

UNIVERSIDAD COMPLUTENSE DE MADRID

FACULTAD DE CIENCIAS MATEMÁTICAS

Departamento de Estadística e Investigación Operativa



TESIS DOCTORAL

**Aproximación bayesiana a los contrastes de
hipótesis múltiples con aplicaciones a los
microarrays**

TESIS DOCTORAL

MEMORIA PARA OPTAR AL GRADO DE DOCTOR

PRESENTADA POR

María Isabel Salazar Mendoza

Directores:

**Miguel Ángel Gómez Villegas
Luis Sanz San Miguel**

Madrid, 2011

ISBN: 978-84-694-6254-6

© María Isabel Salazar Mendoza, 2011

UNIVERSIDAD COMPLUTENSE DE MADRID

FACULTAD DE CIENCIAS MATEMÁTICAS

Departamento de Estadística e Investigación Operativa



**APROXIMACIÓN BAYESIANA A LOS
CONTRASTES DE HIPÓTESIS MÚLTIPLES
CON APLICACIONES A LOS
MICROARRAYS**

**MEMORIA PRESENTADA PARA OPTAR AL GRADO DE
DOCTOR POR**

M^a Isabel Salazar Mendoza

Bajo la dirección de los Doctores:

Miguel Ángel Gómez Villegas
Luis Sanz San Miguel

Madrid, 2011



**APROXIMACIÓN BAYESIANA A LOS
CONTRASTES DE HIPÓTESIS MÚLTIPLES
CON APLICACIONES A LOS
MICROARRAYS**

Tesis Doctoral de M^a Isabel Salazar Mendoza

Realizada bajo la dirección de
Dr. Miguel Ángel Gómez Villegas
Dr. Luis Sanz San Miguel

Madrid, 2011

Departamento de Estadística e Investigación Operativa
FACULTAD DE CIENCIAS MATEMÁTICAS
UNIVERSIDAD COMPLUTENSE DE MADRID

*A mi marido Luis, a mis hijos Luis y Beatriz
y a mis padres Cesáreo y Similia.*

Índice

Prólogo	v
1 Introducción	1
1.1 El problema de los contrastes de hipótesis múltiples	2
1.2 Motivación	5
1.3 Enfoque Frecuentista	9
1.3.1 Tasas de errores de tipo I	9
1.3.2 P-valores y p-valores ajustados	13
1.4 Enfoque bayesiano	18
1.4.1 Inferencia bayesiana	19
1.4.2 Aproximación Empírico Bayes	23
1.4.3 Decisión bayesiana	24
2 Hipótesis simples	33

2.1	Formulación del problema	35
2.2	Criterios de decisión bayesiana	38
2.3	Modelo Gaussiano. Varianza conocida	41
2.3.1	Ejemplo	43
2.3.2	Simulación: resultados y análisis de sensibilidad	44
2.4	Modelo Gaussiano. Varianza desconocida	55
2.4.1	Inferencia basada en los métodos MCMC. Muestreo de Gibbs	58
2.4.2	Ejemplo	65
2.4.3	Simulación: resultados y análisis de sensibilidad	65
2.5	Conclusiones	74
	Apéndice 2A: Modelo Gaussiano. Varianza desconocida. distribuciones finales condicionadas	77
3	Contraste de hipótesis bilaterales. Modelo Empírico Bayes	81
3.1	Introducción	82
3.2	Formulación del problema	84
3.3	Modelo Gaussiano	89
3.3.1	Algoritmo MCMC: muestreo de Gibbs	92
3.3.2	Simulación: resultados y análisis de sensibilidad	94
3.4	Modelo Gaussiano jerárquico	105

3.4.1	Simulación: resultados y análisis de sensibilidad	107
3.5	Modelo Gaussiano. Aproximación Empírico Bayes	110
3.5.1	Simulación: resultados y análisis de sensibilidad	112
3.6	Aplicación a datos reales procedentes de experimentos con biochips de ADN	120
3.7	Conclusiones y extensiones	124
	Apéndice 3A.1: Modelo Gaussiano. distribuciones finales condicionadas . .	128
	Apéndice 3A.2: Modelo Gaussiano jerárquico. distribución final condicionada	131

Prólogo

Es bien conocido que, cuando se contrasta más de una hipótesis a la vez, y cada hipótesis se contrasta de manera individual a un nivel α , la probabilidad de cometer al menos un error de tipo I aumenta rápidamente con el número de hipótesis. Por tanto, si el número de hipótesis es muy grande, siempre se rechazará alguna hipótesis aunque todas sean verdaderas, lo que puede tener graves consecuencias en las conclusiones.

