}
\tilde{\mu}_i &= \frac{(T_i - n_i)c\bar{y}_{u_i} + n_i c\bar{y}_{s_i} + \mu}{T_i c + 1}, \quad i \in \{i_1, \dots, i_k\}, \\
&= \frac{T_i c\bar{y}_{u_i} + \mu}{T_i c + 1}, \quad i \notin \{i_1, \dots, i_k\}, \\
a_1^* &= a_0 + \frac{n}{2} + K, \\
b_1^* &= b_0 + \frac{1}{2c} \left\{ c \sum_{i \in \{i_1, \dots, i_k\}} [(T_i - n_i)(\bar{y}_{u_i} - \theta_i)^2 + \sum_{j \in s_i} (y_{ij} - \theta_i)^2] + \right. \\
&\quad \left. + c \sum_{i \notin \{i_1, \dots, i_k\}} [T_i(\bar{y}_{u_i} - \theta_i)^2] + \sum_{i=1}^K (\theta_i - \mu)^2 \right\}.
\end{aligned} \tag{5.84}$$

Además, se cumple que, las distribuciones  $T_i | \{\bar{y}_{u_i}\}_1^K, \mathbf{y}_s, \theta, \mu, \sigma^2, \{T_i\}_1^k, \lambda, i \notin \{i_1, \dots, i_k\}$  son independientes con funciones de probabilidad log-concavas en los enteros positivos incluyendo el cero.

**Demostración.** De la definición 5.3 del modelo en estudio se deduce directamente la representación gráfica de la figura 5.3, que responde a un modelo DAG. Las distribuciones condicionales se obtienen a partir de la distribución conjunta de  $(\bar{Y}_{u_1}, \dots, \bar{Y}_{u_K}, \mathbf{Y}_s^t, \theta^t, \mu, \sigma^2, \mathbf{T}^t)^t$ , que viene dada por:

$$\begin{aligned}
f((\bar{y}_{u_1}, \dots, \bar{y}_{u_K}, \mathbf{y}_s^t, (\theta)^t, \mu, \sigma^2, \mathbf{T}^t)^t) = \\
\prod_{i=1}^K N(\bar{y}_{u_i} | \theta_i, (T_i - n_i)^{-1} \sigma^2) \times \prod_{i \in \{i_1, \dots, i_k\}} \prod_{j \in s_i} N(y_{ij} | \theta_i, \sigma^2) \times \\
\times \prod_{i=1}^K N(\theta_i | \mu, c\sigma^2) \times N(\mu | \mu_0, \sigma_0^2) \times Ga^{-1}(\sigma^2 | a_0, b_0) \times \prod_{i=1}^K Po(T_i | \lambda) \times Ga(\lambda | \alpha, \beta),
\end{aligned}$$

separando los términos correspondientes a cada condicional (que contienen el vector o variable cuya condicional estamos hallando) de manera análoga a como lo hacíamos en el teorema 5.4. La independencia de las variables condicionadas  $T_i, i \notin \{i_1, \dots, i_k\}$  y la log-concavidad de sus funciones de probabilidad se deducen de la distribución (5.82) (obtenida en el proceso anterior) y la aplicación del lema B.2.

C.S.Q.D.

De la definición del modelo en (5.3) y del teorema 5.10 se deducen los requisitos (ver C.2.1) para poder implementar el algoritmo Gibbs mediante el programa BUGS.

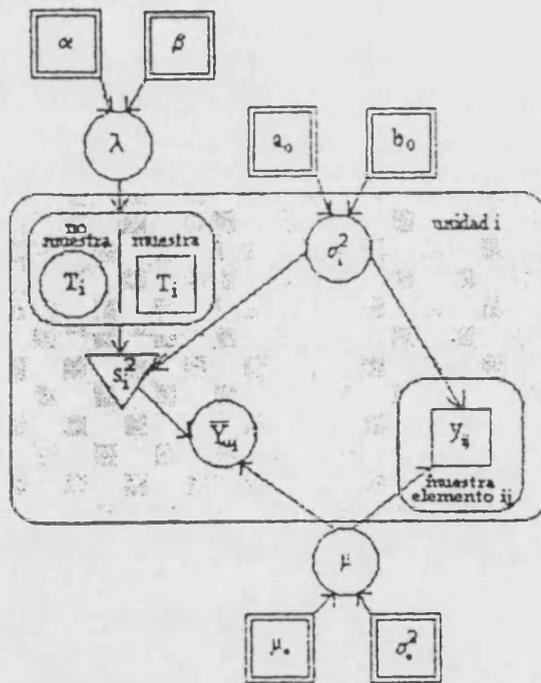


Figura 5.3: Representación gráfica del modelo Bayes para rutas aleatorias homocedástico

### 5.3.4 Modelos para rutas aleatorias mínimo informativos

En la práctica, la asignación de las distribuciones iniciales para los modelos no es siempre fácil y para solucionar este problema o expresar una opinión más neutral, se recurre al uso de distribuciones iniciales mínimo informativas. En esta subsección estudiaremos la problemática que supone la asignación de distribuciones mínimo informativas en los modelos Bayes para rutas aleatorias.

Respecto a la introducción de distribuciones mínimo informativas sobre los parámetros media y varianza tenemos los siguientes resultados para los tres modelos:

#### Modelo Bayes para rutas aleatorias mínimo informativo

Una versión mínimo informativa del modelo Bayes para rutas aleatorias consiste en tomar como distribución inicial para  $\mu$ ,  $\pi(\mu) \propto 1$  y una distribución inicial propia para las variables  $\sigma_i^2$ ,  $i = 1, 2, \dots, K$ . Para este modelo mínimo informativo es inmediato comprobar, tomando  $\sigma_0^{-2} \rightarrow 0$ , que se satisfacen los teoremas 5.2, 5.3 y 5.4 sustituyendo  $\mu_0 = 0$  y  $\sigma_0^{-2} = 0$  (la distribución de  $\mu|(s, \mathbf{y}_s)$  es propia a partir del lema B.5).

Del teorema 5.2, se desprende que toda versión mínimo informativa del modelo Bayes para rutas aleatorias ha de cumplir si  $k < K$ , que la distribución inicial para las variables  $\sigma_i^2$ ,  $i = 1, 2, \dots, K$  ha de ser propia, puesto que, en otro caso la distribución condicional de  $\sigma^2|(s, \mathbf{y}_s), \mu$  no sería propia y en consecuencia la distribución predictiva del vector  $\mathbf{Y}_u|(s, \mathbf{y}_s), \{T_i\}_{i \in \{i_1, \dots, i_k\}}$  y las distribuciones de las funciones lineales del mismo tampoco serían propias.

#### Modelo Bayes para rutas aleatorias con medias iguales mínimo informativo

Una versión mínimo informativa del modelo Bayes para rutas aleatorias con medias iguales consiste en tomar como distribución inicial para  $\mu$ ,  $\pi(\mu) \propto 1$  y una distribución inicial propia para las variables  $\sigma_i^2$ ,  $i = 1, 2, \dots, K$ . Para este modelo mínimo informativo es inmediato comprobar, tomando  $\sigma_0^{-2} \rightarrow 0$ , que se satisfacen los teoremas 5.5, 5.6 y 5.7 sustituyendo  $\mu_0 = 0$  y  $\sigma_0^{-2} = 0$  (la distribución de  $\mu|(s, \mathbf{y}_s)$  es propia a partir del lema B.5).

Del teorema 5.5, se desprende que toda versión mínimo informativa del modelo Bayes para rutas aleatorias ha de cumplir si  $k < K$ , que la distribución inicial para las variables  $\sigma_i^2$ ,  $i = 1, 2, \dots, K$  ha de ser propia por la misma razón que señalábamos al referirnos a los modelos para rutas aleatorias mínimo informativos.

#### Modelo Bayes para rutas aleatorias homocedásticas mínimo informativo

Para el modelo Bayes para rutas aleatorias homocedásticas consideraremos de dos versiones mínimo informativas con la característica añadida de que para estas dos versiones es posible integrar  $\mu$  y  $\sigma^2$ , obteniendo distribuciones estándar (más concretamente t) para el vector  $\mathbf{Y}_u|(s, \mathbf{y}_s), \mathbf{T}$  (y las funciones lineales del mismo).

La primera versión consiste en tomar distribución inicial para  $\mu$  una mínimo informativa de la forma  $\pi(\mu) \propto 1$  y mantener una distribución inicial propia para  $\sigma^2$ ; en este caso se pueden obtener la distribución para el vector  $\mathbf{Y}_u|(s, \mathbf{y}_s), \mathbf{T}$  a partir del corolario 3.1 y la distribución de<sup>10</sup>  $\sigma^2|(s, \mathbf{y}_s)$ , usando el lema B.9.

La segunda versión consiste en tomar una distribución inicial conjunta para  $(\mu, \sigma^2)$  mínimo informativa de la forma  $\pi(\mu, \sigma^2) \propto \frac{1}{\sigma^2}$ ; en este caso, la distribución del vector  $\mathbf{Y}_u|(s, \mathbf{y}_s), \mathbf{T}$  es un caso particular de la obtenida en el corolario 3.2 para el vector  $\mathbf{Y}_u$ .

En ambas versiones, para integrar el vector  $\mathbf{T}_u|\{T_i\}_{i \in \{i_1, \dots, i_k\}}$  (cuya distribución sigue viniendo dada por (5.65)) se pueden aplicar métodos numéricos  $(K - k)$ -variantes o bien aplicar el método de muestreo por Gibbs, en cuyo caso las distribuciones iniciales condicionales vienen dadas por el teorema 5.10, tomando  $\mu_0 = 0$  y  $\sigma_0^{-2} = 0$ , para la primera versión y tomando además  $a_0 = 0$  y  $b_0 = 0$ , en el caso de la segunda versión.

Además en el caso de los tres modelos se puede asumir (en relación a la distribución sobre los tamaños de las rutas), una distribución mínimo informativa sobre  $\lambda$ , esto es,  $\pi(\lambda) \propto 1$ ; en este caso la distribución de  $\mathbf{T}_u|\{T_i\}_{i \in \{i_1, \dots, i_k\}}$  común a los tres modelos viene dada por (5.19) (o equivalentemente por (5.44) y (5.65)), tomando  $\alpha = 1$  y  $\beta = 0$  (tal y como se deduce aplicando el lema B.12) y es propia, con lo que la distribución predictiva del vector  $\mathbf{Y}_u|(s, \mathbf{y}_s), \{T_i\}_{i \in \{i_1, \dots, i_k\}}$  y de sus funciones lineales siguen siendo propias. Respecto a las distribuciones condicionales necesarias para aplicar el método Gibbs en los tres modelos (con distribución inicial para los tamaños también mínimo informativa), éstas se obtienen de manera inmediata a partir de los teoremas 5.4, 5.7 y 5.10 tomando adicionalmente en este caso  $\alpha = 1$  y  $\beta = 0$ .

### 5.3.5 Modelo Bayes para muestras aleatorias con varianzas desconocidas

Una vez analizados los modelos para rutas aleatorias pasaremos a realizar el estudio de los modelos Bayes para muestras aleatorias con varianza desconocida y varianza y tamaño de la población desconocidos propuestos en (5.12) y (5.13). El análisis Bayes de estos dos modelos queda resumido en los siguientes teoremas:

**Teorema 5.11** *Consideremos el modelo Bayes para m.a.s con varianza desconocida definido en (5.12), entonces se tiene que:*

(a)  $\mathbf{Y}_u|(s, \mathbf{y}_s)$  se distribuye según una  $t$ ,  $(N - n)$ -variante dada por:

$$f(\mathbf{y}_u|(s, \mathbf{y}_s)) = t_{N-n}(\mathbf{y}_u | \mathbf{1}_{N-n} \bar{y}_s, s^2 \left( \mathbf{I}_{N-n} + \frac{\mathbf{1}_{N-n} \mathbf{1}_{N-n}^t}{n} \right), n - 1), \quad (5.85)$$

(b)  $\bar{Y}_u|(s, \mathbf{y}_s)$  se distribuye según una  $t$  dada por:

<sup>10</sup>que se obtiene a partir de la distribución de  $\mathbf{y}_s|\mu_0, \sigma^2$  (que se obtiene aplicando dos veces el lema B.8 (a)) y de la distribución inicial de  $\sigma^2$ , aplicando en lema B.10

$$f(\bar{y}_u|(s, \mathbf{y}_s)) = t(\bar{y}_u|\bar{y}_s, \frac{Ns^2}{n(N-n)}, n-1), \quad (5.86)$$

y (c) el estimador Bayes para la media poblacional y su varianza (condicionada a una muestra de la población) vienen dados, para  $n > 3$  por:

$$\begin{aligned} \hat{Y} &= \bar{y}_s, \\ \text{Var}(\bar{Y}|(s, \mathbf{y}_s)) &= \frac{(n-1)s^2}{n-3} \left( \frac{1}{n} - \frac{1}{N} \right), \end{aligned} \quad (5.87)$$

donde  $\bar{y}_s$  es la media muestral y  $s^2 = \frac{\sum_{i \in s} (y_i - \bar{y}_s)^2}{n-1}$ .

**Demostración.** La demostración de este teorema es inmediata a partir del corolario 2.2 tomando  $\mathbf{X} = \mathbf{1}_N$  y  $\mathbf{W} = \mathbf{I}_N$ .

C.S.Q.D.

**Teorema 5.12** Consideremos el modelo Bayes para m.a.s con varianza desconocida y tamaño poblacional desconocidos definido en (5.13), entonces se tiene que:

(a)  $\mathbf{Y}_u|(s, \mathbf{y}_s), T$  se distribuye según una  $t$ ,  $(N-n)$ -variante dada por:

$$f(\mathbf{y}_u|(s, \mathbf{y}_s), T) = t_{T-n}(\mathbf{y}_u | \mathbf{1}_{T-n}\bar{y}_s, s^2 \left( \mathbf{1}_{T-n} + \frac{\mathbf{1}_{T-n}\mathbf{1}_{T-n}^t}{n} \right), n-1), \quad (5.88)$$

(b)  $\bar{Y}_u|(s, \mathbf{y}_s), T$  se distribuye una  $t$  dada por:

$$f(\bar{y}_u|(s, \mathbf{y}_s), T) = t(\bar{y}_u|\bar{y}_s, \frac{Ts^2}{n(T-n)}, n-1), \quad (5.89)$$

(c) la distribución predictiva de  $\bar{Y}_u|(s, \mathbf{y}_s)$  es propia con media y varianza dadas para  $n > 3$ ,  $\alpha > N_s - n + 2$  y  $N_s > n$ , por

$$\begin{aligned} E(\bar{Y}_u|(s, \mathbf{y}_s)) &= \bar{y}_s, \\ \text{Var}(\bar{Y}_u|(s, \mathbf{y}_s)) &= \frac{(n-1)s^2}{(n-3)n} [F(N_s - n, 2) - F(N_s - n - 1, 1)], \end{aligned} \quad (5.90)$$

y (d) la distribución predictiva de  $\bar{Y}|(s, \mathbf{y}_s)$  es propia y el estimador Bayes para la media poblacional y su varianza vienen dados, para  $n > 3$  y  $\alpha > N_s$  por:

$$\begin{aligned}\widehat{Y} &= \bar{y}_s, \\ \text{Var}(\bar{Y}|(s, \mathbf{y}_s)) &= \frac{(n-1)s^2}{n-3} F(N_s - 1, 1),\end{aligned}\quad (5.91)$$

donde  $\bar{y}_s$  es la media muestral,  $s^2 = \frac{\sum_{i \in \mathcal{S}_s} (y_i - \bar{y}_s)^2}{n-1}$  y

$$\begin{aligned}F(x_0, m) &= \sum_{i=0}^{x_0} \binom{x_0}{i} (-1)^i i! t^{-(m+i)} \beta^{m+i} \frac{\Gamma(\alpha - m - i)}{\Gamma(\alpha)}, \quad x_0 > 0, \\ &= \frac{1}{t} \frac{\Gamma(\alpha - m)}{\Gamma(\alpha)} \left[ \beta - \frac{\beta^\alpha}{(\beta + t)^{\alpha - m}} \right], \quad x_0 = 0.\end{aligned}\quad (5.92)$$

**Demostración.** La demostración de estos resultados se apoya en el teorema 5.11 y los lemas B.7 y B.12.

- Los resultados (a) y (b) se deducen de manera inmediata de la definición del modelo Bayes para m.a.s con tamaño y varianza desconocidos y el teorema 5.11.
- Las distribuciones  $\bar{Y}_u|(s, \mathbf{y}_s)$  y  $\bar{Y}|(s, \mathbf{y}_s)$  son propias porque las condicionales  $\bar{Y}_u|(s, \mathbf{y}_s), T$ ,  $\bar{Y}|(s, \mathbf{y}_s), T$  son propias (por los apartados (a) y (b) de este teorema) y la distribución de  $T_u$  es asimismo propia ( $T_u \sim Gp(\alpha, \beta, t)$ ) por aplicación del resultado (c) del lema B.12).
- Los restantes resultados se deducen a partir de las distribuciones señaladas en el punto anterior y la definición de estimador Bayes, aplicando para  $\bar{Y}_u$  (y análogamente para  $\bar{Y}$  con las modificaciones obvias) que:

$$\begin{aligned}E(\bar{Y}_u|(s, \mathbf{y}_s)) &= E^{T_u} E(\bar{Y}_u|(s, \mathbf{y}_s), T) = \bar{y}_s, \\ \text{Var}(\bar{Y}_u|(s, \mathbf{y}_s)) &= E^{T_u} \text{Var}(\bar{Y}_u|(s, \mathbf{y}_s), T) + \text{Var}^{T_u} E(\bar{Y}_u|(s, \mathbf{y}_s), T) = \\ &= E^{T_u} \text{Var}(\bar{Y}_u|(s, \mathbf{y}_s), T) = \\ &= \frac{(n-1)s^2}{(n-3)n} \left( E^{T_u} \left[ \frac{T_u}{T-n} \right] + N_s E^{T_u} \left[ \frac{1}{T-n} \right] \right)\end{aligned}$$

y los resultados (a) y (b) del lema B.7.

C.S.Q.D.

## 5.4 Análisis empírico de los modelos Bayes para rutas

Esta sección se dedicará al estudio, a través de tres poblaciones simuladas, del funcionamiento de los modelos Bayes para rutas propuestos en la sección 5.2, comparando

sus resultados con los obtenidos para los modelos Bayes para muestras aleatorias descritos en la misma sección. Dividiremos, para su exposición, el análisis en varias subsecciones en las que se procederá a: la descripción de las poblaciones simuladas, la descripción de los procedimientos aplicados para realizar las inferencias sobre las mismas y la exposición y discusión de los resultados obtenidos.

### 5.4.1 Descripción de las poblaciones simuladas

Para realizar el presente análisis se han simulado tres poblaciones; estas tres poblaciones responden a una situación de rutas aleatorias, en la que los elementos de la población se clasifican en rutas, y tienen además las siguientes características comunes:

- El número de rutas en las que se clasifican los elementos de la misma es 20 ( $K = 20$ ).
- Los tamaños de cada ruta son una muestra Poisson con media 500 ( $\lambda = 500$ ).

Una vez simulada las poblaciones se procedió en todas ellas a la obtención de dos muestras de las mismas de la siguiente manera<sup>11</sup>:

- elección aleatoria de 4 rutas distintas ( $k = 4$ ) de las 20 existentes y observación de los tamaños correspondientes a estas rutas,
- elección aleatoria de 50 (muestra 50) y 100 (muestra 100) elementos distintos de cada una de las 4 rutas seleccionadas en el paso anterior; para estos elementos se observa el valor de la característica poblacional observada.

Por lo tanto, la inferencia que se realizará sobre estas poblaciones se apoyará en una muestra para rutas aleatorias del tipo descrito en la sección 5.2 (y consecuentemente en los datos observados para la muestra y los tamaños de las rutas muestreadas).

Veamos a continuación, las características propias de cada población y cómo ha sido simulada:

#### Población 3

Para la simulación de la población 3 se aplicó el siguiente procedimiento (que dio lugar con los mismos datos a dos poblaciones que conoceremos por población 3 no ordenada y población 3 ordenada):

1. Obtención de una muestra de tamaño 20 de una distribución Poisson con media 500, ( $\rightsquigarrow \{T_i\}_{i=1}^{20}$ ).
2. Obtención de la suma de la muestra anterior, a fin de obtener el tamaño de la población, ( $\rightsquigarrow N$ ).

---

<sup>11</sup>La elección de las rutas y los elementos a muestrear en cada ruta se hizo de manera individual para cada población y para cada una de las dos muestras

3. Obtención de una muestra de tamaño  $N$  de una normal de media 80 y varianza  $15^2$ . Los datos simulados se guardaron en el orden de simulación y en orden creciente (de menor a mayor valor de la característica simulada).
4. Para la división de la población en rutas, las rutas se construyeron a partir de los dos bancos de datos anteriores, rellenando en el orden del banco, ruta a ruta según los tamaños de cada rutas simulados en el primer punto. Se obtiene así, para el banco no ordenado la población 3 no ordenada y para el banco ordenado en orden creciente la población 3 ordenada, ambas con sus elementos separados en rutas.

Como consecuencia de su construcción tenemos que:

- La población 3 no ordenada cumple las hipótesis para los modelos para muestras aleatorias con varianza desconocida y tamaño desconocidos, en el sentido de que la población tomada en su globalidad está formada por elementos independientes e igualmente distribuidos según una normal y la muestra por rutas obtenida es una muestra aleatoria de la población ya que la división por rutas da lugar, al mantenerse el orden de simulación, a rutas iguales entre sí en medias y varianzas.
- La población 3 ordenada, si bien se obtiene con los mismos datos que la no ordenada, ya no responde tan fielmente a los modelos para m.a.s., ya que la muestra de la población se obtiene a partir de la división en rutas practicada y en este caso esta división no respeta totalmente la aleatoriedad de los elementos de la población, dando lugar a rutas con medias distintas entre sí y de una varianza similar entre sí pero inferior a la de la población global. La figura 5.4 muestra las aproximaciones de kernel a las densidades para los datos poblacionales para 6 grupos de 4 rutas cada uno (entre ellas las correspondientes a las rutas muestreadas), justificando gráficamente la afirmación anterior. Por otra parte, notemos que la población 3 ordenada tampoco se ajusta a los modelos para rutas propuestos.

Como datos adicionales para esta población señalemos que: su media poblacional es  $\bar{Y} = 80.03$ , su varianza poblacional 224.9 y su tamaño  $N = 10131$ .

#### Población 4

La población 4 se simuló (al igual que la población 5) ruta a ruta a diferencia de la población 3 que se simuló globalmente y luego se dividió en rutas. El mecanismo de simulación fue el siguiente:

1. Obtención de una muestra de tamaño 20 de una distribución Poisson con media 500, ( $\rightsquigarrow \{T_i\}_{i=1}^{20}$ ).
2. Obtención para cada ruta de una muestra de tamaño  $T_i$  de una normal de media  $\theta_i$  y varianza  $\sigma_i^2$ ,  $i = 1, 2, \dots, 20$ , siendo  $\theta_i = 50 + 3i$  y  $\sigma_i^2 = (1.5i)^2$ ,  $i = 1, 2, \dots, 20$ .

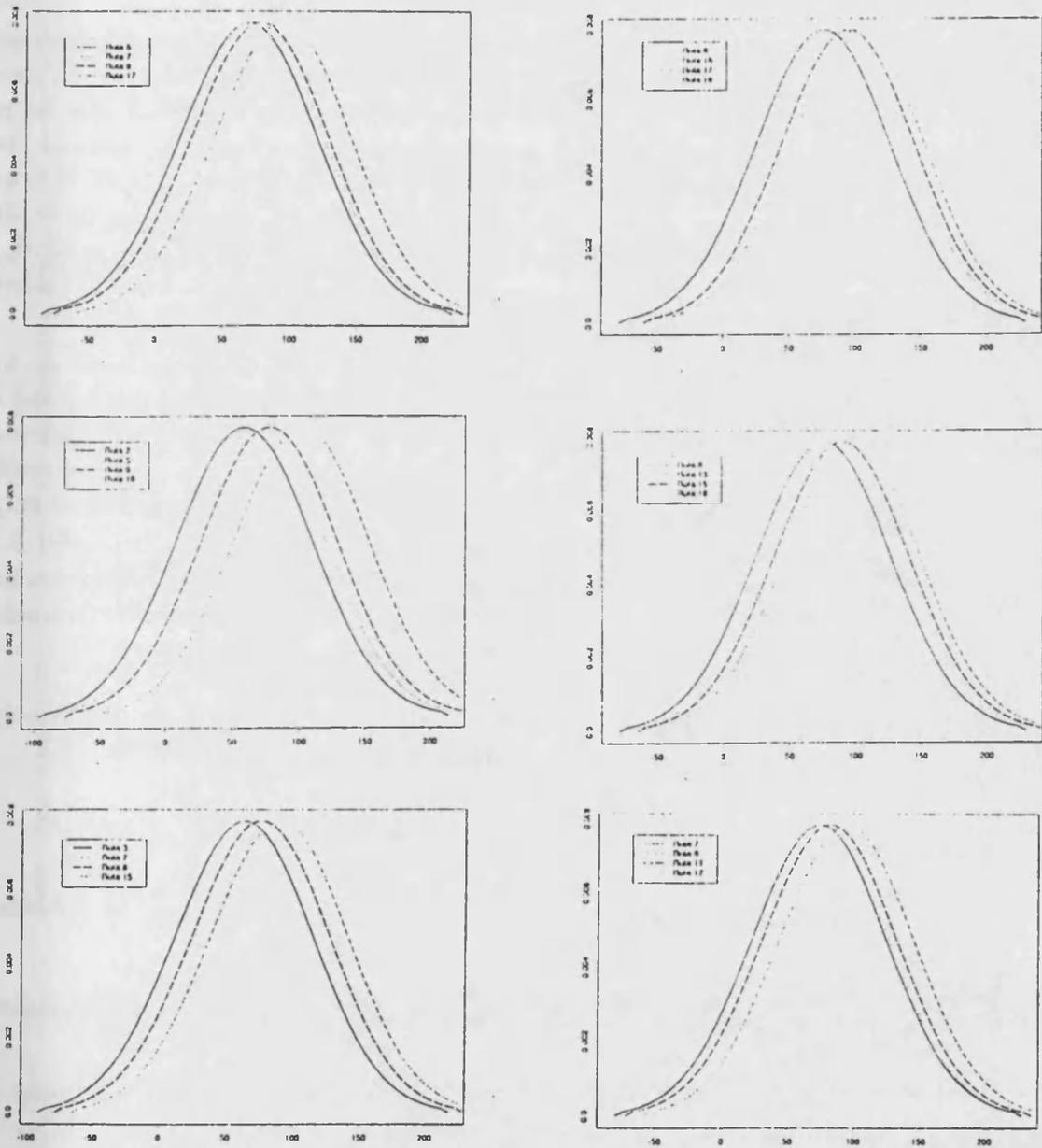


Figura 5.4: Comparación de cuatro rutas distintas de la población 3 ordenada

3. Las muestras obtenidas en el punto anterior, unidas y con su indicación de pertenencia a una ruta determinada constituyen la población 4.