Este problema no es nuevo, las bases de los contrastes de hipótesis múltiples se establecieron a finales de los años cuarenta y principios de los cincuenta por David Duncan, S.N. Roy, Henry Scheffé y John Tukey, aunque mucho antes aparecieron algunas de las ideas en trabajos de Fisher, Student y otros autores. Posteriormente, durante los años setenta, se publicaron un gran número de artículos relacionados con los contrastes de hipótesis múltiples. A partir de finales de los años noventa se ha reactivado de nuevo la investigación sobre este problema debido, fundamentalmente, a las investigaciones que utilizan los biochips de ADN en el campo de la genómica, donde es necesario, en muchas ocasiones, contrastar simultáneamente cientos e incluso miles de hipótesis.

Se han propuesto un gran número de métodos para tratar este problema, sin

embargo, no se ha dado con el método que sea válido en cualquier situación.

En esta memoria se trata el problema de los contrastes de hipótesis múltiples desde una perspectiva bayesiana. El objetivo es proporcionar una metodología general para el problema de los contrastes múltiples, con la finalidad de obtener la probabilidad final de cada hipótesis nula para decidir, en función de dicha probabilidad, qué hipótesis nulas se rechazan y qué hipótesis nulas se aceptan. Con este propósito, también se propone en este trabajo un criterio de decisión bayesiano, basado en la estimación del número de hipótesis nulas falsas.

La memoria está organizada en tres capítulos, cuyo contenido se resume brevemente a continuación.

En el capítulo 1 se introduce el problema de los contrastes de hipótesis múltiples, y se ofrece una visión general de su tratamiento desde una perspectiva frecuentista y desde una perspectiva bayesiana; igualmente se incluye una introducción a los métodos Empírico Bayes y a la teoría de la decisión bayesiana. Asimismo, se analiza la literatura existente en esta materia desde principios de los años cincuenta hasta la actualidad. También se describe, de un modo muy general, en qué consisten los biochips de ADN y sus principales aplicaciones en salud humana, puesto que las investigaciones que utilizan esta tecnología han sido la principal motivación del gran incremento de la literatura sobre contrastes múltiples en la última década.

En el capítulo 2, se trata el problema de los contrastes múltiples cuando tanto la hipótesis nula como la alternativa son simples e idénticas para todas las hipótesis. Este tipo de contrastes, aunque en general son poco realistas, nos van a permitir introducir de forma sencilla la metodología que también será utilizada en el capítulo 3. El objetivo, como se ha comentado anteriormente, es obtener la probabilidad final de cada hipótesis nula, sin embargo esta probabilidad no siempre se puede

obtener de forma analítica. Con el fin de aproximar la probabilidad final de cada hipótesis nula, se utiliza el método de Monte Carlo simulando directamente de la distribución inicial y, cuando esto no es posible, se utilizan los métodos de cadenas de Markov Monte Carlo (MCMC), de los que se incluye una breve introducción. Para decidir las hipótesis nulas que se rechazan y las hipótesis nulas que se aceptan, se describe la regla bayes cuando se considera, para cada acción conjunta, una función de pérdida aditiva y, para cada acción individual, la función de pérdida 0-1 generalizada. Además, en este capítulo, se propone un criterio de decisión bayesiano basado en la estimación del número de hipótesis nulas falsas. La metodología propuesta se aplica al modelo Gaussiano y se ilustra con datos simulados, realizando un análisis de sensibilidad a la elección de los parámetros de la distribución inicial; asimismo, se comparan los resultados con los obtenidos mediante el procedimiento frecuentista de Benjamini y Hochberg (1995) para controlar el FDR.

El capítulo 3 está dedicado al problema de los contrastes múltiples en los que la hipótesis nula es simple y la alternativa es bilateral. Este problema resulta especialmente interesante, ya que son muchas las situaciones, en cualquier campo de la investigación, donde se plantean estudios que se ajustan a este tipo de contrastes. Un ejemplo muy actual son los estudios realizados con biochips de ADN, en los que uno de los objetivos principales es la identificación de genes con expresión diferencial, es decir, estudiar si para cada gen existen diferencias significativas entre el nivel de expresión en tejido sano y el nivel de expresión en tejido enfermo. El objetivo, al igual que en el capítulo 2, es obtener la probabilidad final de cada hipótesis nula y decidir, en función de dicha probabilidad, las hipótesis nulas que se rechazan y las hipótesis nulas que se aceptan. En este capítulo, el modelo que se considera para este problema es evidentemente más complejo que el utilizado en el caso de hipótesis simples, ya que depende de muchos más parámetros, sin embargo, la metodología

es similar a la utilizada en el capítulo precedente, esto es, la inferencia bayesiana se lleva a cabo utilizando los métodos de cadenas de Markov Monte Carlo (MCMC); igualmente se aplican la regla bayes y el criterio basado en la estimación del número de hipótesis nulas falsas, que es el criterio que se propone en este trabajo. El procedimiento descrito se aplica también al modelo Gaussiano, siendo las medias los parámetros de interés. Además, en este capítulo se considera una aproximación Empírico Bayes para el modelo Gaussiano, con la finalidad de estimar el parámetro asociado a la varianza inicial de las medias del modelo, para evitar la sensibilidad detectada a la elección de dicho parámetro puesto que, con anterioridad, se propone un modelo jerárquico que no resuelve este problema de sensibilidad. La metodología que se desarrolla se ilustra con datos simulados y con datos reales procedentes de experimentos con biochips de ADN; igualmente se comparan los resultados con los obtenidos mediante el procedimiento de Benjamini y Hochberg (1995) para controlar el FDR.

Algunos de los contenidos de esta tesis han sido aceptados para su publicación en Ausín et al. (2010).

Finalmente, quiero agradecer a los directores de esta tesis, Miguel Ángel Gómez Villegas y Luis Sanz San Miguel, personas excepcionales tanto desde un punto de vista profesional como humano, la inestimable ayuda que me han proporcionado con sus consejos y aportaciones, así como su dedicación y paciencia. Quiero agradecerles también la gran oportunidad que, con la realización de esta tesis, me han dado de conocer y aprender de excelentes personas como son Concepción Ausín Olivera, Beatriz González Pérez y Teresa Rodríguez Bernal, miembros del Grupo de Investigación Métodos Bayesianos de la U.C.M., sin las que esta tesis no hubiese sido posible.

Mi agradecimiento también a mis compañeros, y sobre todo amigos, del Departamento de Producción Animal: Josefina Ávila, Teresa Castro, Jesús de la Fuente, Elisabet González de Chávarri, Sara Lauzurica, Concepción Pérez de la Unidad Docente de Zoología y muy especialmente a Joaquín Guerra, mi compañero de despacho durante muchos años, por sus ánimos y el interés que siempre han mostrado. Quiero agradecer igualmente a las profesoras de genética de dicho departamento, Blanca Nieto y Concepción Salgado, sus consejos y correcciones sobre los biochips de ADN.

Por último quiero dar las gracias a mi familia por su gran apoyo y comprensión.

Capítulo 1

Introducción

Los contrastes múltiples forman parte de un campo más amplio que es la inferencia múltiple o simultánea, que incluye estimación múltiple además de contrastes, y hace referencia al hecho de contrastar más de una hipótesis a la vez. Este tema no es nuevo, las bases de los contrastes de hipótesis múltiples se establecieron a finales de los años cuarenta y principios de los cincuenta, sin embargo sigue siendo de gran actualidad debido, fundamentalmente, a los rápidos avances tecnológicos que han tenido lugar a partir de finales de los años noventa, especialmente en las investigaciones que utilizan los biochips de ADN en el campo de la genómica, donde es preciso contrastar simultáneamente cientos e incluso miles de hipótesis.