La población 4 construida no responde ni a los modelos por rutas definidos ni a los modelos para m.a.s. Así:

- La población estudiada globalmente no ajusta a las hipótesis de los modelos para m.a.s. (elementos independientes e igualmente distribuidos según una normal), dado que podemos ver gráficamente, a través de la figura 5.5, que las representaciones del histograma y densidad no ajustan a un modelo normal (la densidad empírica no es siquiera simétrica), la mediana (representada en el boxplot) tampoco se sitúa en el centro y la representación de los cuantiles de la distribución muestral (línea) tampoco se ajustan a los cuantiles de una normal, obtenidos asumiendo que los datos son independientes e idénticamente distribuidos según una normal (puntos).
- Por otra parte, la población tampoco se ajusta a las hipótesis de los modelos para rutas, ya que si bien cada ruta se genera a partir de una normal común, las medias  $\{\theta_i\}_{i=1}^{20}$  y las varianzas  $\{\sigma_i^2\}_{i=1}^{20}$  no constituyen respectivamente muestras de una normal y una gamma invertida respectivamente. Esta situación la podemos ver gráficamente en la figura 5.6, que representa la densidad empírica para las muestra de las medias y precisiones ( $\tau_i = \frac{1}{\sigma_i^2}$ ,  $i = 1, 2, \dots, 20$ ) y las desviaciones en la representación de los cuantiles de la distribución para los datos de la población (línea) frente a los obtenidos asumiendo, respectivamente, que los datos son normales y<sup>12</sup>  $Ga(3, 300)$  (puntos).

Como datos adicionales para esta población señalemos que: su media poblacional es  $\bar{Y} = 77.58$ , su varianza poblacional 764.9 y su tamaño  $N = 10087$ .

## Población 5

El mecanismo de simulación aplicado para la construcción de la población 5 fue el siguiente:

1. Obtención de una muestra de tamaño 20 de una distribución Poisson con media 500, ( $\rightsquigarrow \{T_i\}_{i=1}^{20}$ ).
2. Obtención de una muestra de tamaño 20 de una distribución gamma invertida de parámetros 60 y 13000 (media aproximadamente  $15^2$ ), ( $\rightsquigarrow \{\sigma_i^2\}_{i=1}^{20}$ ).
3. Para  $i = 1, 2, \dots, 20$ , se obtiene  $\theta_i$  a partir de la simulación de una  $N(80, \sigma_i^2)$ , ( $\rightsquigarrow \{\theta_i\}_{i=1}^{20}$ ).
4. Obtención para cada ruta de una muestra de tamaño  $T_i$  de una normal de media  $\theta_i$  y varianza  $\sigma_i^2$ ,  $i = 1, 2, \dots, 20$ .

---

<sup>12</sup>los parámetros de la gamma se han elegido, a partir de los parámetros asumidos para la distribución inicial de las varianzas de cada ruta (ver el próximo subapartado)

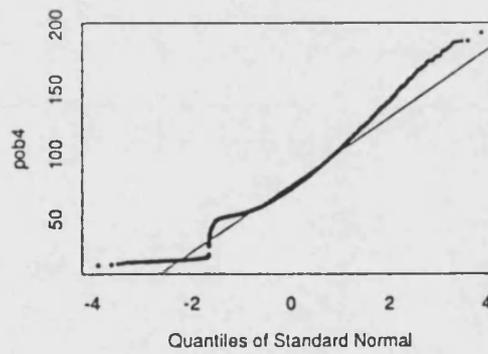
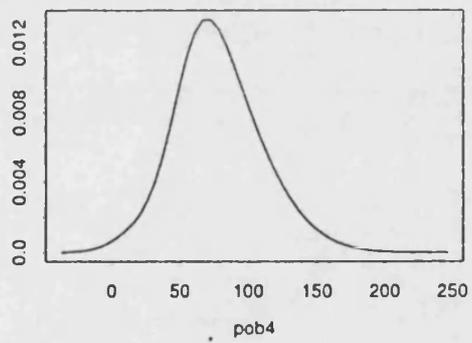
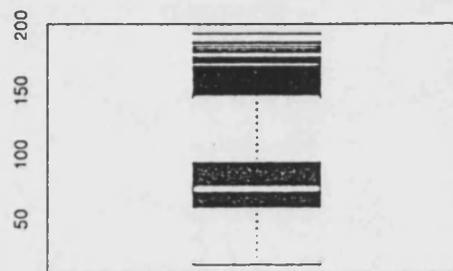
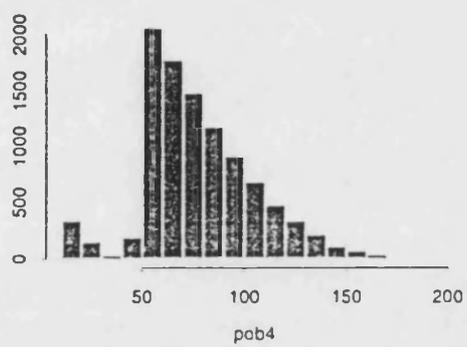


Figura 5.5: Estudio sobre el ajuste de la población 4 a los modelos m.a.s.

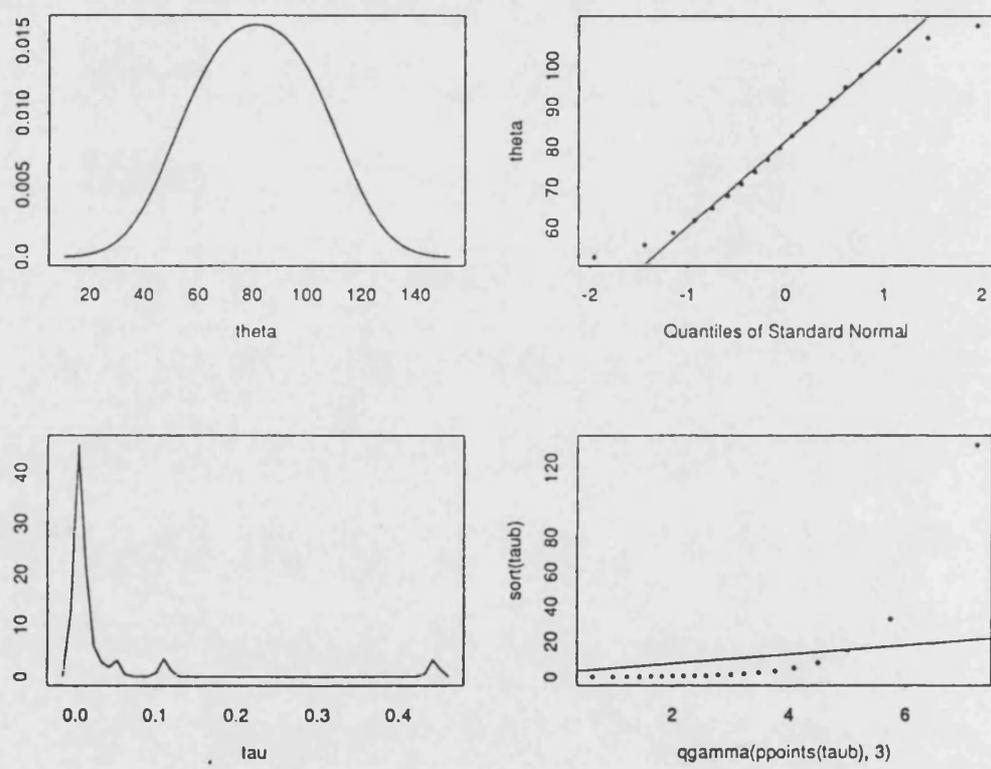


Figura 5.6: .Estudio sobre el ajuste de la población 4 a los modelos de rutas

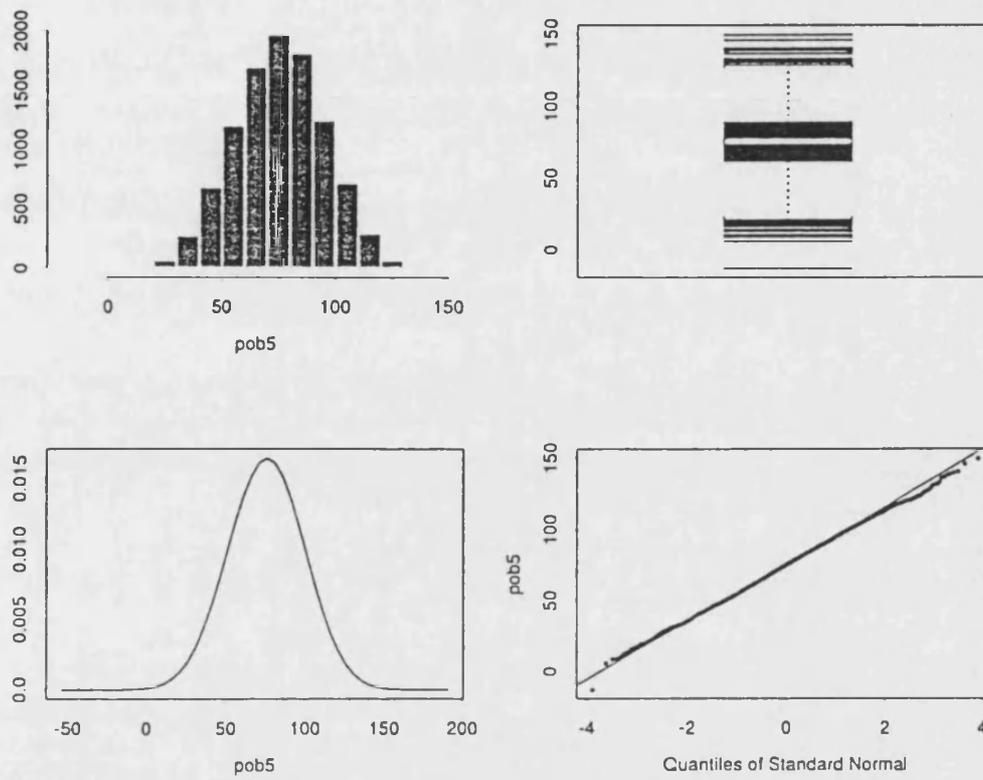


Figura 5.7: Estudio sobre el ajuste de la población 5 a los modelos de m.a.s

5. Las muestras obtenidas en el punto anterior, unidas y con su indicación de pertenencia a una ruta determinada constituyen la población 5.

La población 5, por construcción, se ajusta al modelo Bayes para rutas aleatorias y la población considerada globalmente es “aproximadamente” una población con elementos independientes e igualmente distribuidos normales (ver teorema 5.1, que proporciona el modelo de superpoblación condicionado a  $\mu$  y  $\mathbf{T}$  para este modelo), tal y como se aprecia en la figura 5.7 que muestra el histograma, la aproximación de kernel a la densidad de la población, la variabilidad de la mediana muestral y las diferencias entre los cuantiles de la población y los cuantiles asumiendo que la distribución de la muestra es normal.

Como datos adicionales para esta población señalemos que: su media poblacional es  $\bar{Y} = 75.36$ , su varianza poblacional 387.7 y su tamaño  $N = 10041$ .

### 5.4.2 Determinación de las distribuciones iniciales y mecanismos aplicados para la realización de las inferencias

Una vez obtenidas las poblaciones anteriores y las dos muestras para cada una de ellas, se procedió a la realización de inferencias para todas estas poblaciones asumiendo los cinco modelos descritos en la sección 5.2 y tomando como parámetros iniciales para los tres modelos de rutas los siguientes:

- $c_i = c = 1, i = 1, 2, \dots, 20$ .
- $\mu_0 = 0, \sigma_0^2 = 100^2$ ; ésto es, aproximadamente la distribución mínimo informativa para  $\mu$ .
- $a_0 = 3, b_0 = 300$ ; con el valor de  $a_0$  queda garantizada la existencia de los dos primeros momentos de todas las distribuciones predictivas y finales y el valor de  $b_0$  se toma de modo que  $E(\sigma_i^2) = \frac{b_0}{a_0 - 1} \approx 12^2$ . Para estos valores de  $a_0$  y  $b_0$ , la varianza es  $Var(\sigma_i^2) = \frac{b_0^2}{(a_0 - 1)^2(a_0 - 2)} \approx 150^2$ , siendo por lo tanto la distribución inicial asignada bastante conservadora.
- $\alpha = 300, \beta = 0.5$ ; con lo que  $E(\lambda) = \frac{\alpha}{\beta} = 600$  y  $Var(\lambda) = \frac{\alpha}{\beta^2} \approx 35^2$ , nuevamente, una distribución bastante conservadora.

Y para los modelos para muestras aleatorias podemos distinguir dos situaciones:

- en el caso del modelo Bayes para m.a.s. con varianza desconocida, se tomó como tamaño poblacional el auténtico tamaño de la población, aplicándose los resultados (5.86) y (5.87) para obtener los estimadores Bayes de  $\bar{Y}$  y  $\bar{Y}_u$  y sus varianzas dada la muestra, y
- en el caso del modelo Bayes para m.a.s. con varianza y tamaño poblacional desconocidos, se tomó como parámetros:  $\alpha = 300 + N_s, \beta = 0.5 + k = 4.5$  y  $t = K - k = 16$ , donde  $N_s = \sum_{i \in \{i_1, \dots, i_k\}} T_i$ , que corresponden a la distribución final,  $T_u | N_s$ , para los parámetros iniciales para  $\lambda$  dados con anterioridad para los modelos para rutas; aplicándose las expresiones obtenidas en (5.90) y (5.91) para obtener los estimadores Bayes de  $\bar{Y}$  y  $\bar{Y}_u$  y sus varianzas dada la muestra (y nuestros conocimientos realizado el muestreo por rutas sobre los tamaños de las rutas muestreadas).

(Notemos que se satisfacen las hipótesis para la existencia, en ambos casos, de los dos momentos).

Para la obtención de las inferencias dada la muestra y los tamaños de las rutas muestreadas en los modelos Bayes para rutas, se recurrió a la aplicación del método de muestreo por Gibbs implementado con el BUGS. El programa BUGS es aplicable en este caso por ser los tres modelos DAG, ser todas las condicionales necesarias para la aplicación del muestreo Gibbs (ver teoremas 5.4, 5.7 y 5.10), univariantes y de forma conocida por el programa salvo la distribución sobre  $T_u$ , que ya vimos que estaba definida por variables  $T_i, i \notin \{i_1, \dots, i_k\}$ , independientes con funciones de probabilidad log-concavas, que asimismo son simulables por las rutinas incluidas en el BUGS

y porque todas las distribuciones de la definiciones de los modelos también pueden expresarse a través de distribuciones univariantes conocidas. (Ver C.2.1, para una descripción más detallada de las condiciones que debe satisfacer un modelo para poder implementar Gibbs mediante el programa BUGS.)

El análisis con el programa BUGS (ver C.3) constó de tres fases: la traducción en BUGS de los tres modelos<sup>13</sup> (ver figura 5.8), fase de simulación y obtención de los estadísticos de interés a partir de una muestra de las distribuciones predictivas y finales. La fase de simulación constó de la realización de 1000 iteraciones Gibbs que se emplearon únicamente para la obtención del punto de arranque para nuestra cadena de Markov. A partir de dicho punto de arranque se inició para cada modelo el proceso de simulación Gibbs, valorándose la convergencia del método mediante el estadístico aproximado de Geweke(92)[39] proporcionado por el BUGS para las variables  $\bar{Y}_u$ ,  $T$ ,  $\mu$  y  $\sigma_i^2$  para algún  $i \notin \{1, 2, \dots, 4\}$  (y en el caso del modelo Bayes para rutas homocedástico  $\sigma^2$ ) que se obtuvo tras las primeras 5000 iteraciones del algoritmo Gibbs, y tras la realización de grupos adicionales de 2000 iteraciones más hasta que se considero que podíamos asumir la estabilidad de la cadena (estadístico aproximado de Geweke con un valor en módulo para las cuatro variables consideradas inferior a 1.96). En la mayor parte de los modelos y las muestras asumimos estabilidad (según este criterio) a las 7000 iteraciones desde el punto de arranque y en una minoría (señalada con una \* en las tablas) a las 9000.

Una vez asumida la estabilidad de la cadena de Markov, se realizaron 1000 iteraciones más para obtener así una muestra de tamaño 1000 de la distribución final del vector de interés dada la muestra y los tamaños muestrales conocidos, a partir de la cual se obtuvieron, los estimadores y desviación típica para:  $\bar{Y}$ ,  $\bar{Y}_u$ ,  $T$ ,  $\{\bar{Y}_{u_i}\}_{i \in \{1, 2, \dots, 4\}}$ ,  $\{\theta_i\}_{i \in \{1, 2, \dots, 4\}}$  (que no se obtienen obviamente para el modelo con medias iguales),  $\{\sigma_i^2\}_{i \in \{1, 2, \dots, 4\}}$  ( $\sigma^2$  para el modelo homocedástico),  $\mu$  y  $\theta_j$ ,  $\sigma_j^2$  (en los casos adecuados) y  $T_j$  para algún  $j \notin \{1, 2, \dots, 4\}$ ; todos estos estadísticos se obtuvieron con el comando `stats` del BUGS. En el caso particular del análisis de la población 5, hemos restringido el estudio al del modelo Bayes para rutas aleatorias y los dos modelos para muestras aleatorias. Para el modelo Bayes para rutas aleatorias se procedió a todo el análisis descrito en los dos últimos párrafos y además se obtuvo para la muestra 50 ( $n = 200$ ) una aproximación de kernel de las densidades y funciones de probabilidad (para los tamaños de las rutas y la población) de las distribuciones predictivas y finales marginales condicionadas a la muestra y los tamaños muestreados para  $T$  ( $N$ ),  $T_5$  ( $T[5]$ ),  $\bar{Y}_u$  ( $Y.medu$ ),  $\bar{Y}$  ( $Y.med$ ),  $\bar{Y}_{u_1}$  ( $Y.mdu[1]$ ),  $\bar{Y}_{u_2}$  ( $Y.mdu[2]$ ),  $\bar{Y}_{u_3}$  ( $Y.mdu[3]$ ),  $\bar{Y}_{u_4}$  ( $Y.mdu[4]$ ),  $\bar{Y}_{u_5}$  ( $Y.mdu[5]$ ),  $\theta_1$  ( $theta[1]$ ),  $\theta_2$  ( $theta[2]$ ),  $\theta_3$  ( $theta[3]$ ),  $\theta_4$  ( $theta[4]$ ),  $\theta_5$  ( $theta[5]$ ),  $\sigma_1^2$  ( $sigma2[1]$ ),  $\sigma_2^2$  ( $sigma2[2]$ ),  $\sigma_3^2$  ( $sigma2[3]$ ),  $\sigma_4^2$  ( $sigma2[4]$ ) y  $\sigma_5^2$  ( $sigma2[5]$ ), (en una población que por su construcción cumple todas las hipótesis del modelo Bayes para rutas aleatorias) mediante el programa de Splus `drawdat` incluido en el BUGS.

Una vez obtenidos todos estos resultados se procedió a su análisis según tres líneas de trabajo:

<sup>13</sup>para facilitar las notaciones se han reordenado las rutas de modo que los primeros 4 índices corresponden a las rutas muestreadas y los restantes 16 índices a las rutas no muestreadas

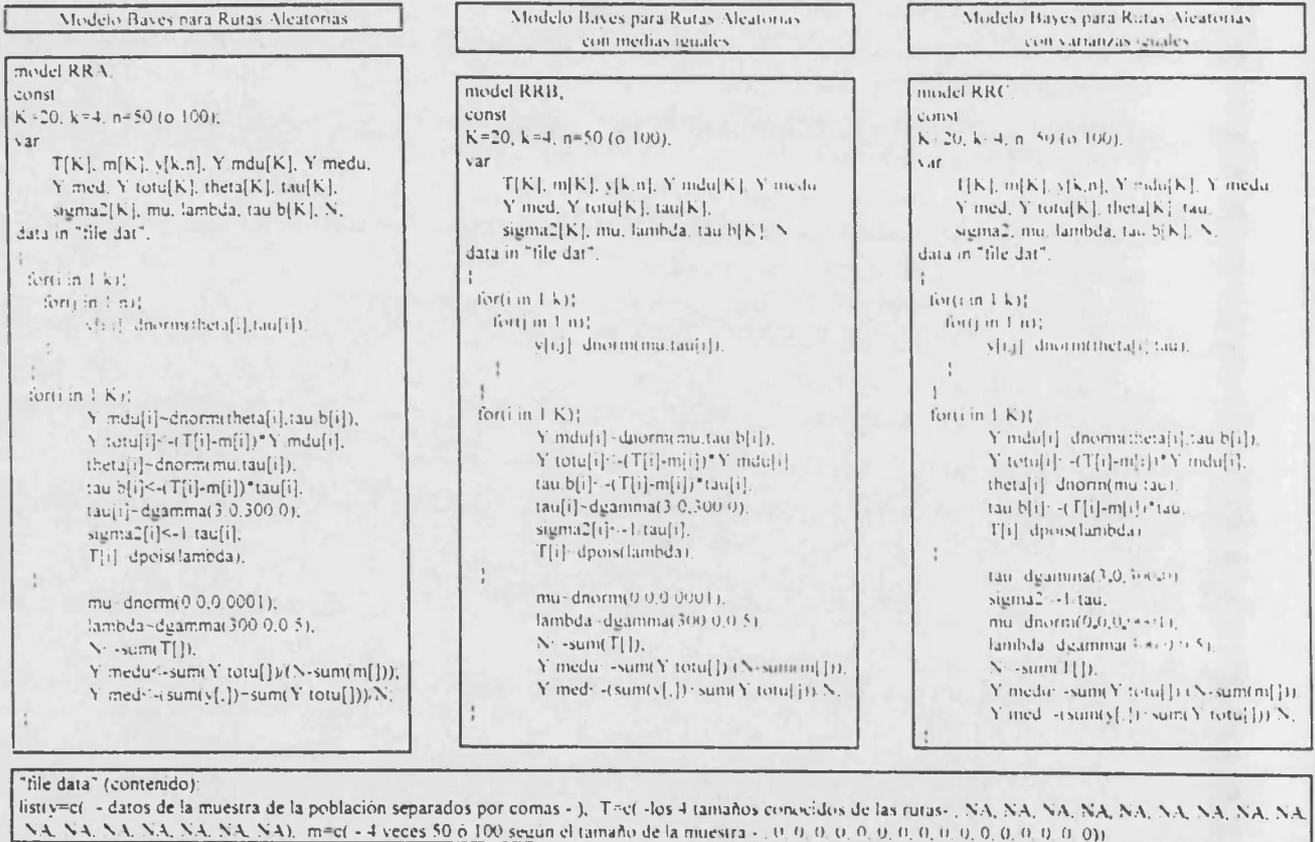


Figura 5.8: Traducción en BUGS de los tres modelos de rutas

1. Comparación de los estimadores Bayes de  $\bar{Y}_u$  y  $\bar{Y}$  y su desviación típica (D.T.) para los tres modelos de rutas y para los dos modelos para muestras aleatorias entre sí y con el auténtico valor de la media poblacional (en este caso conocido) para cada una de las tres poblaciones (subsección 5.4.3).
2. Análisis de los estimadores Bayes para  $T$ ,  $\{\bar{Y}_{u_i}\}_{i=1}^4$ ,  $\{\theta_i\}_{i=1}^4$  (que no se obtienen obviamente para el modelo con medias iguales),  $\{\sigma_i^2\}_{i=1}^4$  ( $\sigma^2$  para el modelo homocedástico),  $\mu$  y  $\theta_j$ ,  $\sigma_j^2$  (en los casos adecuados) y  $T_j$  para algún  $j \notin \{1, 2, 3, 4\}$  (subsección 5.4.4).
3. Estudio de las distribuciones predictivas y finales para las variables del párrafo anterior a través de las aproximaciones de kernel obtenidas a partir de la muestra de las distribuciones predictivas y finales obtenidas por el método Gibbs para el modelo de rutas aleatorias y la muestra 50 de la población 5 (subsección 5.4.4).

### 5.4.3 Exposición y discursión de los resultados para la media no observada y media poblacionales

Centrándonos en la primera línea de trabajo, para la exposición de los resultados nos hemos apoyado en una tabla que contiene para las variables  $\bar{Y}_u$  y  $\bar{Y}$  los valores de su estimador Bayes y desviación típica de la variable condicionada a la muestra y tamaños de las rutas muestreadas (esto es, la raíz cuadrada de la varianza) para cada uno de los cinco modelos implementados (“MBRR”=Modelo Bayes para rutas aleatorias, “MBRR=Med.” =Modelo Bayes para rutas aleatorias con medias iguales, “MBRR=Var.” =Modelo Bayes para rutas aleatorias homocedásticas, “MAS” = Modelo Bayes para m.a.s con varianza desconocida y “MASE” = Modelo Bayes para m.a.s con varianza y tamaño poblacional desconocidos) y que en el caso de  $\bar{Y}$  señala adicionalmente si el valor de la media poblacional (conocido en las tres poblaciones) pertenece ( $\in$ ) o no pertenece ( $\notin$ ) al intervalo de confianza  $\hat{Y} \pm 2Var^{\frac{1}{2}}(\bar{Y}|(s, y_s), \{T_i\}_{i=1}^4)$  (que si la distribución de  $\bar{Y}|(s, y_s), \{T_i\}_{i=1}^4$  fuera normal correspondería aproximadamente al intervalo de confianza al 95 %). Estos resultados se han expuesto para las dos muestras obtenidas para cada población; la muestra 50, se refiere a la muestra obtenida a partir de la selección de 50 elementos para cada una de las 4 rutas seleccionadas (en total, el tamaño de la muestra es 200) y la muestra 100, a la muestra obtenida por la selección de 100 elementos para cada una de las 4 rutas seleccionadas (en total, el tamaño de la muestra es 400). Los resultados obtenidos para cada una de las poblaciones han sido los siguientes:

#### Población 3

Hemos visto en la subsección 5.4.1 que los datos simulados para esta población daban lugar a dos poblaciones: la población 3 no ordenada y la población 3 ordenada con características distintas, por lo que atendiendo a esta dualidad expondremos los resultados en dos tablas: la tabla 5.1 para los resultados para la población 3 no ordenada y la tabla 5.2 para los de la población 3 ordenada.