Este es un capítulo introductorio que consta de cuatro secciones. En la sección 1.1, se introduce el problema de los contrastes de hipótesis múltiples, y se recogen algunas de las referencias bibliográficas más relevantes relacionadas con el tema hasta los años noventa. En la sección 1.2, se describe de un modo muy general en qué consisten los biochips de ADN, así como sus principales aplicaciones en salud humana. En la sección 1.3, se ofrece una visión general del tratamiento de los contrastes

múltiples desde una perspectiva frecuentista; se describen las distintas tasas de errores de tipo I y los p-valores y p-valores ajustados, que son ampliamente utilizados en los contrastes de hipótesis múltiples; también se incluye el q-valor introducido por Storey (2003), así como el procedimiento de Benjamini y Hochberg (1995) para controlar el FDR a un nivel α . En la sección 1.4, se expone una descripción general del tratamiento de los contrastes de hipótesis múltiples desde un punto de vista bayesiano; se incluye una introducción a los métodos Empírico Bayes y a la teoría de la decisión bayesiana. En estas dos últimas secciones, se citan también algunos de los trabajos más importantes sobre contraste de hipótesis múltiples de la última década.

1.1 El problema de los contrastes de hipótesis múltiples

Cuando se contrasta una sola hipótesis la idea básica es la siguiente: sea X una variable aleatoria con función de masa o de densidad $f(x|\theta)$, donde θ es un parámetro desconocido, del que se sabe que pertenece a un espacio paramétrico $\Theta \subset \mathbb{R}$, y se quiere contrastar la hipótesis nula $H_0 : \theta \in \Theta_0$ frente a la hipótesis alternativa $H_1 : \theta \in \Theta_1$, donde $\Theta_0 \cup \Theta_1 = \Theta$ y $\Theta_0 \cap \Theta_1 = \emptyset$, basándose en un estadístico $T(X)$.

Desde un punto de vista frecuentista, especificada una región de rechazo Γ , subconjunto del espacio muestral que conduce al rechazo de la hipótesis nula, se rechaza H_0 si el estadístico observado $T(x) = t$ pertenece a Γ , y no se rechaza H_0 cuando el estadístico no pertenece a dicha región.

La región de rechazo se elige de manera que la probabilidad de cometer error

de tipo II, es decir de aceptar la hipótesis nula cuando ésta es falsa, sea lo más pequeña posible, sujeta a la restricción de que la probabilidad de cometer error de tipo I, es decir de rechazar la hipótesis nula cuando ésta es verdadera, esté acotada por un valor α , fijado previamente y aceptablemente pequeño. A este valor α , se le denomina nivel de significación del test.

En el contraste de más de una hipótesis a la vez, el problema surge al tratar de evaluar todas las hipótesis en conjunto, ya que si se contrasta cada hipótesis de manera individual a un nivel α , sin tener en cuenta la multiplicidad del problema, la probabilidad de cometer al menos un error de tipo I aumenta rápidamente con el número de hipótesis. Es decir, al contrastar simultáneamente varias hipótesis es muy probable que, aunque todas sean ciertas, algunas se rechacen simplemente como consecuencia del azar. En otras palabras, si el número de hipótesis es muy grande, siempre se rechazará alguna aunque todas sean verdaderas y ello puede tener consecuencias importantes ya que las conclusiones deben afectar al conjunto de hipótesis como un todo, no a cada una de forma individual.

Para ilustrar este problema, supongamos que se quiere contrastar simultáneamente N hipótesis nulas H_i , $i = 1, \dots, N$. Si cada una de las hipótesis se contrasta a un nivel $\alpha = 0.05$ y los estadísticos correspondientes a cada una de ellas son independientes, en la tabla 1.1, extraída de Lehmann y Romano (2005, pág. 349), se muestra, para distintos valores de N , la probabilidad de cometer al menos un error de tipo I cuando todas las hipótesis son ciertas. Como se puede observar, cuando se contrastan simultáneamente cincuenta hipótesis, la probabilidad de cometer al menos un error de tipo I es muy elevada.

Se han propuesto numerosos métodos para tratar este problema, pero no se ha dado con la solución única que sea aceptable en cualquier situación. Las bases de

Tabla 1.1: Probabilidad de cometer al menos un error de tipo I cuando todas las hipótesis son ciertas, para diferentes valores de N , número de hipótesis.