		$\bar{y}_0$		$\bar{y}$		
		ESTIMADOR	D T	ESTIMADOR	D T	ESTIM:2DT
M U E S T R A 50	MBRR	81.43	5.953	81.43	5.838	€
	MBRR =Med	81.58	0.9745	81.58	0.9556	€
	MBRR =Var.	81.36	6.411	81.36	6.287	€
	MASe	81.53	0.4392	81.53	0.9706	€
	MAS	81.53	0.9899	81.53	0.9704	€
M U E S T R A 100	MBRR	79.22	6.855	79.23	6.590	€
	MBRR =Med	79.70	0.7748	79.70	0.7448	€
	MBRR =Var.	79.62	7.331	79.62	7.047	€
	MASe	79.69	0.3576	79.69	0.7749	€
	MAS	79.69	0.8064	79.69	0.7746	€

El número de simulaciones realizadas hasta asumir estabilidad han sido 7000. Las estadísticas se obtienen a partir de 1000 simulaciones una vez asumida la estabilidad.

Tabla 5.1: Resultados de la población 3 no ordenada

		$\bar{y}_0$		$\bar{y}$		
		ESTIMADOR	D T	ESTIMADOR	D T	ESTIM:2DT
M U E S T R A 50	MBRR	75.23	3.024	75.23	2.966	€
	MBRR =Med	74.45	0.4140	74.47	0.4060	€
	MBRR =Var.	75.33	1.065	75.33	1.044	€
	MASe	75.32	0.3033	75.32	0.6705	€
	MAS	75.32	0.6838	75.32	0.6703	€
M U E S T R A 100	MBRR	75.23	2.885	75.23	2.774	€
	MBRR =Med *	74.48	0.2371	74.51	0.2279	€
	MBRR =Var.	75.36	0.7944	75.36	0.7637	€
	MASe	75.33	0.2145	75.33	0.4658	€
	MAS	75.33	0.4848	75.33	0.4657	€

El número de simulaciones realizadas hasta asumir estabilidad han sido 7000, salvo en los modelos y muestras seleccionados por \* en cuyo caso han sido 9000. Las estadísticas se obtienen a partir de 1000 simulaciones una vez asumida la estabilidad.

Tabla 5.2: Resultados de la población 3 ordenada

Del análisis de los resultados para la población 3 no ordenada se desprende que:

- los estimadores Bayes de los cinco modelos para la media no observada y media poblacionales son muy similares en las dos muestras,
- los estimadores Bayes para la media poblacional, para esta simulación, para la muestra 50 sobrevaloran en  $\approx 1.5$  unidades el auténtico valor de la media poblacional y para la muestra 100 lo infravaloran en  $\approx 0.5$  unidades,
- los modelos Bayes para rutas aleatorias y para rutas aleatorias homocedásticas sobreestiman el error cometido en la estimación. Ello no obstante para todos ellos el verdadero valor de la media poblacional pertenece al intervalo de confianza estudiado para la media poblacional.

Estos resultados muestran que los estimadores Bayes obtenidos para los modelos de rutas aleatorias son muy semejantes a los estimadores Bayes para los modelos de muestras aleatorias cuando la muestra obtenida por un muestreo por rutas es prácticamente una muestra aleatoria de la población (recordemos las características de la población 3 no ordenada que comentábamos en la subsección 5.4.1). Los modelos para rutas es este caso son menos adecuados que el modelo para m.a.s, pero están cumpliendo su tarea de diagnóstico de la muestra puesto que al proporcionar resultados similares a los de los modelos m.a.s nos están indicando que la muestra analizada es aproximadamente una muestra aleatoria de la población (y por construcción sabemos que efectivamente la muestra es aleatoria).

Y del análisis de los resultados para la población 3 ordenada que:

- los estimadores Bayes de los cinco modelos para la media no observada y media poblacional son similares aunque no tanto como en la población 3 no ordenada sobre todo si comparamos con el estimador Bayes para el modelo Bayes para rutas aleatorias con medias iguales,
- los estimadores Bayes para la media poblacional, para esta simulación, infravaloran en  $\approx 5$  unidades el auténtico valor de la media poblacional para ambas muestras,
- todos los modelos propuestos (especialmente los modelos para muestras aleatorias y el modelo para rutas con medias iguales), subestiman el error cometido en la estimación de modo que, únicamente en el caso del modelo Bayes para rutas aleatorias, el verdadero valor de la media poblacional pertenece al intervalo de confianza estudiado.

En el caso de la población 3 ordenada, ya comentábamos que la muestra por rutas obtenida ya no podía asumirse como una muestra aleatoria de la población (dado que las rutas entre sí constituían muestras de normales con medias distintas tal y como se veía en la figura 5.4). Estas desviaciones de la muestra respecto al concepto de muestra aleatoria, aunque no muy grandes, ya empiezan a ser detectadas por nuestros modelos por rutas puesto que se observan diferencias (aunque no demasiado grandes) entre los estimadores proporcionados por estos modelos comparados entre sí y con respecto a

		$\hat{Y}_0$		$\bar{Y}$		
		ESTIMADOR	DT	ESTIMADOR	DT	ESTIMc2DT
M U E S T R A 50	MBRR*	71.59	5.362	71.75	5.257	€
	MBRR =Med	70.21	1.267	70.39	1.242	€
	MBRR =Var	79.11	8.231	79.11	8.070	€
	MASe	79.37	0.7364	79.37	1.629	€
	MAS	79.37	1.662	79.37	1.629	€
M U E S T R A 100	MBRR	70.62	4.949	70.95	4.755	€
	MBRR =Med	68.52	0.6918	68.93	0.6647	€
	MBRR =Var *	78.32	7.664	78.35	7.362	€
	MASe	79.01	0.4933	79.01	1.069	€
	MAS	79.01	1.113	79.01	1.069	€

El número de simulaciones realizadas hasta asumir estabilidad han sido 2000, salvo en los modelos y muestras señalados por \* en cuyo caso han sido 3000. Las estadísticas se obtienen a partir de 1000 simulaciones una vez asumida la estabilidad.

Tabla 5.3: Resultados de la población 4

los modelos para m.a.s. Debemos destacar también, que el valor de la varianza de las medias poblacional y poblacional no observada dada la muestra y los tamaños de las rutas muestreadas proporcionado por el modelo Bayes para rutas aleatorias es más fiable como medidor del error de nuestras estimaciones que el proporcionado por los modelos para muestras aleatorias; en este sentido el modelo Bayes para rutas aleatorias es el único modelo (de los cinco estudiados) que proporciona un intervalo de confianza para la media que contiene el verdadero valor de la media poblacional.

### Población 4

Los resultados para la media no observada y media poblacionales para los cinco modelos y las dos muestras de esta población aparecen reflejados en la tabla 5.3 y de ellos se desprende que:

- las diferencias entre los estimadores Bayes de los cinco modelos para la media no observada y media poblacional son considerables (de hasta casi 9 unidades),
- para esta simulación, el mejor estimador de la media poblacional es el proporcionado por el modelo Bayes para rutas aleatorias homocedásticas,
- las diferencias entre los errores estimados por cada modelo son importantes entre sí,
- también para esta simulación, el verdadero valor de la media poblacional pertenece al intervalo de confianza de todos los modelos exceptuando el modelo Bayes para

		$\bar{y}_0$		$\bar{y}$		
		ESTIMADOR	D T	ESTIMADOR	D T	ESTIM=2DT
M U E S T R A 50	MBRR	72.50	6.569	72.53	6.438	€
	MASc	74.08	0.5916	74.08	1.3089	€
	MAS	74.08	1.3355	74.08	1.3089	€
M U E S T R A 100	MBRR	72.74	6.528	72.79	6.268	€
	MASc	73.91	0.4427	73.91	0.9596	€
	MAS	73.91	0.9905	73.91	0.9597	€

El número de simulaciones realizadas hasta alcanzar estabilidad han sido 7000. Las estadísticas se obtienen a partir de 1000 simulaciones para ver a modo de estabilidad.

Tabla 5.4: Resultados de la población 5

rutas aleatorias con medias iguales.

Ya comentábamos que la población 4 no se ajustaba a ninguno de los modelos propuestos en este capítulo. En este caso, nuestros modelos para rutas detectan ese desajuste entre las hipótesis y la realidad (modelos con hipótesis bastante similares entre sí dan resultados muy distintos) y nos alertan así sobre la precisión de los resultados de estimación que estamos obteniendo.

### Población 5

En esta población estudiaremos los resultados para la media no observada y media poblacionales únicamente para el modelo Bayes para rutas aleatorias y los dos modelos para muestras aleatorias. Estos resultados aparecen en la tabla 5.4 y de ellos se desprende:

- los estimadores Bayes de los cinco modelos para la media no observada y media poblacional difieren entre sí aproximadamente una unidad (mayores diferencias que en la población 3 ordenada y menores que en la población 4),
- el mejor estimador de la media poblacional, para esta simulación, es el proporcionado por los modelos para muestras aleatorias,
- el verdadero valor de la media poblacional pertenece al intervalo de confianza calculado para los tres modelos.

La población 5, por construcción, se ajusta al modelo Bayes para rutas aleatorias definido en (5.1) y, a la vez al modelo Bayes definido en los modelos para muestras

aleatorias ( $Y \sim N_N(\mathbf{1}_N\mu, \mathbf{I}_N\sigma^2)$ ), ver figura 5.7; de esta dualidad se explica por una parte (primera afirmación), las diferencias existentes entre los estimadores (las diferencias detectan que la muestra es una muestra de rutas y no una muestra aleatoria) y a la vez (por la segunda afirmación), los buenos resultados que proporcionan los modelos para muestras aleatorias.

Del análisis de las poblaciones 3, 4 y 5 podemos deducir que:

- Los modelos Bayes para rutas aleatorias propuestos detectan cuando la muestra se aleja de una muestra aleatoria de la población. Esta detección se observa a través de las diferencias entre los estimadores Bayes obtenidos por los tres modelos y en especial entre los estimadores Bayes para los modelos Bayes para rutas aleatorias y rutas aleatorias homocedásticas en relación con el estimador Bayes para rutas aleatorias con medias iguales; estas diferencias, según nuestras simulaciones, parecen seguir un escala gradual (ver poblaciones 3 ordenada, 4 y 5).
- Los modelos Bayes para rutas propuestos también detectan de manera análoga, cuando la población no responde a la hipótesis de elementos para la característica en estudio independientes e igualmente distribuidos según una normal (ver población 4).
- Se confirma la robustez de los estimadores Bayes para los modelos para muestras aleatorias.

En conclusión, los modelos para rutas propuestos se manifiestan como un instrumento de análisis sencillo para comprobar si la población se ajusta a nuestras hipótesis de elementos independiente e igualmente distribuidos normales y detectar los posibles errores en la valoración de la precisión al emplear el estimador media muestral cuando la muestra empleada es una muestra obtenida a partir de rutas aleatorias y por lo tanto no es muestra aleatoria de la población. Si las diferencias entre las inferencias son grandes, el estudio indica la necesidad de una modelización cuidadosa de las rutas aleatorias, pero en cualquier caso, el estimador Bayes para el modelo Bayes de rutas aleatorias (un estimador puntual que parece ser razonablemente robusto, según se desprende de su comportamiento para las distintas poblaciones simuladas) junto con una medida de su fiabilidad deducida de la desviación típica (o de la varianza) proporcionan una inferencia razonable que supone una mejora sustancial respecto a la práctica habitual.

#### **5.4.4 Exposición y análisis de las inferencias sobre otras variables de los modelos de rutas aleatorias y sobre las distribuciones predictivas y finales dada la muestra y los tamaños de las rutas muestreadas**

En esta subsección estudiaremos algunas características de los modelos de rutas aleatorias a partir de los resultados obtenidos por el análisis de las tres poblaciones anteriores.

Para realizar este estudio nos apoyaremos en primer lugar en los estimadores Bayes de  $T$ ,  $\{\bar{Y}_{u_i}\}_{i=1}^4$ ,  $\{\theta_i\}_{i=1}^4$  (que no se obtienen obviamente para el modelo con medias iguales),  $\{\sigma_i^2\}_{i=1}^4$  ( $\sigma^2$  para el modelo homocedástico),  $\mu$  y  $\theta_j$ ,  $\sigma_j^2$  (en los casos adecuados) y  $T_j$  para algún  $j \notin \{1, 2, 3, 4\}$ , que aparecen resumidas en la tabla 5.5 para las dos muestras ( $n_i = 50$  representa la muestra 50 y  $n_i = 100$ , la muestra 100) de las poblaciones 3, 4 y 5 y cada uno de los modelos de rutas aleatorias estudiados (“General”=Modelo Bayes para rutas aleatorias, “=Media.”=Modelo Bayes para rutas aleatorias con medias iguales y “=Varian.”=Modelo Bayes para rutas aleatorias homocedásticas).

Entre las características para las distribuciones predictivas y finales que se deducen de esta tabla señalemos las siguientes (que se pueden justificar fácilmente a partir de los teoremas 5.2, 5.5 y 5.8):

- La coincidencia para los dos tamaños de muestra y los tres modelos para un población dada de los estimadores Bayes de  $T$  y de  $T_5$ , como consecuencia de la independencia de las distribuciones posteriores de  $\mathbf{T}_u$  y de  $T_u$  respecto a la muestra y por ser estas distribuciones comunes para los tres modelos de rutas (ver (5.19)-(5.20), (5.44)-(5.45) y (5.65)-(5.66)).
- La coincidencia de los estimadores de  $\{\bar{Y}_{u_i}\}_{i=1}^5$  y de  $\mu$  en el modelo Bayes para rutas aleatorias con medias iguales para cada una de las dos muestras de las poblaciones 3 y 4, como consecuencia (ver (5.40)) de que:

$$E(\bar{Y}_{u_i} | (s, \mathbf{y}_s), \{T_i\}_{i=1}^4) = E^{\mathbf{T}_u | T_s} E^{\mu | (s, \mathbf{y}_s)} E(\bar{Y}_{u_i} | (s, \mathbf{y}_s), \mu, \mathbf{T}) = E(\mu | (s, \mathbf{y}_s))$$

para  $i = 1, 2, \dots, 20$ .

- La coincidencia de los estimadores Bayes de  $\{\sigma_i^2\}_{i=5,7}$  para los modelos Bayes para rutas aleatorias y para rutas aleatorias con medias iguales para las poblaciones 3 y 4 (las diferencias observables son pequeñas y se deben al muestreo por Gibbs), como consecuencia (ver (5.17) y (5.42)) de que para los dos modelos:

$$E(\sigma_i^2 | (s, \mathbf{y}_s)) = E^{\mu | (s, \mathbf{y}_s)} E(\sigma_i^2 | (s, \mathbf{y}_s), \mu, \mathbf{T}) = E(\mu | (s, \mathbf{y}_s)) = \frac{b_0}{a_0 - 1} = 150$$

para  $i = 5, 6, \dots, 20$ .

- Y la coincidencia de los estimadores de  $\theta_i$  con los de  $\bar{Y}_{u_i}$ ,  $i = 1, 2, \dots, 20$  (por ruta) para las poblaciones 3 y 4, en los modelos Bayes<sup>14</sup> para rutas aleatorias y para rutas aleatorias homocedásticas, debida a que (ver (5.15)-(5.24) y (5.61)-(5.70)):

$$E(\bar{Y}_{u_i} | (s, \mathbf{y}_s), \mu, \mathbf{T}) = E(\theta_i | (s, \mathbf{y}_s), \mu) = \begin{cases} \tilde{\theta}_i, & i \in \{1, 2, 3, 4\} \\ \mu, & i \notin \{1, 2, 3, 4\} \end{cases}$$

con  $\tilde{\theta}_i$  común en los dos modelos por ser  $c_i = c = 1$ ,  $i = 1, 2, \dots, 20$ .

<sup>14</sup>por separado, dado que la media de la distribución final de  $\mu$  no coincide para los dos modelos

	n	POBLACION 3 (NO ORD)			POBLACION J (ORD)			POBLACION 4			POB 5
		General	-Media	-Varian	General	-Media	-Varian	General	-Media	-Varian	General
I	50	10340	10340	10350	10340	10340	10350	10210*	10190	10200	10040
	100	10350	10340	10350	10350	10350*	10350	10200	10260	10190*	10040
T <sub>3</sub>	50	519.1	519.8	519.2	519.1	519.8	519.2	514.0*	512.3	512.2	504.1
	100	520.7	517.8	519.4	520.7	520.0*	519.4	512.2	513.1	512.4*	504.9
$\bar{f}_n$	50	81.08	81.59	80.99	62.96	74.46	62.95	63.86*	70.20	64.08	72.25
	100	79.47	79.76	79.45	62.97	74.47*	62.97	65.90	68.52	65.72*	69.71
$\bar{f}_{u_1}$	50	81.38	81.58	81.26	73.67	74.45	73.65	72.34*	70.22	72.52	75.91
	100	79.39	79.69	79.32	73.63	74.48*	73.62	69.74	68.53	69.90*	76.30
$\bar{f}_{u_2}$	50	80.21	81.54	80.10	75.52	74.44	75.50	76.78*	70.21	76.97	59.05
	100	80.99	79.67	81.13	75.46	74.48*	75.48	77.92	68.50	77.95*	57.72
$\bar{f}_{u_3}$	50	83.52	81.59	83.38	89.12	74.46	89.09	103.9*	70.27	104.1	88.81
	100	78.90	79.60	79.01	89.25	74.49*	89.26	102.5	68.49	102.4*	91.00
$\bar{f}_{u_4}$	50	81.41	81.59	80.98	75.21	74.46	75.28	70.14*	70.21	79.35	72.07
	100	79.55	79.73	79.34	75.63	74.49*	75.34	68.36	68.52	78.40*	72.10
$\hat{\mu}_1$	50	81.08		81.03	62.96		62.96	63.88*		64.05	72.21
	100	79.49		79.43	62.98		62.97	65.60		65.74*	69.73
$\hat{\mu}_2$	50	81.37		81.30	73.66		73.66	72.35*		72.55	75.90
	100	79.37		79.33	73.63		73.62	69.71		69.90*	76.31
$\hat{\mu}_3$	50	80.22		80.08	75.52		75.50	76.78*		76.95	59.04
	100	80.98		81.09	75.46		75.48	77.90		77.95*	57.72
$\hat{\mu}_4$	50	83.53		83.37	89.12		89.09	103.9*		104.1	88.82
	100	78.87		78.95	89.24		89.26	102.5		102.4*	91.00
$\hat{\sigma}_1$	50	81.42		81.01	75.23		75.28	70.11*		79.38	72.08
	100	79.53		79.26	75.62		75.34	68.33		78.42*	72.10
$\hat{\sigma}_2$	50	209.7	208.0	187.6	14.80	138.6	5.169	66.71*	105.8	305.0	282.0
	100	275.5	277.8	248.0	8.137	137.3*	2.897	40.39	49.14	266.6*	238.6
$\hat{\sigma}_3$	50	168.7	170.0		11.45	12.18		135.4*	130.6		132.7
	100	204.9	206.9		6.101	6.961*		130.9	132.7		168.7
$\hat{\sigma}_4$	50	199.9	200.7		11.49	12.53		198.1*	243.4		182.8
	100	229.0	233.9		6.006	7.040*		217.2	302.6		210.5
$\hat{\sigma}_{-1}$	50	166.5	168.3		15.46	218.4		785.5*	1842		300.4
	100	263.8	266.4		8.235	219.2*		668.0	1793		282.8
$\hat{\sigma}_{-2}$	50	152.1	154.0		152.1	154.0		151.7*	150.0		148.0
	100	149.0	148.1		149.0	150.6*		146.6	152.2		143.6
$\hat{\sigma}_{-3}$	50	144.0	159.3		144.0	159.3		150.4*	145.3		148.7
	100	150.7	151.3		150.7	146.6*		152.0	148.0		149.4
u	50	81.26	81.58	81.29	75.07	74.45	75.33	70.03*	70.21	79.10	72.24
	100	79.07	79.70	79.40	75.15	74.48*	75.35	69.23	68.52	78.27*	72.65

El número de simulaciones realizadas hasta asumir estabilidad han sido 7000, salvo en los muestreos y muestras seleccionados por  $\hat{\sigma}$  en cuyo caso han sido 6000. Las estadísticas se obtienen a partir de 1000 simulaciones una vez asumida la estabilidad.

Tabla 5.5: Otras inferencias sobre las poblaciones 3, 4 y 5

Las distribuciones predictivas y finales (condicionadas a la muestra y los tamaños de las rutas muestreadas) para los modelos para rutas aleatorias no presentan un forma estándar y el cálculo de sus primeros momentos suele presentar problemas de computación, tal y como veíamos en la sección 5.3. A título orientativo sobre cómo son estas distribuciones se han obtenido a partir de la función en Splus `drawdat`, incluida en el programa BUGS las aproximaciones de kernel de las mismas a partir de las muestras simuladas por el muestreo Gibbs para el modelo Bayes para rutas aleatorias y la muestra 50 de la población 5 (que es por construcción la que mejor ajusta al modelo Bayes para rutas aleatorias). Los resultados gráficos se presentan en las figuras 5.9, 5.10, 5.11, 5.12 y 5.13; notemos a partir de ellos:

- La forma acampanada y prácticamente simétrica de las distribuciones predictivas de  $\bar{Y}$ ,  $\bar{Y}_u$ ,  $\bar{Y}_{u_i}$ ,  $i = 1, 2, \dots, 5$  y las finales marginales de  $\mu$  y  $\theta_i$ ,  $i = 1, 2, \dots, 5$ .
- Si nos fijamos en los cuantiles al 95 % estimados para las distribuciones predictivas de  $\bar{Y}$ ,  $\bar{Y}_u$ ,  $\bar{Y}_{u_i}$ ,  $i = 1, 2, \dots, 5$  y las finales marginales de  $\mu$  y  $\theta_i$ ,  $i = 1, 2, \dots, 5$  en comparación con el cuantil de la correspondiente normal, tenemos que, para esta simulación, las distribuciones predictivas para  $\bar{Y}$  y  $\bar{Y}_u$  son más apuntadas que las correspondientes normales, las distribuciones de  $\bar{Y}_{u_5}$  y  $\theta_5$  algo menos apuntadas y las restantes con un apuntamiento muy similar al de la normal (son aproximadamente normales).
- La forma acampanada de  $T$  y  $T_i$ ,  $i = 5, 6, \dots, 20$ ; respecto a este punto señalemos que la densidad de kernel no aproxima en esta caso a la función de densidad de estas variables (que son discretas) sino que une por líneas a su función de probabilidad (definida unicamente para los enteros positivos y cero).
- Y la forma más amplia de las distribuciones finales de  $\sigma_i^2$ ,  $i = 1, 2, 3, 4$ , aunque asimétrica (por efecto de la información introducida por la muestra) y la forma de la distribución de  $\sigma_5^2$ , que corresponde obviamente (ver (5.17)) a una  $Ga^{-1}(3, 300)$ .

Adicionalmente, a la derecha de las aproximaciones de kernel a las funciones de densidad (y probabilidad) se presenta una gráfica con la traza de los valores simulados en las últimas 1000 iteraciones (con las que se obtiene la muestra). A través de ella podemos valorar gráficamente la estabilización de la cadena de Markov.

## 5.5 Comentarios finales

Como comentario final al capítulo destacar dos ideas:

1.- Los modelos Bayes para rutas aleatorias aunque simples constituyen una herramienta eficaz (ver sección 5.4) para detectar en qué medida una muestra obtenida por un muestreo por rutas (de las condiciones descritas en la sección 5.2) se aproxima a un muestreo aleatorio de una población. Proporcionando, además, el modelo para rutas aleatorias un estimador Bayes razonablemente robusto junto con una medida de su precisión (deducida de la varianza condicionada a la muestra y los tamaños de

## Output for Predictive and Posterior Distributions analysis

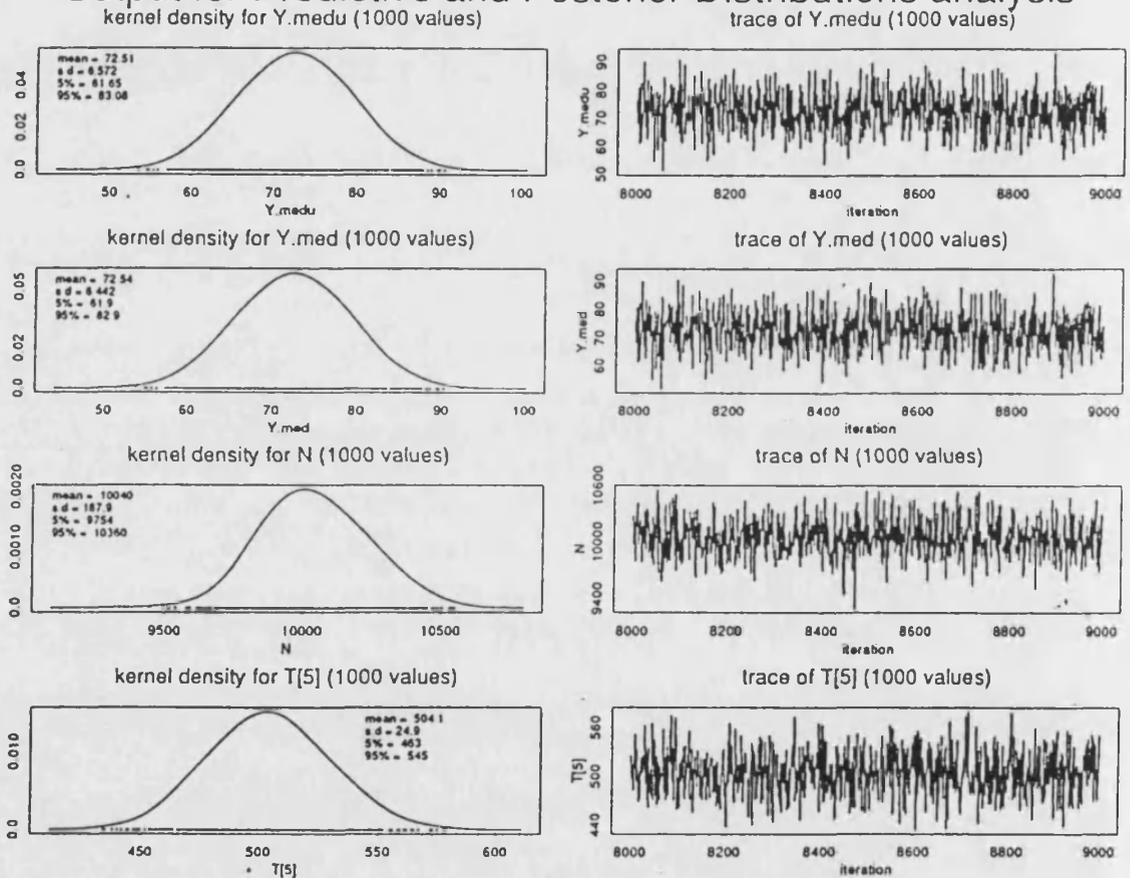


Figura 5.9: Distribuciones predictivas y finales dada la muestra y los tamaños de las rutas muestreadas

## Output for Predictive and Posterior Distributions (2) analysis

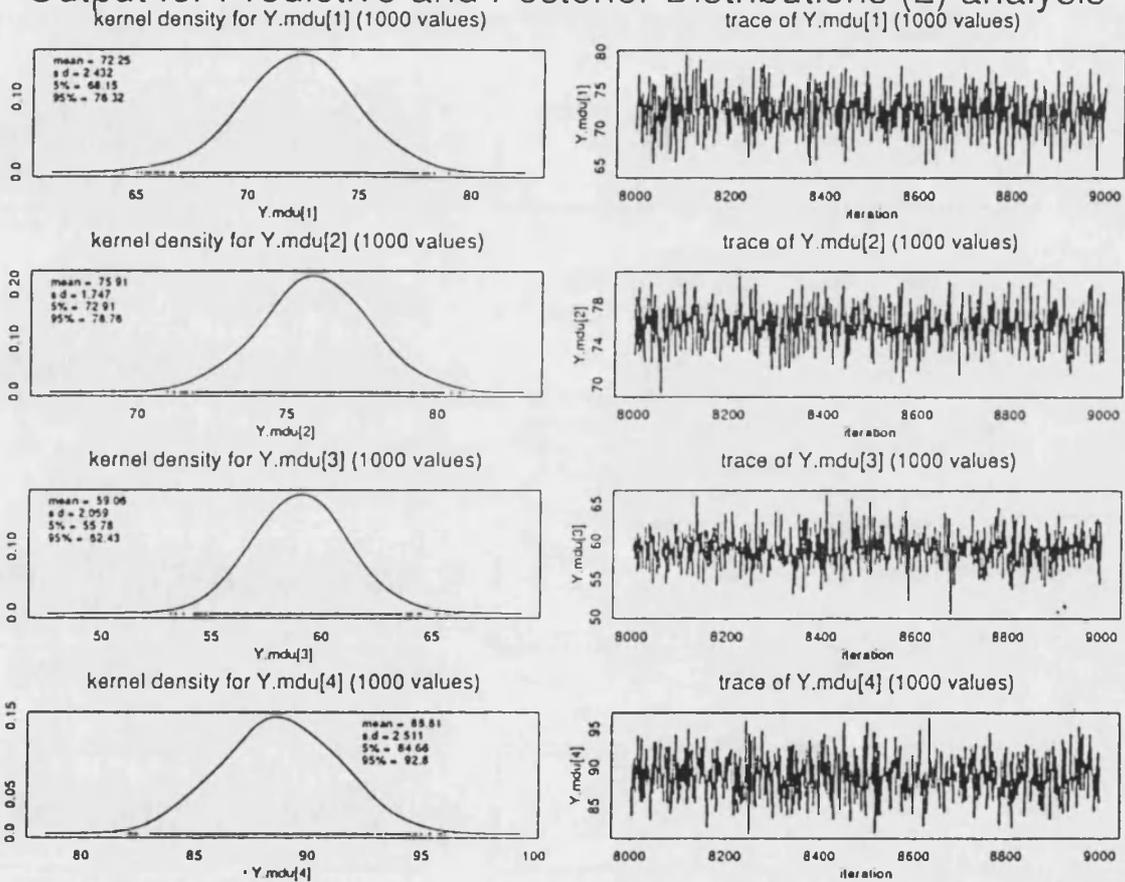


Figura 5.10: Distribuciones predictivas y finales dada la muestra y los tamaños de las rutas muestreadas (2)

### Output for Predictive and Posterior Distributions (3) analysis

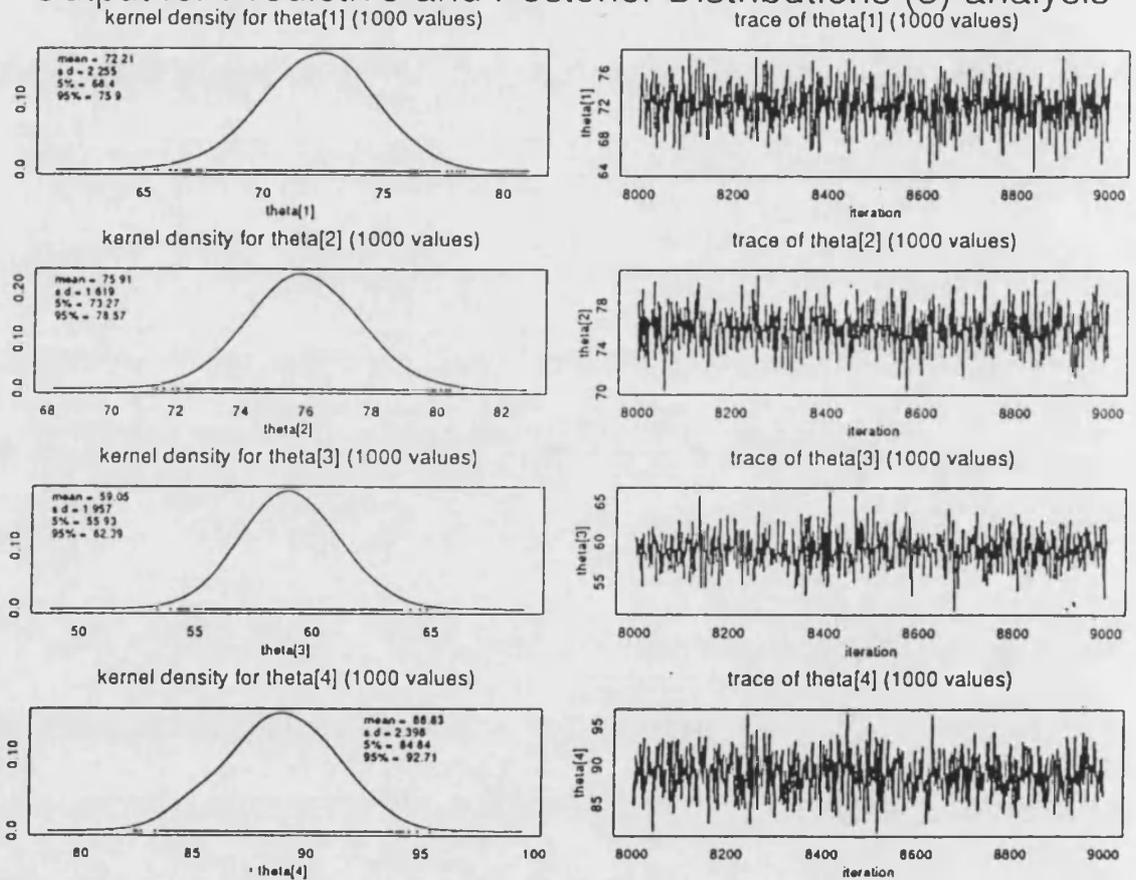


Figura 5.11: Distribuciones predictivas y finales dada la muestra y los tamaños de las rutas muestreadas (3)

## Output for Predictive and Posterior Distributions (4) analysis

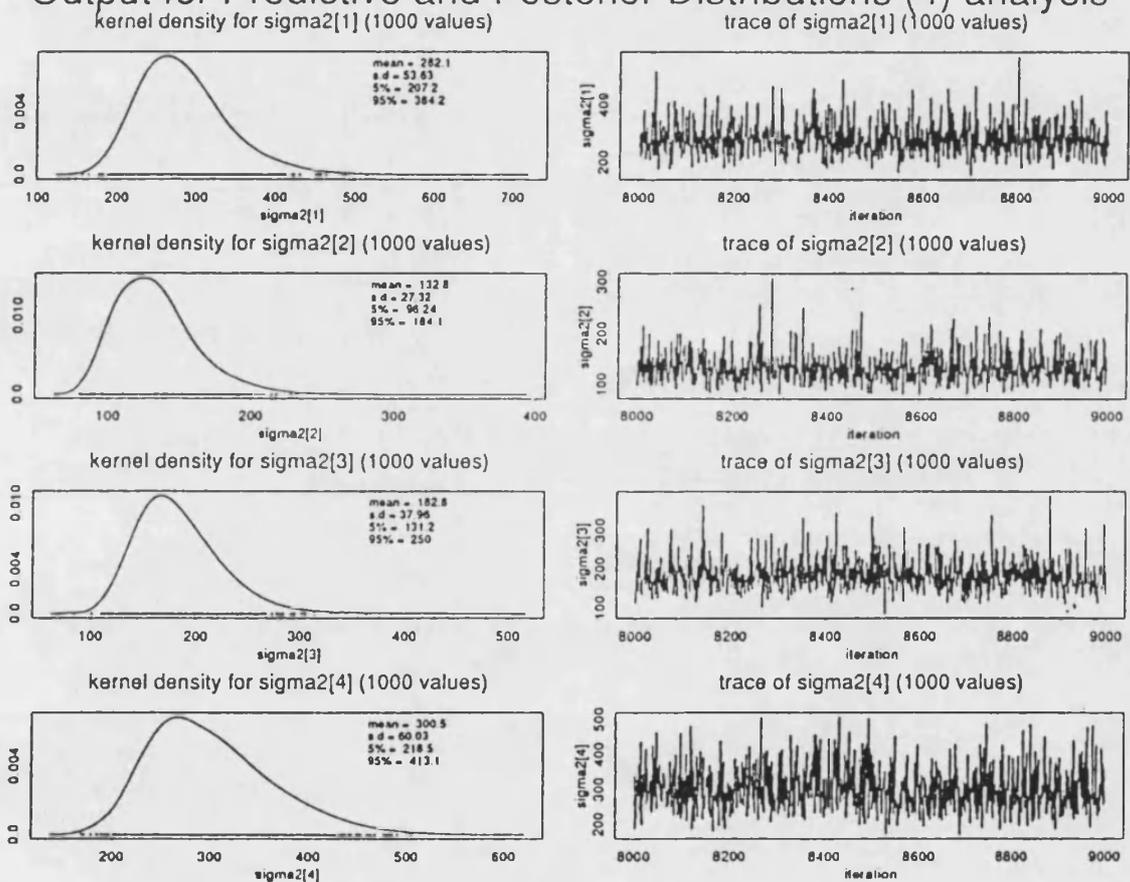


Figura 5.12: Distribuciones predictivas y finales dada la muestra y los tamaños de las rutas muestreadas (4)

## Output for Predictive and Posterior Distributions (5) analysis

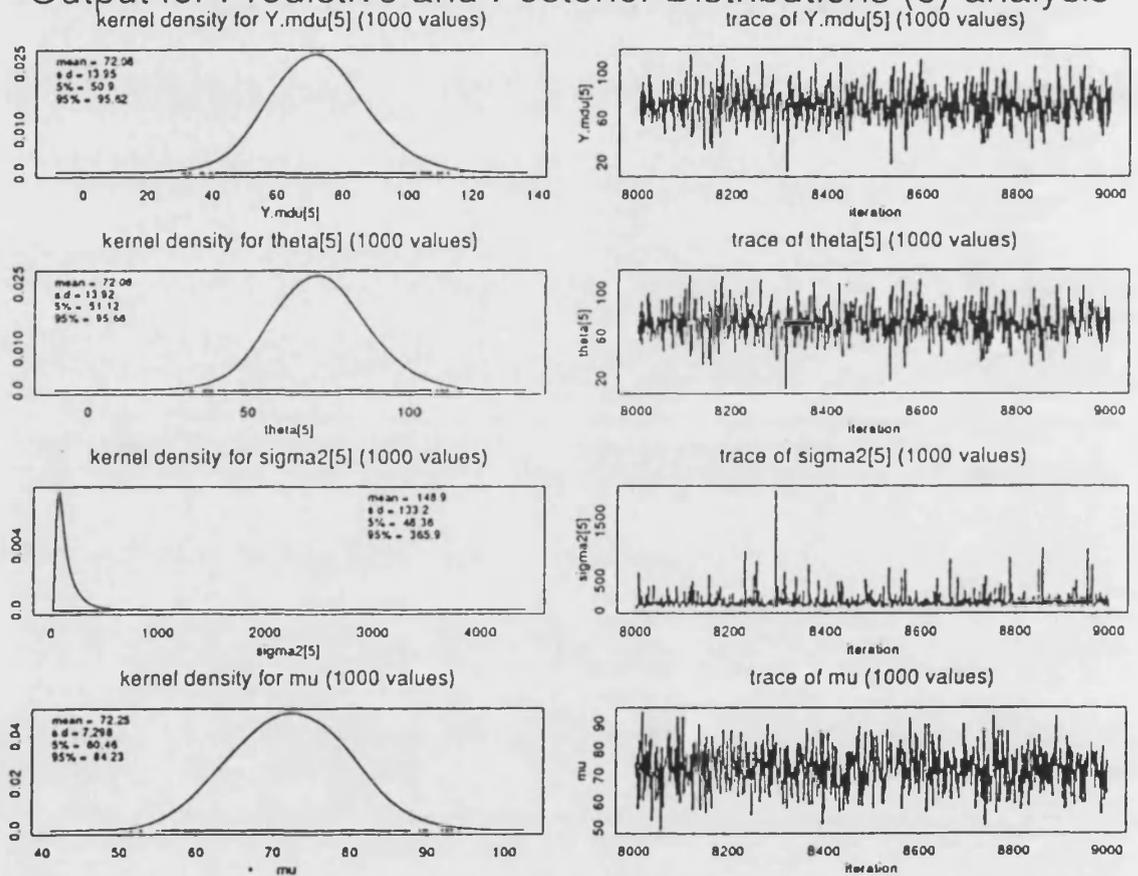


Figura 5.13: Distribuciones predictivas y finales dada la muestra y los tamaños de las rutas muestreadas (5)

las rutas muestreadas) que permite una inferencia mejor que la habitual que aplica procedimientos inferenciales basados en muestras aleatorias.

2.- Todos los resultados, modelos y método de muestreo se apoyan por simplicidad en la hipótesis de conocer los tamaños de las rutas muestreadas. Esta hipótesis, no siempre se satisface en la realidad y supone una restricción a la aplicación del modelo en la medida que no sea posible la obtención de unos buenos estimadores de los mismos. En futuros trabajos podría ensayarse la introducción de otros modelos más complejos, que permitieran un aprendizaje a priori de los tamaños de las rutas, que luego podría trasladarse a un modelo similar a los estudiados condicionado al tamaño de cada ruta. Estos modelos para conocer los tamaños de las rutas podrían apoyarse, por ejemplo, en la modelización de estos tamaños como los resultados de un proceso Poisson en el espacio asumiendo que el número de elementos de la población es constante por unidad de área.

# Apéndice A

## Notación

En esta tesis hemos aplicado una notación básica, que se ha completado al principio de cada capítulo atendiendo a las necesidades notacionales del problema tratado. Describiremos en este apéndice esta notación básica, atendiendo a una estructura monoetápica y bietápica de la población.

### Estructura monoetápica

La población se divide en elementos según la característica en estudio (que se asume en toda la tesis continua). Denotamos por:

- $N$ , número de elementos de la población.
- $Y_i(y_i)$ ,  $i = 1, 2, \dots, N$ , el valor de la característica en estudio para el elemento de la población  $i$ ,  $i = 1, 2, \dots, N$ . En mayúsculas indica que dicha característica es considerada como variable aleatoria, y en minúscula el valor tomado por la variable  $Y_i$  en el contexto de modelos de superpoblación, o simplemente el valor de la característica en estudio en los contextos de población fija.
- $\mathbf{y} = (y_1, y_2, \dots, y_N)^t$ , vector poblacional en población fija.
- $\mathbf{Y} = (Y_1, Y_2, \dots, Y_N)^t$ , vector poblacional (aleatorio) en el contexto de modelos de superpoblación.
- $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{ip})$ ,  $i = 1, 2, \dots, N$ , vector auxiliar de covariables asociado a  $Y_i$ ,  $i = 1, 2, \dots, N$ .
- $\mathbf{X} = (\mathbf{x}_1^t, \mathbf{x}_2^t, \dots, \mathbf{x}_N^t)^t$ , matriz de covariables.
- $\bar{Y} = \frac{\sum_{i=1}^N Y_i}{N}$ , la media poblacional.
- $n$ , número de elementos muestreados de la población.
- $s = \{i_1, i_2, \dots, i_n\}$ , índices de los elementos muestreados.
- $\mathbf{Y}_u = (Y_i : i \notin s)^t$ , vector poblacional no observado.

- $\mathbf{y}_s = (y_{i_1}, y_{i_2}, \dots, y_{i_n})^t$ , vector poblacional observado (conocido).
- $\mathbf{X}_u = (\mathbf{x}_i^t : i \notin s)^t$ , matriz de covariables correspondiente a los elementos no observados (a  $\mathbf{Y}_u$ ).
- $\mathbf{X}_s = (\mathbf{x}_{i_1}^t, \dots, \mathbf{x}_{i_n}^t)^t$ , matriz de covariables correspondiente a los elementos observados (a  $\mathbf{y}_s$ ).
- $\bar{Y}_u = \frac{\sum_{i \notin s} Y_i}{N-n}$ , la media poblacional no observada.
- $\bar{y}_s = \frac{\sum_{i \in s} y_i}{n}$ , la media muestral o media poblacional observada.
- $\hat{Y} (\hat{Y}_u)$ , estimadores de la media poblacional y la media poblacional no observada respectivamente (cuando aparecen con minúsculas tienen el mismo significado en población fija).

## Estructura bietápica

La población se divide en unidades que a su vez se dividen en elementos según la característica en estudio (que se asume en toda la tesis continua). Denotamos por:

- $K$ , número de unidades de la población.
- $N_i, i = 1, 2, \dots, K$ , número de elementos de la unidad  $i, i = 1, 2, \dots, K$ .
- $N$ , número de elementos de la población ( $N = \sum_{i=1}^K N_i$ ).
- $Y_{ij}(y_{ij}), j = 1, 2, \dots, N_i, i = 1, 2, \dots, K$ , el valor de la característica en estudio para el elemento de la población  $(i, j), j = 1, 2, \dots, N_i, i = 1, 2, \dots, K$ . En mayúsculas indica que dicha característica es considerada como variable aleatoria, y en minúscula el valor tomado por la variable  $Y_{ij}$  en el contexto de modelos de superpoblación, o simplemente el valor de la característica en estudio en los contextos de población fija.
- $\mathbf{y}_i = (y_{i1}, y_{i2}, \dots, y_{iN_i})^t, i = 1, 2, \dots, K$ , vector poblacional de la unidad  $i, i = 1, 2, \dots, K$  en población fija.
- $\mathbf{Y}_i = (Y_{i1}, Y_{i2}, \dots, Y_{iN_i})^t, i = 1, 2, \dots, K$ , vector poblacional (aleatorio) de la unidad  $i, i = 1, 2, \dots, K$  en el contexto de modelos de superpoblación.
- $\mathbf{y} = (\mathbf{y}_1^t, \mathbf{y}_2^t, \dots, \mathbf{y}_K^t)^t$ , vector poblacional en población fija.
- $\mathbf{Y} = (\mathbf{Y}_1^t, \mathbf{Y}_2^t, \dots, \mathbf{Y}_K^t)^t$ , vector poblacional (aleatorio) en el contexto de modelos de superpoblación.
- $\mathbf{x}_{ij} = (x_{ij1}, x_{ij2}, \dots, x_{ijp})^t, j = 1, 2, \dots, N_i, i = 1, 2, \dots, K$  vector auxiliar de covariables asociado a  $Y_{ij}, j = 1, 2, \dots, N_i, i = 1, 2, \dots, K$ .
- $\mathbf{X}_i = (\mathbf{x}_{i1}^t, \mathbf{x}_{i2}^t, \dots, \mathbf{x}_{iN_i}^t)^t, i = 1, 2, \dots, K$ , matriz de covariables asociada a  $\mathbf{Y}_i, i = 1, 2, \dots, K$ .
- $\mathbf{X} = (\mathbf{X}_1^t, \mathbf{X}_2^t, \dots, \mathbf{X}_K^t)^t$ , matriz de covariables.

- $\bar{Y}_i = \frac{\sum_{j=1}^{N_i} Y_{ij}}{N_i}$ ,  $i = 1, 2, \dots, K$ , la media (poblacional) de la unidad  $i$ ,  $i = 1, 2, \dots, K$ .
- $\bar{Y} = \frac{\sum_{i=1}^K N_i \bar{Y}_i}{N}$ , la media poblacional.
- $k$ , número de unidades muestreadas de la población.
- $\{i_1, \dots, i_k\}$ , índices de la población muestreados.
- $n_i$ ,  $i \in \{i_1, \dots, i_k\}$ , número de elementos muestreados de la unidad  $i$ ,  $i \in \{i_1, \dots, i_k\}$ .
- $s_i = \{j_1, \dots, j_{n_i}\}$ ,  $i \in \{i_1, \dots, i_k\}$ , índices de la unidad  $i$ ,  $i \in \{i_1, \dots, i_k\}$  muestreados.
- $n$ , número de elementos muestreados de la población ( $n = \sum_{i \in \{i_1, \dots, i_k\}} n_i$ ).
- $s = \{(i, j), j \in s_i, i \in \{i_1, \dots, i_k\}\}$ , índices de los elementos muestreados.
- $\mathbf{Y}_{u_i} = (Y_{ij} : j \notin s_i)^t$ ,  $i = 1, 2, \dots, K$ , vector poblacional no observado de la unidad  $i$ ,  $i = 1, 2, \dots, K$ .
- $\mathbf{Y}_u = (Y_{ij} : (i, j) \notin s)^t$ , vector poblacional no observado.
- $\mathbf{y}_{s_i} = (y_{ij} : j \in s_i)^t$ , vector poblacional observado de la unidad muestreada  $i$ ,  $i \in \{i_1, \dots, i_k\}$  (conocido).
- $\mathbf{y}_s = (\mathbf{y}_{s_i}^t : i \in \{i_1, \dots, i_k\})^t$ , vector poblacional observado (conocido).
- $\mathbf{X}_{u_i} = (\mathbf{x}_{ij}^t : j \notin s_i)^t$ ,  $i = 1, 2, \dots, K$ , matriz de covariables correspondiente a los elementos no observados de la unidad  $i$ ,  $i = 1, 2, \dots, K$  (a  $\mathbf{Y}_{u_i}$ ,  $i = 1, 2, \dots, K$ ).
- $\mathbf{X}_u = (\mathbf{x}_{ij}^t : (i, j) \notin s)^t$ , matriz de covariables correspondiente a los elementos no observados (a  $\mathbf{Y}_u$ ).
- $\mathbf{X}_{s_i} = (\mathbf{x}_{ij}^t : j \in s_i)^t$ , matriz de covariables correspondiente a los elementos observados de la unidad muestreada  $i$ ,  $i \in \{i_1, \dots, i_k\}$  (a  $\mathbf{y}_{s_i}$ ,  $i \in \{i_1, \dots, i_k\}$ ).
- $\mathbf{X}_s = (\mathbf{X}_{s_i}^t : i \in \{i_1, \dots, i_k\})^t$ , matriz de covariables correspondiente a los elementos observados (a  $\mathbf{y}_s$ ).
- $\bar{Y}_{u_i} = \frac{\sum_{j \notin s_i} Y_{ij}}{N_i - n_i}$ ,  $i = 1, 2, \dots, K$ , la media (poblacional) no observada de la unidad  $i$ ,  $i = 1, 2, \dots, K$ .
- $\bar{Y}_u = \frac{\sum_{i=1}^K (N_i - n_i) \bar{Y}_{u_i}}{N - n}$ , la media poblacional no observada.
- $\bar{y}_{s_i} = \frac{\sum_{j \in s_i} y_{ij}}{n_i}$ , la media muestral o media (poblacional) observada de la unidad  $i$ ,  $i \in \{i_1, \dots, i_k\}$ .
- $\bar{y}_s = \frac{\sum_{i \in \{i_1, \dots, i_k\}} n_i \bar{y}_{s_i}}{n}$ , la media muestral o media poblacional observada.
- $\hat{\bar{Y}}_i$  ( $\hat{\bar{Y}}_{u_i}$ ),  $i = 1, 2, \dots, K$ , estimadores de la media (poblacional) y la media (poblacional) no observada de la unidad  $i$ ,  $i = 1, 2, \dots, K$  respectivamente (cuando aparecen con minúsculas tienen el mismo significado en población fija).
- $\hat{\bar{Y}}$  ( $\hat{\bar{Y}}_u$ ), estimadores de la media poblacional y la media poblacional no observada respectivamente (cuando aparecen con minúsculas tienen el mismo significado en población fija).

## Apéndice B

# Lemas aplicados

En este apéndice presentaremos y demostraremos los resultados sobre matrices, funciones y distribuciones aplicados a lo largo de la tesis en la demostración de teoremas y corolarios. Como referencias auxiliares a las demostraciones presentadas señalemos, por ejemplo: Rao(73)[82][pág 1-78], Abramowitz y Stegun(72)[1], Lindley y Smith(72)[70] y DeGroot(70)[20].

### B.1 Resultados sobre matrices

**Lema B.1** *Consideremos: A una matriz  $p \times p$  y C una matriz  $n \times n$ , ambas no singulares y B y D dos matrices de dimensiones respectivas  $p \times n$  y  $n \times p$ , entonces se cumplen los siguientes dos resultados:*

$$(A + BCD)^{-1} = A^{-1} - A^{-1}B(C^{-1} + DA^{-1}B)^{-1}DA^{-1}, \quad (B.1)$$

$$|A + BD| = |A|^{-1}|I_n + CA^{-1}B|, \quad (B.2)$$

donde  $I_n$  es la matriz identidad de dimensión  $n$ .