N	1	2	5	10	50
$P(\text{al menos un error de tipo I})$	0.05	0.10	0.23	0.40	0.92

los contrastes simultáneos de hipótesis se establecieron a finales de los años cuarenta y principios de los cincuenta por David Duncan, S.N. Roy, Henry Scheffé y John Tukey, aunque algunas de las ideas básicas aparecieron mucho antes en trabajos de Fisher, Student y otros autores. Para una revisión histórica más completa ver Tukey (1953), Harter (1980), Miller (1981) y Shaffer (1988). El primer libro que se publicó sobre inferencia múltiple se debe a Miller (1966) que, por un lado, ayudó a popularizar el uso de los procedimientos de contrastes múltiples y por otro sirvió de estímulo para impulsar de nuevo la investigación en este campo.

Durante los años setenta se publicaron un gran número de artículos relacionados con los contrastes de hipótesis múltiples, entre los que mencionamos los de Aitkin (1969), Chew (1976a), Dunnett (1970), Dunnett y Goldsmith (1981), Games (1971), Gill (1973), Miller (1977, 1985), O'Neill y Wetherill (1971), Ryan (1959), Shaffer (1986b), Spjøtvoll (1974) y Thomas (1973). Un tratamiento completo de los contrastes de hipótesis múltiples aparece en Klockars y Sax (1986) y Hochberg y Tamhane (1987) que recogen su evolución en las dos décadas precedentes. Otros trabajos posteriores en este sentido se deben a Bauer et al. (1988), Braun (1994), Hsu (1996), Toothaker (1991, 1993), Hoppe (1993b) y Wesfall y Young (1993).

Desde finales de los años noventa, los rápidos avances tecnológicos, concretamente en el campo de la genómica con la utilización de los biochips de ADN, hicieron que se retomara con fuerza el problema de los contrastes múltiples, debido a que

en muchas de las investigaciones que utilizan esta tecnología se requiere contrastar simultáneamente miles de hipótesis.

1.2 Motivación

El problema de los contrastes de hipótesis múltiples se plantea con frecuencia en cualquier campo de la investigación, como los ensayos clínicos que comparan diferentes fármacos para el tratamiento de una determinada enfermedad, en el campo de la investigación biomédica; o en los experimentos agrícolas que comparan diferentes variedades de cultivos en términos de sus rendimientos, en el campo de la investigación agronómica. Otros muchos ejemplos que utilizan los contrastes múltiples en sus estudios, se presentan también en otras áreas como la investigación en psicología, educación y en general en cualquier campo de la investigación empírica o experimental.

Como ya se ha mencionado antes, en la última década se ha reactivado la investigación en el campo de los contrastes múltiples, principalmente por la necesidad que surge de contrastar simultáneamente un elevado número de hipótesis, en las investigaciones llevadas a cabo en el campo de la genómica con la utilización de los biochips de ADN.

Los biochips de ADN son una tecnología que permite a los investigadores obtener grandes cantidades de datos que corresponden a miles de genes. Uno de los objetivos en este tipo de experimentos es la identificación de genes con expresión diferencial; es decir, identificar los genes que varían su nivel de expresión en función del tejido analizado. Esta cuestión se puede plantear como un problema de contraste de hipótesis múltiples, en el que cada contraste individual estaría referido a un gen

y, por tanto, se contrastarían simultáneamente miles de hipótesis, cada una de ellas correspondiente a un gen.

El ADN contiene toda la información de la estructura y funcionamiento de un organismo. Pequeñas diferencias en la secuencia de ADN pueden tener un efecto importante sobre la salud y la enfermedad. Aunque todas las células poseen el código genético completo, todos los genes no se expresan en todas las células.

La expresión de un gen puede medirse al observar la cantidad de ARN mensajero o de la proteína elaborada con el ARN mensajero que produce ese gen. Los biochips son una herramienta que ayuda a identificar qué genes presentan altos niveles de expresión en un individuo enfermo, comparando los niveles de expresión de los genes con los de un individuo sano, y así poder asociar un grupo de genes con una determinada enfermedad.