Véase por ejemplo: Lindley y Smith(72)[70] (respecto a (B.1)) y Rao(73)[82].

### B.2 Resultados sobre funciones

**Lema B.2** *La función  $f(x) = x^{\frac{1}{2}} \exp\{-\beta x\} \frac{\lambda^x}{x!}$  definida en  $\mathfrak{R}^+$  (incluyendo el cero), con  $\beta$  y  $\lambda$  escalares positivos conocidos, es log-concava<sup>1</sup>.*

---

<sup>1</sup>De la log-concavidad en  $\mathfrak{R}^+$  (incluyendo el cero), se deriva la log-concavidad en los enteros positivos incluyendo el cero de la función de probabilidad  $g(x) \propto x^{\frac{1}{2}} \exp\{-\beta x\} \frac{\lambda^x}{x!}$  en el sentido considerado por Gilks(92)[44], por lo que, para simular una muestra de la distribución de  $\lambda$ , podemos hacer uso del método de aceptación-rechazo adaptado sin derivadas

**Demostración.** La función  $\log f(\cdot)$  es de clase  $C^2$  en  $\mathfrak{R}^+$ , por lo que, para comprobar el resultado basta con demostrar que  $\frac{d^2}{dx^2} \log f(x) < 0$  para  $\forall x \in \mathfrak{R}^+$ . Tenemos que:

- $\log f(x) = \frac{1}{2} \log x - \beta x + x \log \lambda - \log(\Gamma(x+1))$ ,
- $\frac{d}{dx} \log f(x) = \frac{1}{2x} - \beta + \log \lambda - \psi(x+1)$ , donde  $\psi(\cdot)$  es la función digamma,
- $\frac{d^2}{dx^2} \log f(x) = -\frac{1}{2x^2} - \psi'(x+1)$ , donde  $\psi'(\cdot)$  es la función trigamma,

luego, la función  $f(\cdot)$ , será log-concava si  $\psi'(x+1) > -\frac{1}{2x^2}$ . Esta desigualdad es fácil de comprobar ya que (ver 6.4.2, 6.4.6, 6.4.10 y 23.2.18 de Abramowitz y Stegun (72)[1]):

$$\begin{aligned} \psi'(x+1) &= \psi'(x) - x^{-2} = \sum_{t=0}^{\infty} (x+t)^{-2} - x^{-2} = \sum_{t=1}^{\infty} (x+t)^{-2}, \quad x \in (0, \infty) \text{ y} \\ \psi'(1) &= \zeta(2) = \sum_{t=1}^{\infty} x^{-2}, \quad x = 0, \text{ (con } \zeta(\cdot) \text{ la z-función de Riemann),} \end{aligned}$$

y por lo tanto,  $\psi'(x+1) > 0 > -\frac{1}{2x^2}$ ,  $\forall x \in \mathfrak{R}^+$  (incluyendo el cero).

C.S.Q.D.

### B.3 Resultados sobre distribuciones

**Lema B.3** *Dada una variable aleatoria  $X$  continua con rango en los reales positivos excluyendo el cero y función de densidad:*

$$p(x) \propto |\mathbf{I}_n x + \mathbf{A}|^{-\frac{1}{2}} x^{-(a_0+1)} \left( b_0 + \frac{b_1}{x} + \frac{1}{2} \mathbf{a}^t (\mathbf{I}_n x + \mathbf{A})^{-1} \mathbf{a} \right)^{-a_1 + \frac{n}{2}},$$

con  $\mathbf{A}$  una matriz simétrica definida positiva conocida de dimensión  $n \times n$ ,  $\mathbf{a}$  un  $n$ -vector asimismo conocido,  $\mathbf{I}_n$ , la matriz identidad de dimensión  $n$ ,  $n$  un entero conocido y  $a_0$ ,  $a_1$ ,  $b_0$  y  $b_1$  escalares positivos conocidos. Se tiene que, si  $n \geq 2$  y  $a_1 > a_0$ , entonces:

- $p(x)$  es propia,
- si  $a_0 > 1$ , existe  $E(X)$ ,
- si  $a_0 > 2$ , existe  $\text{Var}(X)$ ,
- $\forall a_0 > 0$  existen los siguientes momentos:  $E[(\mathbf{B} + \mathbf{C}x)^{-1}(\mathbf{B}\mathbf{b} + \mathbf{C}cx)]$ ,  $\text{Var}[(\mathbf{B} + \mathbf{C}x)^{-1}(\mathbf{B}\mathbf{b} + \mathbf{C}cx)]$  y también existen si  $a_0 > 1$ ,  $E\left\{ \left( b_2 + \frac{b_1}{x} + \frac{1}{2} \mathbf{a}^t (\mathbf{I}_n x + \mathbf{A})^{-1} \mathbf{a} \right) [m + \mathbf{b}^t (\mathbf{B} + \mathbf{C}x)^{-1} \mathbf{b}] x \right\}$ , donde  $\mathbf{B}$  y  $\mathbf{C}$  son matrices simétricas, definidas positivas conocidas de dimensión  $n \times n$ ,  $\mathbf{b}$  y  $\mathbf{c}$  son  $n$ -vectores conocidos,  $m$  es un escalar conocido y  $b_2$  un escalar positivo.

**Demostración.** Para ver que la distribución es propia, notemos que si  $a_1 > a_0$  y  $n \geq 2$ ,

$$\lim_{\lambda \rightarrow 0^+} |\mathbf{I}_n \mathbf{x} + \mathbf{A}|^{-\frac{1}{2}} x^{-(a_0+1)} \left( b_0 + \frac{b_1}{x} + \frac{1}{2} \mathbf{a}' (\mathbf{I}_n \mathbf{x} + \mathbf{A})^{-1} \mathbf{a} \right)^{-a_1 + \frac{n}{2}} = 0, \quad (\text{B.3})$$

$$\lim_{\lambda \rightarrow \infty} |\mathbf{I}_n \mathbf{x} + \mathbf{A}|^{-\frac{1}{2}} \left( b_0 + \frac{b_1}{x} + \frac{1}{2} \mathbf{a}' (\mathbf{I}_n \mathbf{x} + \mathbf{A})^{-1} \mathbf{a} \right)^{-a_1 + \frac{n}{2}} = 0, \quad (\text{B.4})$$

y que a partir de (B.4) se tiene que:

$$\exists H > 0 : \lambda > H \Rightarrow |\mathbf{I}_n \mathbf{x} + \mathbf{A}|^{-\frac{1}{2}} \left( b_0 + \frac{b_1}{x} + \frac{1}{2} \mathbf{a}' (\mathbf{I}_n \mathbf{x} + \mathbf{A})^{-1} \mathbf{a} \right)^{-a_1 + \frac{n}{2}} < 1. \quad (\text{B.5})$$

Si denotamos por  $p_{\bar{c}}(x)$  al núcleo de la función de densidad de  $X$ , se tiene que:

$$I = \int_0^{\infty} p_{\bar{c}}(x) dx = \int_0^H p_{\bar{c}}(x) dx + \int_H^{\infty} p_{\bar{c}}(x) dx,$$

donde  $I = (\text{constante de integración})^{-1}$ ,  $I_1 = \int_0^H p_{\bar{c}}(x) dx$  e  $I_2 = \int_H^{\infty} p_{\bar{c}}(x) dx$ . Las integrales  $I_1$  e  $I_2$  son finitas, veámoslo:

\*  $I_1 < \infty$ , porque la función  $P(x)$  definida en  $[0, H]$  por:

$$P(x) = \begin{cases} p_{\bar{c}}(x) & 0 < x \leq H \\ 0 & x = 0 \end{cases}$$

es continua en todo el dominio de su definición (ya que  $P(x)$  es continua en  $(0, H]$  por producto y composición de funciones elementales continuas y en 0 por definición y (B.3)).

\*  $I_2 < \infty$ , es propia ya que de (B.5) y multiplicando por  $x^{-(a_0+1)}$  se tiene que:

$$I_2 = \int_H^{\infty} p_{\bar{c}}(x) dx < \int_H^{\infty} x^{-(a_0+1)} dx,$$

donde  $\int_H^{\infty} x^{-(a_0+1)} dx < \infty$  por ser, por ejemplo,  $x^{-(a_0+1)}$  el núcleo de una distribución Pareto( $H, a_0$ ).

En conclusión, si  $a_1 > a_0$  y  $n \geq 2$ ,  $I < \infty$  y por otra parte, obviamente,  $p_{\bar{c}}(x) > 0$ , luego la función de densidad de  $X$  es una autentica función de densidad propia.

Los restantes resultados del lema se obtienen aplicando los argumentos anteriores, (bajo las hipótesis introducidas), sobre:  $\int_0^{\infty} x p_{\bar{c}}(x) dx$ ,  $\int_0^{\infty} x^2 p_{\bar{c}}(x) dx$ ,  $\int_0^{\infty} (\mathbf{B} + \mathbf{C}x)^{-1} (\mathbf{B}\mathbf{b} + \mathbf{C}cx) p_{\bar{c}}(x) dx$ ,  $\int_0^{\infty} (\mathbf{B} + \mathbf{C}x)^{-1} (\mathbf{B}\mathbf{b} + \mathbf{C}cx) (\mathbf{B}\mathbf{b} + \mathbf{C}cx)' (\mathbf{B} + \mathbf{C}x)^{-1} p_{\bar{c}}(x) dx$  y  $\int_0^{\infty} \left( b_2 + \frac{b_1}{x} + \frac{1}{2} \mathbf{a}' (\mathbf{I}_n \mathbf{x} + \mathbf{A})^{-1} \mathbf{a} \right) [m + \mathbf{b}' (\mathbf{B} + \mathbf{C}x)^{-1} \mathbf{b}] x dx$  respectivamente.

C.S.Q.D.

**Lema B.4** Dado un vector aleatorio  $X_1, X_2, \dots, X_K$  continuo con rango  $X_i > 0$ ,  $i = 1, 2, \dots, K$ , y función de densidad:

$$p(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_K) \propto \prod_{i=1}^K |\mathbf{I}_{n_i} \mathbf{x}_i + \mathbf{A}_i|^{-\frac{1}{2}} \times \prod_{i=1}^K x_i^{-(a_0+1)} (b_0 + b_1 \sum_{i=1}^K \frac{1}{x_i} + \frac{1}{2} \sum_{i=1}^K \mathbf{a}_i^t (\mathbf{I}_{n_i} \mathbf{x}_i + \mathbf{A}_i)^{-1} \mathbf{a}_i)^{-(a_1 + \frac{n}{2})},$$

con  $\mathbf{A}_i$ ,  $i = 1, 2, \dots, K$ , matrices simétricas definidas positivas conocidas de dimensiones respectivas  $n_i \times n_i$ ,  $\mathbf{a}_i$ ,  $n_i$ -vectores asimismo conocidos,  $i = 1, 2, \dots, K$ ,  $\mathbf{I}_{n_i}$ , matrices identidad de dimensiones  $n_i$ ,  $i = 1, 2, \dots, K$ ,  $n_i$ ,  $i = 1, 2, \dots, K$  enteros conocido,  $n = \sum_{i=1}^K n_i$  y  $a_0$ ,  $a_1$ ,  $b_0$  y  $b_1$  escalares positivos conocidos. Se tiene que, si  $n \geq 2K$  y  $a_1 > K a_0$ , entonces:

- $p(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_K)$  es propia,
- si  $a_0 > 1$ , existe  $E(X_1, X_2, \dots, X_K)$ ,
- si  $a_0 > 2$ , existe  $\text{Var}(X_1, X_2, \dots, X_K)$ ,
- $\forall a_0 > 0$  existen los siguientes momentos:  $E[(\mathbf{B}_i + \mathbf{C}_i \mathbf{x}_i)^{-1} (\mathbf{B}_i \mathbf{b}_i + \mathbf{C}_i \mathbf{c}_i \mathbf{x}_i)]$ ,  $\text{Var}[(\mathbf{B}_i + \mathbf{C}_i \mathbf{x}_i)^{-1} (\mathbf{B}_i \mathbf{b}_i + \mathbf{C}_i \mathbf{c}_i \mathbf{x}_i)]$ ,  $i = 1, 2, \dots, K$ , y si  $a_0 > 1$  también existen  $E[(\mathbf{B}_i + \mathbf{C}_i \mathbf{x}_i)^{-1} (\mathbf{B}_i \mathbf{b}_i + \mathbf{C}_i \mathbf{c}_i \mathbf{x}_i) (\mathbf{B}_j \mathbf{b}_j + \mathbf{C}_j \mathbf{c}_j \mathbf{x}_j)^t (\mathbf{B}_j + \mathbf{C}_j \mathbf{x}_j)^{-1}]$ ,  $i \neq j$ ,  $i, j = 1, 2, \dots, K$ , y  $E\{(b_0 + b_1 \sum_{i=1}^K \frac{1}{x_i} + \frac{1}{2} \sum_{i=1}^K \mathbf{a}_i^t (\mathbf{I}_{n_i} \mathbf{x}_i + \mathbf{A}_i)^{-1} \mathbf{a}_i) [m_i + \mathbf{b}_i^t (\mathbf{B}_i + \mathbf{C}_i \mathbf{x}_i)^{-1} \mathbf{b}_i] x_i\}$ ,  $i = 1, 2, \dots, K$ , donde  $\mathbf{B}_i$  y  $\mathbf{C}_i$ ,  $i = 1, 2, \dots, K$  son matrices simétricas, definidas positivas conocidas de dimensiones respectivas  $n_i \times n_i$ ,  $\mathbf{b}_i$  y  $\mathbf{c}_i$ ,  $i = 1, 2, \dots, K$ , son  $n_i$ -vectores conocidos y  $\{m_i\}_{i=1}^K$  son escalares conocidos.

**Demostración.** La demostración de este lema se apoya en el lema B.3 y en la siguiente acotación,

$$(b_0 + b_1 \sum_{i=1}^K \frac{1}{x_i} + \frac{1}{2} \sum_{i=1}^K \mathbf{a}_i^t (\mathbf{I}_{n_i} \mathbf{x}_i + \mathbf{A}_i)^{-1} \mathbf{a}_i)^{-a_1 + \frac{n}{2}} \leq K^{-(a_1 + \frac{n}{2})} \prod_{i=1}^K (\frac{b_0}{K} + b_1 \sum_{i=1}^K \frac{1}{x_i} + \frac{1}{2} \sum_{i=1}^K \mathbf{a}_i^t (\mathbf{I}_{n_i} \mathbf{x}_i + \mathbf{A}_i)^{-1} \mathbf{a}_i)^{-\frac{1}{K}(a_1 + \frac{n}{2})},$$

que se apoya en la desigualdad  $G \leq A$ , entre la media geométrica,  $G$ , y la media aritmética,  $A$ , de  $K$  números positivos. Notemos que la expresión anterior permite acotar la función de densidad del vector  $X_1, X_2, \dots, X_K$  por el producto de densidades del tipo descrito en el lema B.3 y por tanto las integrales necesarias para el análisis de este lema pueden ser acotadas por integrales que sabemos finitas, bajo las hipótesis introducidas en dicho lema.

C.S.Q.D.

**Lema B.5** Dada una variable aleatoria  $X$  continua con rango en  $\mathbb{R}$  y función de densidad:

$$p(x) \propto \prod_{i=1}^k \left[ 1 + \frac{1}{2b_i} (x - x_i)^2 \right]^{-a_i} \times \exp \left\{ -\frac{1}{2b_0} (x - x_0)^2 \right\},$$

con  $k$  un entero positivo conocido y  $b_i > 0$ ,  $i = 1, 2, \dots, k$  y  $x_i$ ,  $i = 0, 1, 2, \dots, k$  escalares conocidos. Se tiene que para  $b_0 > 0$  finito y para  $b_0^{-1} \rightarrow 0$  se cumple que:

- $p(x)$  es propia,
- si para algún  $i$ ,  $a_i > 1$  ó  $b_0 > 0$  finito, entonces existe  $E(X)$ ,
- si para algún  $i$ ,  $a_i > 2$  ó  $b_0 > 0$  finito, entonces existe  $Var(X)$ .

**Demostración.** La demostración de este resultado se apoya en las propiedades de la distribución normal y  $t$  de Student.

\* Para ver que la distribución es propia basta notar que la función de densidad antes descrita es proporcional al producto de funciones acotadas respecto a  $x$ , en la recta real (si  $b_0 > 0$ , producto de  $k$  núcleos  $t$  y de un núcleo normal y si  $b_0^{-1} \rightarrow 0$ , producto de  $k$  núcleos  $t$ ), luego acotada como función de  $x$  en la recta real y por lo tanto integrable en la recta real.

\* La existencia de primer momento se apoya en la integral  $\int_{-\infty}^{\infty} xp(x)dx$ , esta integral es propia si  $b_0 > 0$  ó si  $a_i > 2$ , para algún  $i \in \{1, 2, \dots, k\}$ .

Efectivamente:

$$\begin{aligned} \int_{-\infty}^{\infty} xp(x)dx &\leq C \int_{-\infty}^{\infty} x \exp \left\{ -\frac{1}{2b_0} (x - x_0)^2 \right\} dx \quad \text{si } b_0 > 0, \\ &\leq C \int_{-\infty}^{\infty} x \left[ 1 + \frac{1}{2b_i} (x - x_i)^2 \right]^{-a_i} dx \quad \text{si } b_0^{-1} \rightarrow 0, \end{aligned}$$

donde  $C$  es la constante de integración de la función de densidad  $p(x)$  (los restantes términos de la función de densidad se acotan por 1) e  $i$  es un índice cualquiera entre 1 y  $k$ . La integral de interés es convergente, en el caso  $b_0 > 0$  porque la normal tiene siempre primer momento, y en el caso  $b_0^{-1} \rightarrow 0$  si  $a_i > 1$  ya que entonces la distribución  $t$  tiene primer momento.

\* Análogamente al apartado anterior se comprueba que el segundo momento existe siempre si  $b_0 > 0$  y cuando  $a_i > 0$ , para algún  $i$  si  $b_0^{-1} \rightarrow 0$ , apoyándonos en el estudio de la convergencia de la integral  $\int_{-\infty}^{\infty} x^2 p(x)dx$  y la existencia del segundo momento de un normal y una  $t$  respectivamente.

C.S.Q.D.

**Lema B.6** Dado un vector aleatorio  $\mathbf{X} = (X_1, X_2, \dots, X_k)$  discreto con rango en los enteros positivos incluyendo el cero y función de probabilidad:

$$p(x_1, x_2, \dots, x_k) = \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{\Gamma(\sum_{i=1}^k x_i + \alpha)}{(\beta + k) \sum_{i=1}^k x_i + \alpha} \prod_{i=1}^k \frac{1}{x_i!},$$

con  $k$  un entero positivo conocido y  $\alpha > 0$  y  $\beta > 0$  escalares positivos conocidos, se tiene que existe la media y matriz de varianzas-covarianzas de  $\mathbf{X}$  y los siguientes valores esperados<sup>2</sup>:  $E \left[ \frac{1}{\sum_{i=1}^k X_i + x_0} \right]$ ,  $E \left[ \frac{\sum_{i=1}^k X_i}{\sum_{i=1}^k X_i + x_0} \right]$ ,  $E \left[ \frac{1}{(\sum_{i=1}^k X_i + x_0)^2} \right]$ ,  $E \left[ \frac{\sum_{i=1}^k X_i}{(\sum_{i=1}^k X_i + x_0)^2} \right]$ ,  $E \left[ \frac{\sum_{i=1}^k x_i}{\sum_{i=1}^k x_i + x_0} \right]$ ,  $E \left[ \frac{X_i}{\sum_{i=1}^k X_i + x_0} \right]$ ,  $i = 1, 2, \dots, k$ ,  $E \left[ \frac{X_i^2}{\sum_{i=1}^k X_i + x_0} \right]$ ,  $i = 1, 2, \dots, k$ , con  $x_0$  un entero positivo conocido.

**Demostración.** La demostración de este resultado se apoya en la convergencia de las siguientes dos series:

$$\sum_{x \geq 0} \frac{\Gamma(\alpha+x)}{\beta^{\alpha+x}} \frac{1}{x!}, \quad (\text{B.6})$$

$$\sum_{x \geq 0} \sum_{y \geq 0} \frac{\Gamma(\alpha+x+y)}{(\beta+2)^{\alpha+x+y}} \frac{1}{x!} \frac{1}{y!}. \quad (\text{B.7})$$

La serie (B.6) es convergente, para  $\beta > 1$ , por tratarse del núcleo de una distribución gamma Poisson de parámetros  $\alpha$ ,  $\beta - 1$  y 1, y la serie (B.7) es el núcleo de un caso particular de la distribución enunciada en este lema para  $k = 2$  (que es propia por el resultado (b) del lema B.12 tomando  $t = 1$ ). Efectivamente:

1.-  $E(\mathbf{X})$  existe dado que existe  $E(X_i)$ ,  $i = 1, 2, \dots, k$ . Veámoslo:

$$E(X_i) = \sum_{x_i \geq 0} x_i p(\mathbf{x}) = \frac{\beta^\alpha}{\Gamma(\alpha)} \prod_{j \neq i} \frac{1}{x_j!} \sum_{x \geq 0} \frac{\Gamma(x + \sum_{j \neq i} x_j + \alpha + 1)}{(\beta + k)^{x + \sum_{j \neq i} x_j + \alpha + 1}} \frac{1}{x!} < \infty$$

ya que la última serie es del tipo dado en (B.6).

2.-  $Var(\mathbf{X})$  existe dado que existen:  $E(X_i^2)$ ,  $i = 1, 2, \dots, k$  y  $E(X_i X_j)$ ,  $i \neq j$ ,  $i, j = 1, 2, \dots, k$ . Veámoslo:

<sup>2</sup> $E \left[ \frac{1}{\sum_{i=1}^k X_i + x_0} \right]$  y  $E \left[ \frac{\sum_{i=1}^k X_i}{\sum_{i=1}^k X_i + x_0} \right]$  pueden calcularse de acuerdo con el lema B.7 dado que  $\sum_{i=1}^k x_i \sim Gp(\alpha, \beta, k)$

2.1.- Por linealidad del operador esperanza se tiene que  $E(X_i^2) = E(X_i(X_i - 1)) + E(X_i)$ , de donde se sigue que  $E(X_i^2)$  existirá si existe  $E(X_i(X_i - 1))$ . Esta última esperanza existe ya que:

$$E(X_i(X_i-1)) = \sum_{x_i \geq 0} x_i(x_i-1)p(\mathbf{x}) = \frac{\beta^\alpha}{\Gamma(\alpha)} \prod_{j \neq i} \frac{1}{x_j!} \sum_{x \geq 0} \frac{\Gamma(x + \sum_{j \neq i} x_j + \alpha + 2)}{(\beta + k)^{x + \sum_{j \neq i} x_j + \alpha + 2}} \frac{1}{x!} < \infty$$

y la última serie es convergente por ser del tipo dado en (B.6).

2.2.- Por otra parte también existe  $E(X_i X_j)$ ,  $i \neq j$ ,  $i, j = 1, 2, \dots, k$ ,

$$\begin{aligned} E(X_i X_j) &= \sum_{x_i \geq 0} \sum_{x_j \geq 0} x_i x_j p(\mathbf{x}) = \\ &= \frac{\beta^\alpha}{\Gamma(\alpha)} \prod_{l \neq i, j} \frac{1}{x_l!} \sum_{x \geq 0} \sum_{y \geq 0} \frac{\Gamma(x+y + \sum_{l \neq i, j} x_l + \alpha + 2)}{(\beta + k)^{x+y + \sum_{l \neq i, j} x_l + \alpha + 2}} \frac{1}{x!} \frac{1}{y!} < \infty \end{aligned}$$

ya que la última serie es del tipo dado en (B.7) y por lo tanto convergente.

3.- Los demás valores esperados existen por tratarse de valores esperados de funciones acotadas por otras funciones cuyos valores esperados existen. Así:

- $E \left[ \frac{1}{\sum_{i=1}^k X_i + x_0} \right] < E \left[ \frac{1}{x_0} \right] = \frac{1}{x_0}$ , (dado que  $p(\mathbf{X})$  es propia).
- $E \left[ \frac{\sum_{i=1}^k X_i}{\sum_{i=1}^k X_i + x_0} \right] < E[1] = 1$ , (dado que  $p(\mathbf{X})$  es propia).
- $E \left[ \frac{1}{(\sum_{i=1}^k X_i + x_0)^2} \right] < E \left[ \frac{1}{x_0^2} \right] = \frac{1}{x_0^2}$ , (dado que  $p(\mathbf{X})$  es propia).
- $E \left[ \frac{\sum_{i=1}^k X_i}{(\sum_{i=1}^k X_i + x_0)^2} \right] < E \left[ \frac{1}{\sum_{i=1}^k X_i + x_0} \right] < \infty$ .
- $E \left[ \frac{\sum_{i=1}^k X_i}{\sum_{i=1}^k X_i + x_0} \right]^2 < E[1] = 1$ , (dado que  $p(\mathbf{X})$  es propia).
- $E \left[ \frac{X_i}{\sum_{i=1}^k X_i + x_0} \right] < E \left[ \frac{X_i}{x_0^2} \right] < \infty$ , (dado que existe  $E(\mathbf{X})$ ),  $i = 1, 2, \dots, k$ .
- $E \left[ \frac{X_i^2}{\sum_{i=1}^k X_i + x_0} \right] < E[1] = 1$ , (dado que  $p(\mathbf{X})$  es propia),  $i = 1, 2, \dots, k$ .

C.S.Q.D.

**Lema B.7** Sea  $X$  una variable aleatoria discreta gamma Poisson,  $X \sim Gp(\alpha, \beta, t)$  y  $x_0$  un entero positivo conocido, entonces se tiene que,

(a) Si  $\alpha > x_0$ ,

$$\begin{aligned} E\left[\frac{1}{X+x_0}\right] &= \sum_{i=0}^{x_0-1} \binom{x_0-1}{i} (-1)^i i! t^{-(1+i)} \beta^{1+i} \frac{\Gamma(\alpha-1-i)}{\Gamma(\alpha)}, \quad x_0 > 1, \\ &= \frac{1}{t} \frac{\Gamma(\alpha-1)}{\Gamma(\alpha)} \left[ \beta - \frac{\beta^\alpha}{(\beta+t)^{\alpha-1}} \right], \quad x_0 = 1. \end{aligned} \quad (\text{B.8})$$

(b) Si  $\alpha > x_0 + 2$ ,

$$E\left[\frac{X}{X+x_0}\right] = \sum_{i=0}^{x_0} \binom{x_0}{i} (-1)^i i! t^{-(2+i)} \beta^{2+i} \frac{\Gamma(\alpha-2-i)}{\Gamma(\alpha)}. \quad (\text{B.9})$$

**Demostración.** Para demostrar este resultado nos apoyaremos en que la distribución gamma-Poisson,  $X \sim Gp(\alpha, \beta, t)$  se puede obtener como resultado de calcular la marginal de una Poisson,  $X|\lambda \sim Po(t\lambda)$  y una gamma,  $\lambda \sim Ga(\alpha, \beta)$ , tal y como se obtiene en el resultado (a) del lema B.12, y en la definición de la siguiente función:

$$\tilde{G}_x(s) = \sum_{x \geq 1} s^{x-1} \exp\{-t\lambda\} \frac{(t\lambda)^x}{x!}, \quad s > 0.$$

La función  $\tilde{G}_x(s)$  está definida a partir de una serie convergente para  $s > 0$ , ya que aplicando las propiedades de la distribución Poisson:

$$\tilde{G}_x(s) = s^{-1} (\exp\{(s-1)t\lambda\} - \exp\{-t\lambda\}) < \infty$$

y por lo tanto (por tratarse de una serie de potencias) es infinitamente derivable para  $s > 0$  y en particular para  $s = 1$ , cumpliéndose que:

$$\tilde{G}_x^{(n)}(1) = \sum_{x \geq n+1} (x-1)(x-2)\cdots(x-n) \exp\{-t\lambda\} \frac{(t\lambda)^x}{x!}$$

con

$$\tilde{G}_x^{(n)}(1) = \frac{d^n}{ds^n} [s^{-1} (\exp\{(s-1)t\lambda\} - \exp\{-t\lambda\})]_{s=1} = \sum_{i=0}^n \binom{n}{i} (-1)^i i! (t\lambda)^{n-i}$$

A partir de estos resultados se tiene que:

1.-  $E \left[ \frac{1}{X+x_0} \right] = E^\lambda E^{X|\lambda} \left[ \frac{1}{X+x_0} \right]$ , con

$$E^{X|\lambda} \left[ \frac{1}{X+x_0} \right] = \begin{cases} \frac{1}{t\lambda}(1 - \exp\{-t\lambda\}) & \text{si } x_0 = 1 \\ (t\lambda)^{-x_0} \tilde{G}_x^{(x_0-1)}(1) & \text{si } x_0 > 1 \end{cases}$$

y tomando esperanza respecto a  $\lambda$ :

$$E \left[ \frac{1}{X+x_0} \right] = \begin{cases} t^{-1}[E(\lambda^{-1}) - E(\lambda^{-1}\exp\{-t\lambda\})] & \text{si } x_0 = 1 \\ \sum_{i=0}^{x_0-1} \binom{x_0-1}{i} (-1)^i i! t^{-(1+i)} E(\lambda^{-(1+i)}) & \text{si } x_0 > 1 \end{cases}$$

con

$$\begin{aligned} E(\lambda^{-(1+i)}) &= \frac{\beta^{1+i} \Gamma(\alpha-1-i)}{\Gamma(\alpha)} & \text{si } \alpha > i+1, i = 0, 1, \dots, x_0-1, \\ E(\lambda^{-1}\exp\{-t\lambda\}) &= \frac{\beta^\alpha}{(\beta+t)^{\alpha-1}} \frac{\Gamma(\alpha-1)}{\Gamma(\alpha)} & \text{si } \alpha > 1, \end{aligned}$$

de donde se obtiene el resultado (a).

2.-  $E \left[ \frac{X}{X+x_0} \right] = E^\lambda E^{X|\lambda} \left[ \frac{X}{X+x_0} \right]$ , con  $E^{X|\lambda} \left[ \frac{X}{X+x_0} \right] = (t\lambda)^{-1} E^{X|\lambda} \left[ \frac{1}{X+x_0+1} \right]$ , de donde se obtiene el resultado (b) aplicando los resultados obtenidos en el punto 1 de la presente demostración.

C.S.Q.D.

## B.4 Resultados sobre la obtención de distribuciones finales y marginales

**Lema B.8** *Supongamos que,*

$$\mathbf{Y}|\theta_1 \sim N_{p_0}(\mathbf{A}_1\theta_1, \mathbf{V}_1), \quad (\text{B.10})$$

$$\theta_1 \sim N_{p_1}(\mathbf{A}_2\theta_2, \mathbf{V}_2), \quad (\text{B.11})$$

entonces (a) la distribución marginal de  $\mathbf{Y}$  es

$$\mathbf{Y}|\theta_2 \sim N_{p_0}(\mathbf{A}_1\mathbf{A}_2\theta_2, \mathbf{V}_1 + \mathbf{A}_1\mathbf{V}_2\mathbf{A}_1^t), \quad (\text{B.12})$$

y (b) la distribución de  $\theta_1|y$  es  $N_{p_1}(\mathbf{D}d, \mathbf{D})$  con

$$\begin{aligned} \mathbf{D}^{-1} &= \mathbf{A}_1^t\mathbf{V}_1^{-1}\mathbf{A}_1 + \mathbf{V}_2^{-1}, \\ \mathbf{d} &= \mathbf{A}_1^t\mathbf{V}_1^{-1}\mathbf{y} + \mathbf{V}_2^{-1}\mathbf{A}_2\theta_2, \end{aligned} \quad (\text{B.13})$$

donde  $\theta_2$  es un  $p_2$ - vector de parámetros conocido,  $\mathbf{A}_1$  y  $\mathbf{A}_2$  son matrices de dimensiones  $p_0 \times p_1$  y  $p_1 \times p_2$  conocidas y  $\mathbf{V}_1$  y  $\mathbf{V}_2$  matrices de varianzas-covarianzas de dimensiones  $p_0 \times p_0$  y  $p_1 \times p_1$  definidas positivas conocidas.

**Demostración.**

Para probar (a) notemos que (B.10) puede escribirse dado  $\theta_1$  como  $\mathbf{Y} = \mathbf{A}_1\theta_1 + \mathbf{e}$  con  $\mathbf{e} \sim N_{p_0}(\mathbf{0}_{p_0}, \mathbf{V}_1)$  y (B.11) como  $\theta_1 = \mathbf{A}_2\theta_2 + \mathbf{v}$  con  $\mathbf{v} \sim N_{p_1}(\mathbf{0}_{p_1}, \mathbf{V}_2)$ , con  $\mathbf{e}$  y  $\mathbf{v}$  independientes. Juntado estas dos expresiones:

$$\mathbf{Y} = \mathbf{A}_1(\mathbf{A}_2\theta_2 + \mathbf{v}) + \mathbf{e},$$

esto es,  $\mathbf{Y}$  es combinación lineal de vectores independientes normales de donde se deduce, dado que toda combinación lineal de normales es normal, el resultado (B.12).

Para probar (b) aplicaremos el teorema de Bayes,

$$p(\theta_1|\mathbf{y}) \propto p(\mathbf{Y}|\theta_1)p(\theta_1),$$

el producto de la derecha es igual a  $e^{-\frac{1}{2}\mathbf{Q}}$  donde  $\mathbf{Q}$  viene dado por:

$$\begin{aligned} \mathbf{Q} &= (\mathbf{y} - \mathbf{A}_1\theta_1)' \mathbf{V}_1^{-1}(\mathbf{y} - \mathbf{A}_1\theta_1) + (\theta_1 - \mathbf{A}_2\theta_2)' \mathbf{V}_2^{-1}(\theta_1 - \mathbf{A}_2\theta_2) = \\ &= \theta_1' \mathbf{D}^{-1}\theta_1 - 2\mathbf{d}'\theta_1 + \text{constantes respecto a } \theta_1, \end{aligned}$$

de donde completando cuadrados respecto a  $\theta_1$  se tiene que

$$\mathbf{Q} = (\theta_1 - \mathbf{D}\mathbf{d})' \mathbf{D}^{-1}(\theta_1 - \mathbf{D}\mathbf{d}) + \text{constantes respecto a } \theta_1.$$

La parte de la derecha del teorema de Bayes es pues proporcional al núcleo de una normal con media  $\mathbf{D}\mathbf{d}$  y varianza  $\mathbf{D}$ , y por lo tanto  $\theta_1|\mathbf{y}$  se distribuye tal y como se indicaba en (b) con  $\mathbf{D}$  y  $\mathbf{d}$  los definidos en (B.13).

(Este resultado aparece demostrado en Lindley y Smith(72)[70].)

C.S.Q.D.

**Lema B.9** *Supongamos que,*

$$\mathbf{Y}|\sigma^2 \sim N_{p_0}(\theta_1, \sigma^2\mathbf{W}_1), \tag{B.14}$$

$$\sigma^2 \sim Ga^{-1}(a_0, b_0), \tag{B.15}$$

entonces (a) la distribución marginal de  $\mathbf{Y}$  es

$$\mathbf{Y} \sim t_{p_0}(\theta_1, b_0 a_0^{-1}\mathbf{W}_1, 2a_0), \tag{B.16}$$

y (b) la distribución de  $\sigma^2|y$  es  $Ga^{-1}(a_1, b_1)$  con

$$\begin{aligned} a_1 &= \frac{p_0}{2} + a_0, \\ b_1 &= \frac{1}{2}(\mathbf{y} - \boldsymbol{\theta}_1)^t \mathbf{W}_1^{-1}(\mathbf{y} - \boldsymbol{\theta}_1) + b_0, \end{aligned} \quad (\text{B.17})$$

donde  $\boldsymbol{\theta}_1$  es un  $p_0$ -vector de parámetros conocido,  $\mathbf{W}_1$  una matriz  $p_0 \times p_0$  definida positiva y  $a_0$  y  $b_0$  escalares positivos.

**Demostración.**

Para probar (a) aplicaremos que,

$$p(\mathbf{Y}) = \int_0^\infty p(\mathbf{Y}|\sigma^2)p(\sigma^2)d\sigma^2,$$

obteniendo que,

$$p(\mathbf{Y}) \propto \int_0^\infty [\sigma^2]^{-(a_1+1)} \exp\left\{-b_1 \left(\frac{1}{\sigma^2}\right)\right\} d\sigma^2,$$

notemos que la integral de la derecha corresponde como función de  $\sigma^2$  al núcleo de una  $Ga^{-1}(a_1, b_1)$ , por lo que aplicando sus propiedades tenemos que,

$$\begin{aligned} p(\mathbf{Y}) &\propto (b_1)^{-a_1} \propto \\ &\propto \left[1 + \frac{1}{2a_0} (\mathbf{Y} - \boldsymbol{\theta}_1)^t a_0 b_0^{-1} \mathbf{W}_1^{-1} (\mathbf{Y} - \boldsymbol{\theta}_1)\right]^{-\frac{1}{2}[p_0+2a_0]}, \end{aligned}$$

que es el núcleo de la  $t$  descrita en (B.16).

Para probar (b) aplicaremos el teorema de Bayes,

$$p(\sigma^2|y) \propto p(\mathbf{Y}|\sigma^2)p(\sigma^2),$$

obteniendo que,

$$p(\sigma^2|y) \propto [\sigma^2]^{-(a_1+1)} \exp\left\{-b_1 \left(\frac{1}{\sigma^2}\right)\right\},$$

que corresponde como función de  $\sigma^2$  al núcleo de una  $Ga^{-1}(a_1, b_1)$  con  $a_1$  y  $b_1$  los definidos en (B.17).

**C.S.Q.D.**

**Lema B.10** *Supongamos que,*

$$\mathbf{Y}|\theta_1, \sigma^2 \sim N_{p_0}(\mathbf{A}_1\theta_1, \sigma^2\mathbf{W}_1), \quad (\text{B.18})$$

$$p(\theta_1, \sigma^2) \propto \frac{1}{\sigma^2}, \quad (\text{B.19})$$

entonces (a) la distribución de  $\theta_1|\mathbf{y}, \sigma^2$  es  $N_{p_1}(\mathbf{D}\mathbf{d}, \sigma^2\mathbf{D})$  con

$$\begin{aligned} \mathbf{D}^{-1} &= \mathbf{A}_1^t \mathbf{W}_1^{-1} \mathbf{A}_1, \\ \mathbf{d} &= \mathbf{A}_1^t \mathbf{W}_1^{-1} \mathbf{y}, \end{aligned} \quad (\text{B.20})$$

y (b) la distribución de  $\sigma^2|\mathbf{Y} = \mathbf{y}$  es <sup>3</sup>,

$$\sigma^2|\mathbf{Y} = \mathbf{y} \sim Ga^{-1}\left(\frac{p_0 - p_1}{2}, \frac{1}{2}[\mathbf{y}^t(\mathbf{W}_1^{-1} - \mathbf{W}_1^{-1}\mathbf{A}_1(\mathbf{A}_1^t\mathbf{W}_1^{-1}\mathbf{A}_1)^{-1}\mathbf{A}_1^t\mathbf{W}_1^{-1})\mathbf{y}]\right), \quad (\text{B.21})$$

donde  $\theta_1$  es un  $p_1$ -vector de parámetros desconocido,  $\mathbf{A}_1$  una matriz de dimensiones  $p_0 \times p_1$  conocida,  $\mathbf{W}_1$  una matriz de varianzas-covarianzas de dimensión  $p_0 \times p_0$  definida positiva conocida y  $p_0 > p_1$ .

**Demostración.**

Para probar este resultado aplicaremos el teorema de Bayes,

$$p(\theta_1, \sigma^2|\mathbf{y}) \propto p(\mathbf{Y}|\theta_1, \sigma^2)p(\theta_1, \sigma^2),$$

el producto de la derecha es  $(\frac{1}{\sigma^2})^{\frac{p_0}{2}+1} e^{-\frac{1}{2\sigma^2}\mathbf{Q}}$  donde  $\mathbf{Q}$  viene dado por:

$$\begin{aligned} \mathbf{Q} &= (\mathbf{y} - \mathbf{A}_1\theta_1)^t \mathbf{W}_1^{-1} (\mathbf{y} - \mathbf{A}_1\theta_1) = \\ &= \theta_1^t \mathbf{D}^{-1} \theta_1 - 2\mathbf{d}^t \theta_1 + \mathbf{y}^t \mathbf{W}_1^{-1} \mathbf{y}, \end{aligned}$$

de donde completando cuadrados respecto a  $\theta_1$  se tiene que

---

<sup>3</sup>El segundo parámetro de la distribución es positivo. Aplicando el resultado (B.1) del lema B.1 sobre  $\mathbf{W}_1 + \mathbf{A}_1 \mathbf{W}_2 \mathbf{A}_1^t$  con  $\mathbf{W}_1$  y  $\mathbf{W}_2$  matrices definidas positivas de las dimensiones adecuadas y tomando límites para  $\mathbf{W}_2^{-1} \rightarrow 0$  se obtiene que

$$\lim_{\mathbf{W}_2^{-1} \rightarrow 0} \mathbf{y}^t (\mathbf{W}_1 + \mathbf{A}_1 \mathbf{W}_2 \mathbf{A}_1^t)^{-1} \mathbf{y} = \mathbf{y}^t (\mathbf{W}_1^{-1} - \mathbf{W}_1^{-1} \mathbf{A}_1 (\mathbf{A}_1^t \mathbf{W}_1^{-1} \mathbf{A}_1)^{-1} \mathbf{A}_1^t \mathbf{W}_1^{-1}) \mathbf{y},$$

luego como  $\mathbf{y}^t (\mathbf{W}_1 + \mathbf{A}_1 \mathbf{W}_2 \mathbf{A}_1^t)^{-1} \mathbf{y}$  es positivo, el segundo parámetro (segundo término del límite) es asimismo positivo.

$$\mathbf{Q} = (\boldsymbol{\theta}_1 - \mathbf{D}\mathbf{d})^t \mathbf{D}^{-1} (\boldsymbol{\theta}_1 - \mathbf{D}\mathbf{d}) - \mathbf{d}^t \mathbf{D}\mathbf{d} + \mathbf{y}^t \mathbf{W}_1^{-1} \mathbf{y},$$

y aplicando que,

$$p(\boldsymbol{\theta}_1, \sigma^2 | \mathbf{y}) = p(\boldsymbol{\theta}_1 | \sigma^2, \mathbf{y}) p(\sigma^2 | \mathbf{y}),$$

se obtienen la distribución de  $\boldsymbol{\theta}_1 | \sigma^2, \mathbf{y}$  definida en el apartado (a) con  $\mathbf{D}$  y  $\mathbf{d}$  los definidos en (B.20) y la de  $\sigma^2 | \mathbf{y}$  dada en (B.21).

C.S.Q.D.

**Lema B.11** Supongamos  $\mathbf{Y} = (\mathbf{Y}_1^t, \mathbf{Y}_2^t, \dots, \mathbf{Y}_k^t)^t$ , un vector aleatorio de dimensión  $p_0$ , con  $\mathbf{Y}_i$  de dimensión  $p_i$ ,  $i = 1, 2, \dots, k$ , que se distribuye según el siguiente modelo:

$$\mathbf{Y} | \sigma^2 \sim N_{p_0}(\boldsymbol{\theta}, \text{diag}_{i=1}^k \{\mathbf{W}_i \sigma_i^2\}) \quad (\text{B.22})$$

$$\sigma_i^2 \sim \text{Ga}^{-1}(a_0^i, b_0^i) \quad i = 1, 2, \dots, k, \text{ independientes}, \quad (\text{B.23})$$

entonces:

(a) los vectores  $\mathbf{Y}_i$ ,  $i = 1, 2, \dots, k$  son independientes y se distribuyen según una  $t$   $p_i$ -variante dada por:

$$\mathbf{Y}_i \sim t_{p_i}(\boldsymbol{\theta}_i, \frac{b_0^i}{a_0^i} \mathbf{W}_i, 2a_0^i), \quad i = 1, 2, \dots, k, \quad (\text{B.24})$$

y (b) las variables  $\sigma_i^2 | \mathbf{y}$ , se distribuyen  $\text{Ga}^{-1}(a_1^i, b_1^i)$  independientes para  $i = 1, 2, \dots, k$  con:

$$a_1^i = a_0^i + \frac{p_i}{2}, \quad i = 1, 2, \dots, k, \quad (\text{B.25})$$

$$b_1^i = b_0^i + (\mathbf{Y}_i - \boldsymbol{\theta}_i)^t \mathbf{W}_i^{-1} (\mathbf{Y}_i - \boldsymbol{\theta}_i), \quad i = 1, 2, \dots, k, .$$

donde  $\boldsymbol{\sigma}^2 = (\sigma_1^2, \sigma_2^2, \dots, \sigma_k^2)^t$ ,  $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^t, \boldsymbol{\theta}_2^t, \dots, \boldsymbol{\theta}_k^t)^t$  es un  $p_0$ -vector conocido ( $p_0 = \sum_{i=1}^k p_i$ ),  $\mathbf{W}_i$ ,  $i = 1, 2, \dots, k$  son matrices  $p_i \times p_i$  definidas positivas conocidas y  $a_0^i$ ,  $b_0^i$ ,  $i = 1, 2, \dots, k$  son escalares positivos conocidos.

**Demostración.** Para probar (a) aplicamos que,

$$\begin{aligned} p(\mathbf{y}) &= \int_0^\infty \int_0^\infty \dots \int_0^\infty \prod_{i=1}^k [p(\mathbf{y}_i | \sigma_i^2) p(\sigma_i^2)] d\sigma_1^2 d\sigma_2^2 \dots d\sigma_k^2 \propto \\ &\propto \int_0^\infty \int_0^\infty \dots \int_0^\infty \prod_{i=1}^k [(\sigma_i^2)^{-(a_1^i+1)} \exp\{-b_1^i (\sigma_i^2)^{-1}\}] d\sigma_1^2 d\sigma_2^2 \dots d\sigma_k^2, \end{aligned}$$

la integral de la derecha, como función de  $\sigma_i^2$  es el núcleo de una  $Ga^{-1}(a_1^i, b_1^i)$ , por lo que, aplicando sus propiedades para  $i = 1, 2, \dots, k$ , tenemos que:

$$p(\mathbf{y}) \propto \prod_{i=1}^k \left[ 1 + \frac{1}{2} (\mathbf{y}_i - \boldsymbol{\theta}_i)^t \mathbf{W}_i^{-1} (\mathbf{y}_i - \boldsymbol{\theta}_i) \right]^{-\frac{1}{2}[p_i + 2a_0^i]},$$

y por lo tanto el resultado dado en (B.24).

Para probar (b) aplicamos el teorema de Bayes,

$$\begin{aligned} p(\sigma^2 | \mathbf{y}) &\propto \prod_{i=1}^k [p(\mathbf{Y}_i | \sigma_i^2) p(\sigma_i^2)] \propto \\ &\propto \prod_{i=1}^k [(\sigma_i^2)^{-(a_1^i + 1)} \exp\{-b_1^i (\sigma_i^2)^{-1}\}], \end{aligned}$$

de donde se deduce el resultado dado en (b) con  $a_0^i$  y  $b_0^i$ ,  $i = 1, 2, \dots, k$ , definidos en (B.25).

C.S.Q.D.

**Lema B.12** Sea  $(X_1, X_2, \dots, X_k)^t$  un vector aleatorio discreto con rango en los enteros positivos incluyendo el cero y  $t > 0$ ,  $a_0 > 0$  y  $b_0 > 0$  escalares conocidos y supongamos que:

$$X_i | \lambda \sim Po(t\lambda), \quad i = 1, 2, \dots, k, \text{ independientes}, \quad (\text{B.26})$$

$$\lambda \sim Ga(a_0, b_0), \quad (\text{B.27})$$

entonces:

(a)  $X_i \sim Gp(a_0, b_0, t)$ ,  $i = 1, 2, \dots, k$ ,

(b)  $\mathbf{X} = (X_1, X_2, \dots, X_k)^t$  se distribuye con función de probabilidad,

$$p(x_1, x_2, \dots, x_k) = \frac{b_0^{a_0}}{\Gamma(a_0)} \frac{\Gamma(\sum_{i=1}^k x_i + a_0)}{(b_0 + kt)^{\sum_{i=1}^k x_i + a_0}} \prod_{i=1}^k \frac{t^{x_i}}{x_i!},$$

(c)  $\lambda | \{x_i\}_{i=1}^k \sim Ga(a_0 + \sum_{i=1}^k x_i, b_0 + kt)$ ,

y (d) si sustituimos (B.27) por una distribución mínimo informativa de la forma  $p(\lambda) \propto 1$ , entonces  $\lambda | \{x_i\}_{i=1}^k \sim Ga(\sum_{i=1}^k x_i + 1, kt)$ .

**Demostración.** Para probar (a) y (b) aplicamos respectivamente que,

$$\begin{aligned} p(x_i) &= \int_0^\infty p(x_i | \lambda) p(\lambda) d\lambda \propto \\ &\propto \frac{t^{x_i}}{x_i!} \int_0^\infty \exp\{-(t + b_0)\lambda\} \lambda^{a_0 + x_i - 1} d\lambda, \quad i = 1, 2, \dots, k, \end{aligned}$$

$$\begin{aligned} p(\mathbf{x}) &= \int_0^\infty \prod_{i=1}^k p(x_i | \lambda) p(\lambda) d\lambda = \\ &= \frac{b_0^{a_0}}{\Gamma(a_0)} \prod_{i=1}^k \frac{t^{x_i}}{x_i!} \int_0^\infty \exp\{-(kt + b_0)\lambda\} \lambda^{a_0 + \sum_{i=1}^k x_i - 1} d\lambda, \end{aligned}$$

y que las dos integrales de la derecha corresponden a los núcleos de dos gammas.

Y para probar (c) y (d), basta con aplicar el teorema de Bayes para las dos distribuciones para  $\lambda$  propuestas.

**C.S.Q.D.**

## Apéndice C

# Métodos computacionales aplicados

El estudio de modelos Bayesianos no conjugados hace necesario el empleo de métodos computacionales que permiten el estudio de la distribución final: el conocimiento de la constante de integración, obtención de sus primeros momentos (si los tuviese), su mediana y de algunos de sus cuantiles, aproximación gráfica de la función de densidad (o probabilidad) de esa distribución,... La literatura en este terreno es muy amplia y variada (como referencias generales podemos señalar, por ejemplo: Smith(91)[111] y O'Hagan(94)[78][capítulo 8]), recogiendo técnicas computacionales muy distintas: aproximaciones asintóticas a la distribución final y algoritmos EM (estimación-maximización), métodos de integración numérica (metodos de Laplace y Naylor-Smith), métodos de Monte Carlo y finalmente, métodos basados en cadenas de Markov (Markov Chain Monte Carlo (MCMC)), entre los que se encuentra el algoritmo Gibbs Sampling, estos últimos métodos y en particular el algoritmo Gibbs, han adquirido un gran auge en la actualidad por permitir estudiar distribuciones finales no analíticas de dimensiones grandes.

Este apéndice se centrará en la descripción de la técnica de muestreo por Gibbs y su implementación a partir del programa BUGS, con el objetivo de establecer los fundamentos de las técnicas y procedimientos de análisis aplicados en esta tesis para el estudio de poblaciones finitas mediante modelos no conjugados.

## C.1 Muestreo de la distribución por Gibbs Sampling

El método de muestreo por Gibbs se apoya en la teoría sobre las cadenas de Markov<sup>1</sup> y consiste en la obtención en el límite de una muestra de la distribución final del vector de interés a partir de un valor inicial del mismo y mediante un proceso iterativo que incluye la obtención de simulaciones de las distribuciones de cada variable del vector condicionadas a las restantes.

A continuación pasaremos a describir el algoritmo Gibbs, su convergencia, cómo inferir a partir del mismo resultados sobre la distribución final y algunos procedimientos de simulación univariante que facilitan su implementación en un ordenador.

### C.1.1 El algoritmo Gibbs Sampling

El algoritmo Gibbs Sampling que describiremos a continuación fue propuesto por Geman y Geman(84)[38].

Supongamos que deseamos obtener una muestra de la distribución  $f(\theta)$ ,  $\theta = (\theta_1, \theta_2, \dots, \theta_p)$ , y que conocemos las distribuciones marginales de cada una de las variables  $\theta_i$ ,  $i = 1, 2, \dots, p$ , condicionadas a las restantes, esto es las distribuciones  $f(\theta_i | \theta_1, \dots, \theta_{i-1}, \theta_{i+1}, \dots, \theta_p)$ ,  $i = 1, 2, \dots, p$ . Entonces si denotamos por  $\theta^0 = (\theta_1^0, \theta_2^0, \dots, \theta_p^0)$ , a un primer valor de  $\theta$  (punto de arranque del algoritmo Gibbs), generamos  $\theta^1$  a partir de  $\theta^0$ ,  $\theta^2$  a partir de  $\theta^1$ , y en general,  $\theta^{i+1}$  a partir de  $\theta^i$ , de modo que:

$$\begin{aligned} \theta_1^{i+1} & \text{ es una simulación de } \theta_1 \sim f(\theta_1 | \theta_2^i, \theta_3^i, \dots, \theta_p^i) \\ \theta_2^{i+1} & \text{ es una simulación de } \theta_2 \sim f(\theta_2 | \theta_1^{i+1}, \theta_3^i, \dots, \theta_p^i) \\ & \vdots \\ \theta_p^{i+1} & \text{ es una simulación de } \theta_p \sim f(\theta_p | \theta_1^{i+1}, \theta_2^{i+1}, \dots, \theta_{p-1}^{i+1}). \end{aligned} \quad (C.1)$$

Los vectores generados por el proceso anterior  $\{\theta^i\}_{i>0}$ , son realizaciones de una cadena de Markov homogénea (notemos que se cumple, en particular la propiedad de Markov y por tanto, la generación de  $\theta^{i+1}$  solo depende de  $\theta^i$  y no del conjunto  $\{\theta^j\}_{j=0}^{i-1}$ ), con función de transición de probabilidades dada por:

$$h(\theta^i, \theta^{i+1}) = f(\theta_1^{i+1} | \theta_2^i, \theta_3^i, \dots, \theta_p^i) \times f(\theta_2^{i+1} | \theta_1^{i+1}, \theta_3^i, \dots, \theta_p^i) \times \dots \times f(\theta_p^{i+1} | \theta_1^{i+1}, \theta_2^{i+1}, \dots, \theta_{p-1}^{i+1}) \quad (C.2)$$

<sup>1</sup>La teoría de los procedimientos para la integración por Monte Carlo desarrollada hasta el momento, es clásica. Ver O'Hagan(87)[77] para una discusión sobre este punto

### C.1.2 Convergencia del algoritmo Gibbs a la distribución final

Geman y Geman(84)[38] apoyándose en la teoría sobre cadenas de Markov demuestran que, bajo condiciones suaves, se tiene que:

$$(\theta_1^i, \theta_2^i, \dots, \theta_p^i) \xrightarrow{d} (\theta_1, \theta_2, \dots, \theta_p) \sim f(\theta_1, \theta_2, \dots, \theta_p), \quad (C.3)$$

$$\lim_{m \rightarrow \infty} \frac{1}{m} \sum_{i=1}^m T(\theta_1^i, \theta_2^i, \dots, \theta_p^i) \stackrel{\text{a.s.}}{=} E[T(\theta_1, \theta_2, \dots, \theta_p)], \quad (C.4)$$

siendo la velocidad de convergencia geométrica en  $i$  aplicando la norma supremo.

Con posterioridad, Tierney(91)[117] y Roberts y Smith(92)[87] establecen dos condiciones: la irreducibilidad y la aperioidicidad<sup>2</sup> de la cadena de Markov, para asegurar el cumplimiento de (C.3) y (C.4).

Respecto a la velocidad de convergencia y a la valoración de la misma respecto al número de simulaciones podemos señalar los resultados de Roberts y Polson(92)[86] (para  $p = 1$ ) y de Schervish y Carlin(90)[105].

Las condiciones de irreducibilidad y aperioidicidad (que permiten una interpretación fácil en los términos de la nota a pie de página) y la existencia de una distribución final  $f(\theta)$  propia (ver ejemplos de Casella y George(92)[15]), permiten establecer la convergencia efectiva del Gibbs Sampler y el problema se concreta entonces en establecer el número de iteraciones  $m$  del tipo (C.1), que son necesarias para poder considerar que  $\theta^m = (\theta_1^m, \theta_2^m, \dots, \theta_p^m)$  se distribuye según  $f(\theta)$ .

Los resultados sobre la velocidad de convergencia de Geman y Geman(84)[38] y sobre la valoración de la misma en los términos de los resultados de Roberts y Polson(92)[86] y Schervish y Carlin(90)[105], son de difícil aplicación en la práctica, por lo que se suele adoptar una postura al respecto más empírica. En este sentido, debemos destacar dos líneas de trabajo conducentes a establecer cuando podemos asumir la convergencia del muestreo por Gibbs:

1.- La primera línea de trabajo se basa en la realización de varias simulaciones por Gibbs Sampling en paralelo (parallel runs), esto es, seleccionar varios puntos iniciales de arranque y aplicar para cada uno de ellos el algoritmo de Gibbs. La convergencia del método Gibbs es, entonces, la valoración combinada de la convergencia de cada cadena y entre las cadenas. Esta valoración sobre la convergencia se basará en procedimientos gráficos (gráficos aproximativos de las posibles distribuciones finales marginales a partir de los últimos elementos simulados en cada cadena, gráficos de las trazas de las observaciones simuladas de cada variable, ...), o bien en procedimientos de diagnóstico más formales (ver Gelman y Rubin(92a,b)[36,37]).

<sup>2</sup>de una manera algo informal, una cadena es aperiódica cuando no existen valores iniciales distintos tales que el conjunto de valores que puedan ser generados a partir de ellos sean disjuntos y es irreducible cuando sea cual sea el valor inicial de partida es posible generar un valor para cualquier región con probabilidad positiva según la distribución  $f(\theta)$ . Roberts y Smith(92)[87] presentan definiciones formales y establecen condiciones suficientes para que una cadena sea aperiódica y/o irreducible

Entre los defensores de esta línea de trabajo señalemos: Gelfand y Smith(90)[35], Gelfand et al(90)[34], Casella y George(92)[15] y los antes citados Gelman y Rubin(92) [36,37].

2.- Otro grupo de autores, defienden la comodidad (y probablemente mayor ahorro) del empleo de un muestreo por Gibbs basado en un único punto de arranque del algoritmo, presentando procedimientos de diagnóstico de convergencia formales bajo esta circunstancias. Este es el caso de Geweke(92)[39], Raftery y Lewis(92)[81] y Roberts(92)[85].

Geweke(92)[39], se apoya para establecer su criterio de convergencia en (C.4) y en el siguiente resultado asintótico. Sean  $\bar{t}_m^A = \frac{1}{m_A} \sum_{i=1}^{m_A} T(\theta^i)$ ,  $\bar{t}_m^B = \frac{1}{m_B} \sum_{i=m-m_B+1}^m T(\theta^i)$ , los estimadores asintóticamente eficientes de  $E[T(\theta)]$ , basados en las primeras  $m_A$  y las últimas  $m_B$  observaciones de una simulación con  $m$  iteraciones del tipo (C.1), y sean  $\hat{S}_T^A(0)$  y  $\hat{S}_T^B(0)$ , estimadores consistentes<sup>3</sup> de las densidades espectrales de  $\{T(\theta^i)\}_{i=1}^{m_A}$  y  $\{T(\theta^i)\}_{i=m-m_B+1}^m$ ; entonces, para  $\frac{m_A}{m}$  y  $\frac{m_B}{m}$  constantes<sup>4</sup> y con  $m_A + m_B < m$ , se tiene que:

$$\mathbf{Z} = \frac{\bar{t}_m^A - \bar{t}_m^B}{[m_A^{-1}\hat{S}_T^A(0) + m_B^{-1}\hat{S}_T^B(0)]^{\frac{1}{2}}} \xrightarrow{d} \mathbf{Z}_c \sim N(0, \mathbf{I}_p). \quad (\text{C.5})$$

A partir del resultado anterior y apoyándose en el estadístico  $\mathbf{Z}$ , Geweke(92)[39] acepta la hipótesis de convergencia para un  $m$  dado cuando dicho estadístico toma valores propios de una  $N(0, \mathbf{I}_p)$  (esto es, tomando el intervalo de confianza al 95 % de una normal, cuando el valor de  $\mathbf{Z}$ , toma para cada componente valores inferiores en módulo a 1.96).

### C.1.3 Obtención de una muestra de la distribución final, estimación de sus primeros momentos y representación gráfica de la misma

Una vez asumida la estacionaridad de la cadena de Markov tras  $m$  iteraciones del tipo (C.1), podemos señalar tres aproximaciones para la obtención de una muestra de  $N$  observaciones de la distribución  $f(\theta)$ .

1.- El primer método consiste en la simulación hasta asumir convergencia (en  $m$ ), de  $N$  muestreos por Gibbs Sampling a partir de  $N$  puntos de arranque distintos (parallel runs); los  $\theta^m$  correspondientes a cada uno de los  $N$  muestreos por Gibbs constituyen obviamente una muestra independiente de la distribución  $f(\theta)$ .

2.- Supongamos que estamos trabajando con un único muestreo por Gibbs que se asume estabilizado en la iteración  $m$ , entonces una alternativa a (1) es tomar como

<sup>3</sup>Geweke(92)[39] propone calcular  $\hat{S}_T^m(0)$ , a partir de un periodograma de los  $\{T(\theta^i)\}_{i=1}^m$ , utilizando una ventana de Daniell para frecuencias  $(-\frac{2\pi}{3\sqrt{m}}, \frac{2\pi}{3\sqrt{m}})$ .

<sup>4</sup>Los valores propuestos por Geweke(92)[39] para  $m_A$  y  $m_B$  son:  $m_A = 0.1m$  y  $m_B = 0.5m$ .

muestra  $\{\theta^{m+it}\}_{i=0}^{N-1}$ , con  $t$  suficientemente grande para que las correlaciones entre las observaciones sea muy pequeña y pueda considerarse nula, con lo que la muestra anterior puede considerarse como una muestra independiente de la distribución  $f(\theta)$ .

Hills y Smith(92)[60] afirman que comunmente  $t < m$  y por lo tanto que este procedimiento es menos costoso que el anterior.

3.- La última alternativa consiste en emplear un único muestreo por Gibbs (que asumimos que se estabiliza en  $m$ ) y tomar como muestra  $\{\theta^{m+i}\}_{i=0}^{N-1}$ , aceptando la correlación existente entre las observaciones.

A partir de las muestras anteriores, es fácil obtener estimadores de la media, varianza (si existen), mediana o cualquier cuantil de la distribución  $f(\theta)$ , o de cualquiera de las marginales (ya que si  $\theta^0 = (\theta_1^0, \theta_2^0, \dots, \theta_p^0)$  es una simulación de  $f(\theta)$ , entonces  $\theta_i$  es una simulación de la distribución marginal  $f(\theta_i)$ ,  $i = 1, 2, \dots, p$ ). Dada  $T$  una función sobre  $\mathbb{R}^p$  y  $\{\theta_i\}_{i=1}^N$  una muestra de tamaño  $N$  (obtenida por cualquiera de las tres aproximaciones anteriores), el estimador insesgado habitual de  $E[T(\theta)]$ , viene dado por:

$$\bar{t}_N = \frac{1}{N} \sum_{i=1}^N T(\theta_i), \quad (\text{C.6})$$

cuya varianza viene dada por:

a.-) Si la muestra es obtenida por los procedimientos (1) ó (2):

$$\text{Var}(\bar{t}_N) = \frac{1}{N} \text{Var}[T(\theta)]. \quad (\text{C.7})$$

b.-) Si la muestra es obtenida por el procedimiento (3):

$$\begin{aligned} \text{Var}(\bar{t}_N) &= \frac{1}{N^2} \{N \text{Var}[T(\theta)] + 2 \sum_{i < j} \text{Cov}[T(\theta_i), T(\theta_j)]\} = \\ &= \frac{1}{N} \mathbf{S}_T(0) \end{aligned} \quad (\text{C.8})$$

donde  $\mathbf{S}_T(0)$  es la densidad espectral para frecuencia 0 (si ésta existe y es continua en 0).

Debemos notar que una muestra de tamaño  $Nt$  obtenida por el procedimiento (3) (aunque las observaciones estén correladas), es al menos tan eficiente como una de tamaño  $N$  obtenida por los procedimientos (1) o (2). El razonamiento para establecer la observación anterior es muy simple: si denotamos por  $\{\theta_i\}_{i=1}^{Nt}$ , a una muestra del tamaño  $Nt$  obtenida por el procedimiento (3) entonces, las muestras

$M_i = \{\theta_{i+(j-1)t}\}_{j=1}^N$ ,  $i = 1, 2, \dots, t$ , son del tipo del procedimiento (2) (o del (1)), cuyas observaciones son independientes entre sí, y

$$\begin{aligned} \text{Var}(\bar{t}_N) &= \frac{1}{(Nt)^2} \text{Var}\left[\sum_{i=1}^t \sum_{\theta_i \in M_i} T(\theta_i)\right] = \\ &= \frac{1}{(Nt)^2} \sum_{i=1}^t \sum_{j=1}^t \text{Cov}\left[\sum_{\theta_i \in M_i} T(\theta_i), \sum_{\theta_j \in M_j} T(\theta_j)\right] \leq \\ &\leq \frac{1}{N} \text{Var}[T(\theta)]. \end{aligned} \quad (\text{C.9})$$

Respecto al estudio de las distribuciones finales marginales  $f(\theta_j)$ ,  $j = 1, 2, \dots, p$ , considerada una muestra  $\{\theta_i\}_{i=1}^N$  de la distribución final, la literatura ofrece dos posibilidades: la representación gráfica de la misma a partir de la muestra  $\{\theta_{ij}\}_{i=1}^N$  (empleando las aproximaciones de kernel a la función de densidad) o su estimación mediante:

$$\widehat{f(\theta_j)} = \frac{1}{N} \sum_{i=1}^N f(\theta_j | \theta_{i1}, \dots, \theta_{ij-1}, \theta_{ij+1}, \dots, \theta_{ip}). \quad (\text{C.10})$$

(Notemos que  $f(\theta_j) = E[f(\theta_j | \theta_1, \dots, \theta_{j-1}, \theta_{j+1}, \dots, \theta_p)]$ .)

### C.1.4 Mecanismos para la simulación de variables aleatorias

El algoritmo Gibbs Sampling requiere en cada iteración  $i+1$  de la simulación de  $p$  variables aleatorias con distribuciones dadas por  $f(\theta_1 | \theta_2^i, \theta_3^i, \dots, \theta_p^i)$ ,  $f(\theta_2 | \theta_1^{i+1}, \theta_3^i, \dots, \theta_p^i)$ ,  $\dots$ ,  $f(\theta_p | \theta_1^{i+1}, \theta_2^{i+1}, \dots, \theta_{p-1}^{i+1})$ , por lo que para su implementación será necesario poseer algoritmos que posibiliten la generación de esas variables aleatorias.

Para la generación de las variables aleatorias anteriores podemos destacar cuatro grupos de métodos:

**1.-Densidades o funciones de probabilidad estándares.** Si las densidades o funciones de probabilidad son estándares, para simularlas poseemos una gama de algoritmos extensamente estudiados en Devroye(86)[25] y Ripley(87)[84].

**2.-Métodos de Aceptación-Rechazo (Rejection Methods).** Supongamos que queremos generar una distribución sobre  $\theta$ , con función de densidad (o probabilidad) no estándar  $g(\theta) = C_g g^*(\theta)$ , con  $C_g$  la constante de normalización y  $s(\theta) = C_s s^*(\theta)$ , con  $C_s$  la constante de normalización, una densidad (o probabilidad) correspondiente a una distribución que sabemos simular (por ejemplo, la densidad de una distribución estándar), y supongamos que el cociente  $\frac{g^*(\theta)}{s^*(\theta)}$  está acotado y podemos determinar para ellos una cota superior  $A$ , entonces el algoritmo para la simulación de  $g(\theta)$  es el siguiente: simular  $\theta \sim s(\theta)$  y  $U \sim Un(0, 1)$  hasta que se cumpla:

$$AU \leq \frac{g^*(\theta)}{s^*(\theta)}, \quad (\text{C.11})$$

entonces aceptar  $\theta$  como una simulación de la distribución sobre  $\theta$ . La eficiencia del algoritmo de aceptación-rechazo anterior, depende de la cota  $A$  y de la elección de la densidad (o función de probabilidad)  $s(\theta)$ .

Gilks y Wild(92)[47], Gilks(92)[44] y Gilks, Best y Tan(92)[45], proponen tres variantes distintas del algoritmo anterior, que mejoran la eficiencia del proceso de simulación con la construcción de una función de densidad o probabilidad  $s(\theta)$ , fácil de simular y un test rápido de rechazo que evita el cálculo del cociente dado en (C.11). Las variantes de Gilks y Wild(92)[47] y Gilks(92)[44] se denominan respectivamente método de aceptación-rechazo adaptado (adaptive rejection method) y método de aceptación-rechazo adaptado sin derivadas (derivative free adaptive rejection) y se aplican sobre función de densidad (o probabilidad) log-concavas, requiriendo el primero la evaluación de derivadas de la log-densidad (log-probabilidad). La variante de Gilks, Best y Tan(92)[45] no requiere, en cambio, la hipótesis de log-concavidad.

**3.- Métodos basados en cocientes de uniformes (ratio-of-uniform sampling).** Sea  $g(\theta) = C_g g^*(\theta)$ , con  $C_g$  la constante de normalización, la función de densidad (o probabilidad) de la distribución que queremos simular y  $A_g = \{(u, v) : 0 < u < \sqrt{g^*(\frac{u}{v})}\}$ , entonces el algoritmo de cociente de uniformes consiste en simular  $(U, V)$  uniformes sobre  $A_g$  y tomar  $\theta = \frac{U}{V}$  como una simulación de la distribución sobre  $\theta$ .

**4.- Método de la transformada inversa (inversion method).** Sea  $G(\theta)$ , la función de distribución que queremos simular y  $G^{-1}(u) = \inf\{\theta : \theta \in A_u\}$  con  $A_u = \{\theta : F(\theta) \geq u\}$ , entonces el método de la transformada inversa consiste en simular  $U \sim Un(0, 1)$  y tomar  $\theta = G^{-1}(U)$  como una simulación de la distribución  $G(\theta)$ .

## C.2 El programa BUGS (Bayesian Inference using Gibbs Sampling)

El programa BUGS creado por Gilks, Spiegelhalter y Thomas, consiste en una implementación del algoritmo de Gibbs Sampling abierta, que permite estudiar mediante esta técnica una amplia gama de modelos (siempre que estos cumplan unas determinadas hipótesis). Este programa trabaja a partir de la definición del modelo (especificado en términos de BUGS), generando las distribuciones condicionales y eligiendo un mecanismo de simulación para las mismas.

El programa BUGS cuyas principales características describiremos a continuación, ha reducido considerablemente el esfuerzo computacional para el análisis de los modelos no conjugados presentados y analizados en esta tesis. Como referencias sobre el BUGS señalemos los artículos de Thomas, Spiegelhalter y Gilks(92)[115] y Gilks, Thomas y Spiegelhalter(94)[46] y los dos manuales del BUGS[13,14].

### C.2.1 Especificación del modelo

El programa BUGS en su versión 0.30 es aplicable en modelos que cumplan las siguientes condiciones:

- el modelo ha de ser DAG (Directed Acyclic Graphs).
- las distribuciones a especificar en la descripción del modelo y las condicionales necesarias para aplicar el algoritmo Gibbs han de ser univariantes ya que el BUGS sólo genera distribuciones univariantes.
- solo admite como máximo tres bucles de subíndices para describir las variables del modelo ( $Y_{ijk}$ , como mucho).
- las distribuciones condicionales a especificar en la descripción del modelo han de ser estándar y las distribuciones condicionales para aplicar el Gibbs han de ser estándar o bien log-concavas.
- si se define alguna variable discreta, el número de posibles estados es como máximo 101.

La condición DAG es de gran importancia para la implementación correcta de nuestro modelo, ya que el programa BUGS se basa en esta propiedad para la construcción de las condicionales necesarias para el muestreo Gibbs. Los modelos DAG son modelos jerárquicos en los que la dependencia probabilística entre las variables es directa de modo que la distribución conjunta de todo el modelo es producto de las distribuciones conjuntas de cada una de sus jerarquías (hipótesis de cadena de Markov directa (Lauritzen et al(90)[68] ) ).

Estos modelos, admiten una representación gráfica con las siguientes características:

- Se distinguen tres tipos de nodos: 1) constantes o datos, que se representan con cuadrados, 2) nodos estocásticos, son variables del problema que tienen asociada una distribución y se representan con círculos, y 3) nodos deterministas, que son funciones de otros nodos y se representan con triángulos.
- Las relaciones entre los nodos (ejes del gráfico), se representan con flechas de modo que el nodo del extremo “depende del” nodo del origen.

Con la particularidad (Whitaker(90)[118]), descrita en el nombre, de que todos los ejes son directos y no es posible siguiendo la dirección de las flechas retornar a un nodo una vez éste ha sido abandonado.

La representación anterior del modelo ayuda a trasladar el modelo al lenguaje BUGS. La descripción en BUGS del modelo (para ver más detalles consultar el manual), consta de dos partes:

1. Una parte declarativa en la que se introduce: el nombre del modelo, los nombres de las constantes, los nombres de las variables, y se indica el nombre del archivo en el que están guardados los datos (variables auxiliares o variables muestreadas) y (opcionalmente) el punto de arranque de la simulación.

2. La descripción en sí del modelo, esta descripción queda encerrada entre llaves  $\{\dots\}$ . El signo “ $\sim$ ” se aplica para representar relaciones con extremo en un nodo estocástico y significa “distribuido como”, y el signo “ $\leftarrow$ ” se aplica para representar relaciones deterministas y significa “igual a”. Para representar bucles (por ejemplo, un conjunto de variables con la misma distribución), se emplea el signo  $for(i \text{ in } \square : \square)\{\dots\}$ .

## C.2.2 Descripción del funcionamiento

Una vez arrancado el programa e introducido el modelo, el BUGS: examina el modelo introducido, carga los datos y valores iniciales (si se introducen) del modelo, genera internamente un gráfico del modelo, genera las distribuciones condicionales para aplicar el Gibbs Sampling, genera el punto de arranque del mismo y elige los algoritmos para simular las condicionales anteriores.

Las distribuciones condicionales para aplicar el Gibbs son calculadas por el programa atendiendo a que el modelo propuesto es DAG (y por tanto se cumple la hipótesis de cadena de Markov directa), aplicando que:

$$f(\theta) = \prod_{i=1}^p f(\theta_i | \text{padres}[\theta_i]), \quad (\text{C.12})$$

y que entonces:

$$\begin{aligned} f(\theta_i | \theta_1, \dots, \theta_{i-1}, \theta_{i+1}, \dots, \theta_p) &\propto f(\theta_1, \theta_2, \dots, \theta_p) & (\text{C.13}) \\ &\propto \text{términos de } f(\theta) \text{ que contienen } \theta_i = \\ &= f(\theta_i | \text{padres}[\theta_i]) \prod_{\theta_j \in \text{padres}[\theta_i]} f(\theta_j | \text{padres}[\theta_j]), \end{aligned}$$

donde por  $\text{padres}[\theta_i]$  nos referimos a los nodos de los que depende a un primer nivel  $\theta_i$ ,  $i = 1, 2, \dots, p$ .

Una vez obtenidas por el programa las distribuciones condicionales, aplicando (C.13), el BUGS selecciona el algoritmo para su simulación empleando si la distribución es estándar (bernoulli, beta, binomial, Dirichlet, exponencial, gamma, log-normal, multinomial, normal, Pareto, Poisson, uniforme o Weibull), un algoritmo de simulación de la misma y en otro caso el algoritmo de aceptación-rechazo adaptado sin derivadas de Gilks(92)[44]. En este punto adquiere su importancia la hipótesis de condicionales estándar o log-concavas puesto que si esta hipótesis no se da el programa BUGS se parará dando error (lo más probable) o puede que continúe pero proporcionando unos resultados erróneos.

Respecto al punto de arranque para la simulación por Gibbs, el programa BUGS trabaja con el punto suministrado en una base de datos (cuyo nombre queda indicado

en la descripción del modelo) y en caso de ausencia lo genera mediante una primera iteración “falsa” del muestreo por Gibbs seguida por un chequeo del mismo a través de la función de verosimilitud, que en el caso de resultar negativo detiene el programa causando error.

### C.2.3 Diagnóstico de la convergencia y obtención de resultados

Realizados los pasos anteriores, el programa BUGS está en condiciones de realizar el muestreo por Gibbs Sampling. Para estudiar la convergencia del mismo, el BUGS proporciona archivos con todas las simulaciones y mediante el comando `diag` una aproximación al estadístico de Geweke(92)[39] descrito en (C.5). El estadístico suministrado  $\tilde{Z}$ , para cada variable se obtiene de la siguiente manera, para cada serie de  $m$  simulaciones de cada variable se consideran los grupos  $A$  y  $B$  formados por las  $\frac{m}{4}$  primeras simulaciones y por las  $\frac{m}{2}$  últimas respectivamente, que se dividen en 25 subgrupos cada uno para los cuales se obtienen las medias de cada subgrupo, esto es,  $\{m_i^A\}_{i=1}^{25}$  y  $\{m_i^B\}_{i=1}^{25}$ , respectivamente y se calcula  $\tilde{Z}$ , haciendo:

$$\tilde{Z} = \frac{\bar{m}_A - \bar{m}_B}{\sqrt{\frac{s_A^2 + s_B^2}{25}}}, \quad (C.14)$$

donde  $\bar{m}_A$ ,  $\bar{m}_B$ ,  $s_A^2$  y  $s_B^2$ , son las medias y varianzas muestrales de los  $\{m_i^A\}_{i=1}^{25}$  y  $\{m_i^B\}_{i=1}^{25}$ .

Una vez asumida la estacionariedad de la cadena de Markov, el programa BUGS facilita la obtención de los estadísticos más comunes para las distribuciones finales marginales mediante el comando `stats`. Este comando obtiene los estimadores habituales de la media, desviación típica, mediana y cuantiles 2.5 % y 97.5 % de la distribución marginal final de cada variable a partir de una muestra simulada de la misma. Esta muestra construida una vez obtenida la estabilidad de la cadena puede responder a las aproximaciones (2) o (3) a las que nos referíamos en la subsección C.1.3. En el segundo caso, para valorar la correlación entre las observaciones y por tanto precisión de nuestras estimaciones (ver (C.8)) nos podemos apoyar como método aproximado para estimar la densidad espectral en cero en el valor de  $s_B^2$  proporcionado como dato auxiliar por el comando `diag`, comentado con anterioridad.

Para completar el análisis a partir de la muestra de las distribuciones marginales finales, el BUGS suministra un programa de Splus, que denomina `drawdat`, que obtiene la aproximación de kernel (más o menos suave según empleemos el intervalo recomendado por Silverman(86)[108][ecuaciones 3.28 y 3.30] o por Hardle(91)[54][ecuación 4.1.3] de la densidad (o de la función de probabilidad) de las distribuciones marginales finales y una representación de la traza de la muestra, que permite una valoración gráfica de la posible estabilidad del muestreo Gibbs.

Para terminar la sección, señalemos que aunque el planteamiento sobre la convergencia, la obtención de la muestra y estimaciones desarrollado en esta subsección responde a las aproximaciones basadas en una única cadena de Markov (one run), el programa

BUGS permite realizar un estudio basado en la aplicación del muestreo Gibbs para un grupo de puntos de arranque (paralell runs), siguiendo la filosofía de Gelman y Rubin(92)[36], sin más que, haciendo funcionar el programa BUGS sucesivamente con cada uno de los puntos de arranque propuestos.

### C.3 Aplicación del programa BUGS al análisis de los modelos Bayesianos no conjugados sobre poblaciones finitas estudiados en esta tesis

Dedicaremos esta sección a comentar brevemente la metodología que se ha empleado para analizar los modelos no conjugados planteados en esta tesis (en los capítulos 2, 3 y 5). Podemos distinguir en este análisis tres fases:

1.- *Estudio de los modelos.* En esta primera fase se estudian los modelos presentados: se garantiza que las distribuciones predictivas y posteriores de interés son propias y se establecen las condiciones suficientes para la existencia de los primeros dos momentos (media y varianza).

Este análisis garantiza la existencia de una distribución final propia a la que el algoritmo Gibbs Sampling pueda converger. Esta propiedad unida a que estos modelos (apoyándonos en los conceptos informales señalados anteriormente), no parece que puedan dar lugar a cadenas periódicas o reducibles, nos lleva a creer en principio en la convergencia del muestreo Gibbs.

2.- *Obtención de las distribuciones condicionales y estudio de las condiciones que deben satisfacer los modelos programables por BUGS.* En esta fase, se obtienen las distribuciones condicionales necesarias para aplicar el Gibbs Sampling, estudiándose en el caso de distribuciones no estándar, la log-concavidad de las densidades o funciones de probabilidad correspondientes. Y se comprueba que los modelos propuestos son modelos DAG obteniendo sus representaciones gráficas.

El estudio anterior y la propia definición del modelo demuestra para todos los modelos propuestos que se cumplen los requisitos para poder aplicar el programa BUGS.

3.- *Estudio de los modelos presentados con el programa BUGS.* Para el estudio a partir del Gibbs Sampling, se empleó la aproximación basada en una única cadena de Markov, esto es, un único muestreo por Gibbs basado en un único punto de arranque. Las fases de este análisis fueron:

- la traducción en BUGS del modelo,
- la obtención del punto de arranque y estudio del número de simulaciones para asumir la estabilidad del algoritmo Gibbs.
- obtención de una muestra y estudio de las características de las distribuciones predictivas y finales marginales a través de la misma.

Para la obtención del punto de arranque, se dejó libertad al programa BUGS en la elección de un primer punto de arranque a partir del cual se realizaron 1000 iteraciones del algoritmo Gibbs, el punto final resultante fue considerado como el punto de arranque definitivo de nuestro muestreo por Gibbs. Con este procedimiento se ha eludido la elección de un punto arbitrario de arranque y a la vez se ha intentado el establecimiento de un punto de arranque adecuado a las características del modelo.

A partir del anterior punto de arranque, se realizaron 5000 iteraciones Gibbs y se obtuvieron las aproximaciones del estimador de convergencia de Geweke(92) (ver (C.14)), para distintas variables del modelo (un grupo de 4 ó 5 variables, que se consideraron significativas), asumiendo la estacionariedad de la cadena cuando este valor pertenecía para todas ellas al intervalo al 95 % de una normal (0,1), y rechazando en caso contrario. En caso de rechazo se repitió hasta la aceptación de la estacionariedad, el proceso anterior añadiendo grupos con 2000 iteraciones adicionales.

Una vez asumida la estabilidad del proceso mediante el método anterior, se obtuvo una muestra de 1000 observaciones realizando 1000 iteraciones más del algoritmo Gibbs, a partir de la que se realiza un estudio de las distribuciones final y predictivas marginales; en este sentido se obtuvieron: estimaciones de la media y varianza (cuando teníamos garantizada su existencia) y de la mediana y algunos cuantiles (en los otros casos), así como en algunas ocasiones, gráficas de las densidades (o funciones de probabilidad en su caso) de las mismas, basadas en la aproximación de kernel para el intervalo definido por Silverman(86)[108].

La muestra anterior es una muestra correlada (método de muestreo (3)), si empleamos  $s_B^2$  (ver (C.14)) como una aproximación de la densidad espectral de frecuencia cero y analizamos la precisión de nuestras estimaciones de acuerdo con una estimación de (C.8) basada en esta aproximación, tenemos que:

- las estimaciones de las características poblacionales, pendientes y variables media, se realizan con una varianza del orden de  $10^{-4}$  y  $10^{-5}$ ,
- en las estimaciones de las varianzas para los modelos con cociente entre varianzas desconocido, del orden del  $10^{-4}$ , y para los modelos para rutas aleatorias de  $10^{-3}$ ,
- y en las estimaciones sobre los tamaños de la población y de las rutas aleatorias no muestreadas, del orden de  $10^{-3}$ .

Estas valoraciones de la varianza del estimador, a título orientativo, muestran que la "calidad" de la muestra en relación con la autocorrelación entre las observaciones varía de una variable a otra. En este sentido las muestras correspondientes a las características poblacionales, pendientes y variables media son las que proporcionan mejores resultados. Pero, en cualquier caso, se desprende que el tamaño de la muestra es suficiente (a pesar de la correlación entre las observaciones), ya que el error cometido en las estimaciones y estudios realizados a partir de la muestra es en el peor de los casos del orden de  $10^{-3}$ , reduciéndose en el caso de las distribuciones predictivas a un orden de  $10^{-5}$ .

# Bibliografía

- [1] Abramowitz, M. y Stegun, L. (1972). *Handbook of Mathematical Functions*. Dover: New York.
- [2] Aitchison, J. y Dunsmore, I.R. (1975). *Statistical prediction analysis*. Cambridge University Press. Cambridge, London: New York: Melbourne.
- [3] Basu, D. (1971). An essay on the logical foundations of survey sampling, part I. *Foundations of Statistical Inference*. Eds. V.P. Godambe and D.A. Sprott. Holt, Rinehart and Winston, Toronto, 203-233.
- [4] Basu, D. y Ghosh, J.K. (1967). Sufficient Statistics in Sampling from a finite universe. *Proc. 36th Session Internat. Statist. Inst.*. 850-859.
- [5] Bayarri, M.J. y Font, B. (1994). Bayesian Analysis of Random Routes. *Technical Report N.94-31, Depart. of Statist., Purdue University*. 1-9.
- [6] Bayarri, M.J. y Font, B. (1994). A (Bayesian) Note on Non-Random Samples from Finite Populations. *1994 Proceedings of the Section on Bayesian Statistical Science*. 246-251.
- [7] Bernardo, J.M. y Girón, J. (1991). Robust Sequential Prediction from Non-random Samples: The Election Night Case. *Bayesian Statistics 4*, 61-77. Clarendon Press, Oxford, UK.
- [8] Bolfarine, H., Pereira, C.A.B. y Rodrigues, J. (1987). Robust linear prediction in finite populations, a Bayesian perspective. *Sankhyā*. 49, 23-35.
- [9] Bolfarine, H. y Zacks, S. (1991). Bayes and Minimax Prediction in Finite Populations. *J. Statist. Planning and Inference*. 28(3), 139-151.
- [10] Bolfarine, H. y Zacks, S. (1992). *Prediction Theory for Finite Populations*. Springer-Verlag.
- [11] Brewer, K.R.W. (1963). Ratio estimation and finite populations. Some results deducible from the assumption of an underlying stochastic process. *Aust. J. Statist.*. 5, 93-105.
- [12] Brunk, H.D. (1980). Bayesian Least Squares Estimates of Univariate Regression Functions. *Communications in Statistics*. 9, 1101-1136.
- [13] BUGS Manual 0.30, *MRC Biostatistics Unit Cambridge*, 1994.

- [14] BUGS Examples 0.30.1, *MRC Biostatistics Unit Cambridge*, 1994.
- [15] Casella, G. y George, E.I. (1992). Explaining the Gibbs Sampler *J. Americ. Statist. Assoc.* **46**(3), 167-174.
- [16] Cassel, C.M , Särndal, C.E y Wretman, J.H (1976). Some results on generalized difference estimation and generalized regression estimation for finite populations. *Biometrika*. **63**(3), 615-620.
- [17] Cassel, C.M , Särndal, C.E y Wretman, J.H (1977). *Foundations of Inference in Survey Sampling*. John Wiley and Sons. New York.
- [18] Cochran, W.G. (1977). *Sampling Techniques. Third edition*. John Wiley and Sons. New York.
- [19] Cumberland, W.G. y Royall, R.M. (1981). Prediction Model and Unequal Probability Sampling. *J. Roy. Statist. Soc. (Ser. B)*. **43**, 353-367.
- [20] DeGroot, M.H. (1970). *Optimal Statistical Decisions*. New York: McGraw-Hill Book Co.
- [21] Devroye, L. (1986). *Non-uniform Random Variate Generation*. Springer-Verlag, New York.
- [22] Datta, G.S. y Ghosh, M. (1991). Bayesian prediction in linear models: applications to small area estimation. *Annals of Statistics*. **19**(4), 1748-1770.
- [23] Datta, G.S. y Lahiri, P. (1994). Robust hierarchical Bayes estimation of small area characteristics in presence of covariates. Preprint
- [24] DeGroot, M.H. (1970). *Optimal Statistical Decisions*. New-York: McGraw-Hill Book Co.
- [25] Devroye, L. (1986). *Non-uniform Random Variate Generation*. Springer-Verlag: New York.
- [26] Ericson, W.A. (1969). Subjective Bayesian Models in Sampling Finite Population. *J. Roy. Statist. Soc. (Ser. B)*. **31**, 195-233.
- [27] Ericson, W.A. (1969). A Note on the Posterior Mean of a Population Mean. *J. Roy. Statist. Soc. (Ser. B)*. **31**, 332-334.
- [28] Ericson, W.A. (1970). On the Posterior Mean and Variance of a Population Mean. *J. Americ. Statist. Assoc.* **65**, 649-652.
- [29] Ericson, W.A. (1983). A Bayesian approach to regression estimation in finite populations. *Technical Report N.120, Depart. of Statist., University of Michigan*. 1-21.
- [30] Ericson, W.A. (1988). Bayesian Inference in Finite Populations. *Handbook of Statistics*. **6**, 213-246.
- [31] Fay, R.E. y Herriot, R.A. (1979). Estimates of income for small places: An application of James-Stein procedures to census data. *J. Amer. Statist. Assoc.* **74**, 269-277.
- [32] Finetti, B de. (1937). La prévision: Ses lois logiques, ses sources subjectives. *Annales de l'Institut Henri Poincaré*. **7**, 1-68.

- [33] Font, B. (1995). Inferencia Bayesiana en Poblaciones Finitas: Un análisis comparativo. *Estadística Española*. Próxima publicación.
- [34] Gelfand, A.E., Hills, S.E., Racine-Poon, A. y Smith, A.F.M. (1990). Illustration of Bayesian inference in normal data models using Gibbs sampling. *J. Amer. Statist. Assoc.* **85**, 972-985.
- [35] Gelfand, A.E. y Smith, A.F.M. (1990). Sampling-based approach to calculating marginal densities. *J. Amer. Statist. Assoc.* **85**, 398-409.
- [36] Gelman, A. y Rubin, D.B. (1992). A single series from the Gibbs sampler provides a false sense of security. *Bayesian Statistics 4*. 625-631. Clarendon Press, Oxford, UK.
- [37] Gelman, A. y Rubin, D.B. (1992). Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science*. **7**(4), 457-511.
- [38] Geman, S. y Geman, D. (1984). Stochastic Relaxation, Gibbs Distributions and the Bayesian Restoration of Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. **6**, 721-741.
- [39] Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. *Bayesian Statistics 4*, 169-194. Clarendon Press, Oxford, UK.
- [40] Ghosh, M. (1992). Hierarchical and empirical Bayes multivariate estimation. *Current Issues in Statistical Inference. Essays in honor of D. Basu*. Ed. M. Ghosh y P.K. Pathak. **17**, 151-177.
- [41] Ghosh, M. y Lahiri, P. (1987). Robust Empirical Bayes Estimation of Means From Stratified Samples. *J. Americ. Statist. Assoc.* **82**, 1153-1162.
- [42] Ghosh, M. y Meeden, G. (1986). Empirical Bayes estimation in finite population sampling. *J. Americ. Statist. Assoc.* **81**, 1058-1062.
- [43] Ghosh, M. y Sinha, B.K. (1990). On the consistency between model and design-based estimators in Survey Sampling. *Communications in Statistics. Theory and Methods*. **19**(2), 689-702.
- [44] Gilks, W.R. (1992). Derivative-free adaptive rejection sampling for gibbs sampling. In *Bayesian Statistics 4*, 641-665. Clarendon Press, Oxford, UK.
- [45] Gilks, W.R., Best, N. y Tan, K. (1992). Adaptive rejection sampling from densities which are not necessarily log-concave. Technical Report, MRC. Cambridge, UK.
- [46] Gilks, W.R., Thomas, A. y Spiegelhalter, D.J. (1994). A language and program for complex Bayesian modelling. *The Statistician*. **43**, 169-178.
- [47] Gilks, W.R. y Wild, P. (1992). Adaptive rejection sampling for gibbs sampling. *Appl. Statist.* **41**, 337-348.
- [48] Godambe, V.P. (1955). A unified theory of sampling from finite population. *J. Roy. Statist. Soc. (Ser. B)*. **17**, 269-278.
- [49] Godambe, V.P. (1965). A review of the contributions towards a unified theory of sampling from finite populations. *Rev. Int. Statist. Inst.* **33**, 242-258.

- [50] Godambe, V.P. (1992). Survey Sampling -As I understand it. *Current Issues in Statistical Inference. Essays in honor of D. Basu*. Ed. M. Ghosh y P.K. Pathak. **17**, 187-195.
- [51] Godambe, V.P. y Joshi, V.M. (1965). Admissibility and Bayes estimation in sampling finite population I. *Ann. Math. Statist.* **36**, 1707-1722.
- [52] Goel, P.K. y DeGroot, M.H. (1980). Only Normal Distribution Have Linear Posterior Expectations In Linear Regression. *J. Amer. Statist. Assoc.* **75**, 895-900.
- [53] Hájek, J. (1949). Cluster sampling method of two phases (In Czech). *Statist. Obzor.* **29**, 384-394.
- [54] Hardle, W. (1991). *Smoothing techniques with implementation*. Springer-Verlag, New York.
- [55] Hartigan, J.A. (1969). Linear Bayesian Methods. *J. Roy. Statist. Soc. (Ser. B)*. **31**, 446-454.
- [56] Hansen, M.H., Dalenius, T. y Tepping, B.J. (1985). The development of sample surveys of finite populations. *Celebration of Statistics: The ISI Centenary Volume*. Eds. A.G. Atkinson y S.E. Fienberg, 327-354.
- [57] Hansen, M.H. y Hurwitz, W.N. (1943). On the theory of sampling from finite population. *Ann. Math. Statist.* **14**, 333-362.
- [58] Hess, I (1985). *Sampling for Social Research Surveys. 1947-1980*. Institute for Social Research Ann. Arbor. Mich.
- [59] Hewitt, E. y Savage, L.J. (1955). Symmetric measures on Cartesian products. *Trans. Amer. Math. Soc.* **80**, 470-501.
- [60] Hills, S.E. y Smith, A.F.M. (1992). Parameterization issues in Bayesian inference. *Bayesian Statistics 4*, 227-246. Clarendon Press, Oxford, UK.
- [61] Hinkley, D.V. (1979). Predictive likelihood. *Annals of Statistics*. **7**. 718-728.
- [62] Horvitz, D.G. y Thompson, D.J. (1952). A generalisation of sampling without replacement from a finite universe. *J. Amer. Statist. Ass.* **47**, 663-685.
- [63] Iglesias, P.L. (1993). Formas finitas do teorema de de Finetti: a visão preditivista da inferência estadística em populações finitas. Tesis doctoral. Instituto de Matemática e Estadística, Universidade de São Paulo.
- [64] Iglesias, P.L., Pereira, C.A.B. y Wechsler, S. (1994). Preditivistic statistical inference in finite populations. Presentado en *the Fifth Valencia International Meeting on Bayesian Statistics*, Alicante 5-9 junio de 1994.
- [65] Kish, L (1965). *Survey Sampling*. Wiley. New York.
- [66] Lahiri, D.B. (1951). A method of sample selection providing unbiased ratio estimates. *Bull. Inst. Statist. Inst.* **33**(2), 133-140.
- [67] Lahiri, P. (1990). Adjusted Bayes and empirical Bayes estimation in finite population sampling. *Sankhyā. Ser. B*. **42**, 50-66.

- [68] Lauritzen, S.L., Dawid, A.P., Larsen, B.N. y Leimer, H.G. (1990). Independence properties of directed Markow fields. *Networks*. 20, 491-505.
- [69] Lindley, D.V. (1961). The use of prior probability distributions in statistical inference an decisions. In *Proceedings of the Fourth Berkeley Symposium (Vol. 1)*. Berkeley: Univ. of California Press. 453-468.
- [70] Lindley, D.V. y Smith, A.F.M. (1972). Bayes Estimates for the Linear Model. *J. Roy. Statist. Soc. (Ser. B)*. 34, 1-41.
- [71] Louis, T.A. (1984). Estimating a population parameter values using Bayes and empirical Bayes methods. *J. Amer. Statist. Association*. 79, 393-398.
- [72] Malec, D. y Sedransk, J. (1985). Bayesian Inference for Finite Population Parameters in Multistage Cluster Sampling. *J. Amer. Statist. Association*. 80, 897-902.
- [73] Midzuno, H. (1952). On the sampling system with probability proportional to sum of sizes. *Ann. Inst. Statist. Math.*. 3, 99-107.
- [74] Morris, C. (1983). Parametric Empirical Bayes Inference: Theory and Applications. *J. Americ. Statist. Assoc.* 78, 47-65.
- [75] Mosteller y Tukey (1977). *Data Analysis and Regression*, Addison-Wesley.
- [76] Murgui, J.S. (1982). *Diseño e Inferencia en Poblaciones Finitas: Modelos de Superpoblación*. Tesis Doctoral. Universidad de Valencia.
- [77] O'Hagan, A. (1987). Monte Carlo is fundamentally unsound. *The Statistician*. 36. 247-249
- [78] O'Hagan, A. (1994). *Bayesian Inference*. Kendall's Library of Statistics. Editorial board: Green, Little, Ord, Scott y Weisberg.
- [79] Padmawar, V.R. (1981). A note on the comparison of certain sampling strategies. *J. Roy. Statist. Soc. (Ser. B)*. 43, 321-326.
- [80] Pérez, M.E. y Pericchi, L. R. (1991). Analysis of Multistage Survey as a Bayesian Hierarchical Model. *Reporte N.91-05 Dpto. de Matemáticas Puras y Aplicadas Univ. Simon Bolívar*.
- [81] Raftery, A.E. y Lewis, S.M. (1992). How many iterations in the Gibbs sampler?. *Bayesian Statistics 4*, 763-773. Clarendon Press, Oxford, UK.
- [82] Rao, C.R., (1973). *Linear Statistical Inference and its Applications*. Wiley, New York.
- [83] Rao, J.N.K., Hartley, H.O. y Cochran, W.G. (1962). On a simple procedure of unequal probability sampling without replacement. *J. Roy. Statist. Soc. (Ser. B)*. 24, 482-491.
- [84] Ripley, B.D. (1987). *Stochastic Simulation*. Wiley: New York
- [85] Roberts, G.O. (1992). Convergence diagnostic of the Gibbs sampler. *Bayesian Statistics 4*, 775-782. Clarendon Press, Oxford, UK.
- [86] Roberts, G.O. y Polson, N.G. (1992). A note on the geometric convergence of the Gibbs sampler. Sin publicar.

- [87] Roberts, G.O. y Smith, A.F.M (1992). Some convergence theory for Markow chain Monte Carlo. Sin publicar.
- [88] Rodrigues, J. (1988). Some results on restricted Bayes least squares predictors for finite populations. *South African Statistics Journal*. **22**, 45-53.
- [89] Rodrigues, J. ,Bolfarine, H. y Rogakto, A. (1985). A general theory of prediction in finite populations. *International Statistical Review*. **53**, 239-254.
- [90] Rodrigues, J. y Elian, S.N. (1989). The coordinate free estimation in finite population sampling. *Statistics and Probability Letters*. **7**, 293-295.
- [91] Royall, R.M. (1970). On finite population sampling theory under certain linear regression models. *Biometrika*. **57**, 377-387.
- [92] Royall, R.M. (1971). Linear regression models in finite population sampling theory. In *Foundations of Statistical Inference* (V.P. Godambe and D.A. Sprott Eds.). 259-274. Holt, Rinehart and Winston, Toronto.
- [93] Royall, R.M. (1976). The Linear Least-Squares Prediction Approach to Two-Stage Sampling. *J. Amer. Statist. Association*. **71**, 657-664.
- [94] Royall, R.M (1976). Likelihood function in finite population sampling theory. *Biometrika*. **63**, 605-614.
- [95] Royall, R.M. (1988). The Prediction Approach to Sampling Theory. *Handbook of Statistics*. **6**, 399-413.
- [96] Royall, R.M. (1992). The model based (prediction) approach to finite population sampling theory. *Current Issues in Statistical Inference. Essays in honor of D. Basu*. Ed. M. Ghosh y P.K. Pathak. **17**, 225-240.
- [97] Royall, R.M. y Pfeffermann, D. (1982). Balanced samples and robust Bayesian inference in finite population sampling. *Biometrika*. **69**, 401-409.
- [98] Rueda, M y Arcos, A. (1994). Un nuevo estimador de razón en muestreo estratificado. Presentado en forma de poster en el XXI Congreso Nacional de Estadística e Investigación Operativa, Calella 18-21 de abril de 1994.
- [99] Ruiz, M.C. (1986). El estimador de traslación para muestreos postcensales. *Estadíst. Española*. **111**, 81-86.
- [100] Ruiz, M.C. (1988). El teorema de Rao-Blackwell en poblaciones finitas. *Actas XVII. Reunión Nacional de Estadística, I.O. e Informática. Benidorm*.
- [101] Ruiz, M.C. (1988). Estimación Insesgada Generalizada en Poblaciones Finitas. *Questió*. **12(3)**, 315-321.
- [102] Ruiz, M.C. (1988). Un Ensayo sobre criterios de Inferencia en Poblaciones Finitas. *Questió*. **12(3)**, 323-332.
- [103] Sanchez - Crespo, J.L. (1977). A new sampling scheme: selection with graduate variable probabilities without replacement. *Bull. Internat. Statist. Inst.* **47**, 458-461.
- [104] Särndal, C.E. (1978). Design-based and Model-based Inference in Survey Sampling. *Scandinavian Journal of Statistics*. **5**, 27-52.

- [105] Schervish, M.J. y Carlin, B.P. (1990). On the convergence of successive substitution sampling. Sin publicar.
- [106] Scott, A. y Smith, T.M.F. (1969). Estimation in Multi-stage Surveys. *J. Amer. Statist. Association.* **64**, 830-840.
- [107] Sen, A.R. (1953). On the estimate of the variance in sampling with varying probabilities. *J. Indian. Soc. Agric. Statist.* **5**, 119-127.
- [108] Silverman, B.W. (1986). *Density Estimation for Statistics and Data Analysis*. Chapman and Hall: London.
- [109] Skiener, C.J. (1983). Multivariate prediction from selected samples. *Biometrika.* **70**, 289-292.
- [110] Smith, T.M.F. (1976). The foundations of survey sampling: a review. *J. Roy. Statist. Soc. (Ser. A)*. **139**(2), 183-204.
- [111] Smith, A.F.M. (1991). Bayesian computational methods. *Phil. Trans. R. Soc. Lond. A.* **337**, 369-386.
- [112] Smouse, E.P. (1982). Bayesian Least Squares Inference for Finite Population Models. Artículo presentado en *Institute for Mathematical Statistic meeting in Cincinnati, Ohio, August*
- [113] Smouse, E.P. (1984). A Note on Bayesian Least Squares Inference for Finite Population Models. *J. Amer. Statist. Association.* **79**, 390-392.
- [114] Sundberg, R. (1983). The Predictive Approach and Randomized Population Type Models for Finite Population Inference from two-stage Samples. *Scandinavian Journal Statistics.* **10**, 223-238.
- [115] Thomas, A., Spiegelhalter, D.J. y Gilks, W.R. (1992). BUGS: A program to perform Bayesian inference using Gibbs sampling. In *Bayesian Statistics 4*. 837-842. Clarendon Press, Oxford, UK.
- [116] Thomsen, I. y Tesfu, D. (1988). On the Use of Models in Sampling from Finite Populations. *Handbook of Statistics.* **6**, 369-397. P.H. Krishnaiah and C.R. Rao, eds.
- [117] Tierney, L. (1991). Markov chains for exploring posterior distributions. Sin publicar.
- [118] Whittaker, J. (1990). *Graphical Models in Applied Multivariate Statistics*. New York: Wiley.
- [119] Wild, P. y Gilks, W.R. (1993). AS 287. Adaptative rejection sampling from log-concave density functions. *Appl. Statist.* **42**, 701-709.



UNIVERSIDAD DE VALENCIA

FACULTAD DE CIENCIAS MATEMÁTICAS

Reunido el Tribunal que suscribe, en el día de la fecha,  
acordó otorgar, por unanimidad, a esta Tesis doctoral de

D. Maria Begonia Font Belauze

la calificación de Apto cum laude

Valencia, a 21 de Octubre de 1995

El Secretario,



Presidente

Verbatim  
DataLife  
MF  
**2HD**  
IBM FORMAT



ANALISIS BAYESIANO DE  
MUESTRAS NO ALEATORIAS EN  
POBLACIONES FINITAS

por M<sup>a</sup> Begoña Font Belaïre