Un biochip de ADN consiste en un gran número de moléculas de ADN ordenadas sobre un soporte sólido (vidrio, plástico, membranas, etc.), de manera que forman una matriz de secuencias en dos dimensiones, que permite la automatización simultánea de miles de ensayos con el fin de analizar la expresión y la variabilidad génica. A estos fragmentos de ADN inmovilizados en el soporte se les denomina sondas y cada una representa un gen diferente. En cada una de las casillas del chip se depositan millones de copias de una de estas sondas, por lo que cada casilla del chip representa un gen. Los ácidos nucleicos de las muestras a analizar se marcan por diversos métodos (enzimáticos, fluorescentes, etc.), el más utilizado es el fluorescente, y se incuban sobre el panel de sondas permitiendo la hibridación (reconocimiento y unión entre moléculas complementarias) de secuencias homólogas. Durante la hibridación, las muestras de material genético marcadas se unirán a sus complementarias inmovilizadas en el soporte del chip; posteriormente se procede al

revelado, que consiste en introducir el chip en un escáner provisto de un láser capaz de captar la fluorescencia. El resultado es una imagen con la fluorescencia de cada una de las casillas del chip, que permite observar si cada gen se ha expresado o no. Después se procede, mediante herramientas informáticas, a procesar la imagen que se ha generado. El propósito es obtener una medida cuantitativa de la intensidad, o los radios de cada uno de los puntos de hibridación, para obtener un valor que represente el nivel de expresión de cada uno de los genes. Los datos que se obtienen se preprocesan y normalizan con objeto de hacerlos más interpretables y prepararlos para el análisis posterior, así como para minimizar la variación no biológica propia de la aplicación de esta tecnología, permitiendo comparar los datos procedentes de distintos biochips.

Esta tecnología ha revolucionado la investigación en el campo de la genómica. Una de las limitaciones que existían era la necesidad de centrar los estudios en un pequeño número de genes o incluso, en cada gen de forma individual. La principal ventaja de esta nueva tecnología, frente a los métodos tradicionales, reside en la posibilidad de analizar simultáneamente miles de genes gracias a la automatización del proceso que permite depositar estas sondas moleculares, en forma de microgotas, en una superficie de pocos centímetros cuadrados. En poco más de una década, desde que Schena et al. (1995) utilizaron los primeros biochips de ADN para medir los niveles de expresión genética en plantas, se ha pasado de trabajos basados en el estudio de uno o unos pocos genes, a poder medir simultáneamente la expresión de todos los genes de un genoma, o al menos de una parte considerable de éste.

La utilización de los biochips de ADN está teniendo un gran impacto en muchas áreas de la investigación biomédica, un ejemplo es su utilización en la investigación sobre las alteraciones genéticas, que están en el origen de enfermedades tan comunes como el cáncer en seres humanos.

Actualmente los biochips están siendo aplicados, entre otros, en

- *Monitorización de la expresión génica.* Permite determinar cuál es el patrón de expresión génica y cuantificar el nivel de expresión de manera simultánea para un elevado número de genes. Esta técnica se utiliza, además, para realizar estudios comparativos de activación de determinados genes en tejidos sanos y enfermos y determinar así la función de los mismos.

- *Toxicología de fármacos.* Con el uso de los biochips es posible analizar los cambios de expresión génica que se dan durante la administración de un fármaco, así como la localización de nuevas posibles dianas terapéuticas y los efectos toxicológicos asociados.

- *Seguimiento de terapia.* Mediante el uso de los biochips se puede valorar cómo la herencia genética afecta a la respuesta individual a los fármacos, con el fin de personalizar la terapia para poblaciones específicas de pacientes, e incluso a nivel individual.

- *Medicina preventiva.* El conocimiento y posible diagnóstico de ciertos caracteres genéticos asociados a determinadas patologías, permite la prevención de las mismas antes de que aparezcan los síntomas.

Para una información más detallada ver Microarrays y Biochips de ADN y Aplicaciones de los Microarrays y Biochips en salud humana, Informe de Vigilancia Tecnológica. GENOMA ESPAÑA/CIBT-FGUAM.

Tabla 1.2: Posibles resultados cuando se contrastan simultáneamente N hipótesis.

Nº de hipótesis	aceptadas	rechazadas	Total
Verdaderas	U	V	N_0
Falsas	T	S	N_1
	W	R	N

1.3 Enfoque Frecuentista

Desde un punto de vista frecuentista, así como en el caso de una sola hipótesis la región de rechazo se obtiene con el fin de controlar el error de tipo I, en el caso de los contrastes de hipótesis múltiples es preciso, por lo ya comentado en la sección 1.1, controlar alguna medida relativa a la tasa de errores de tipo I que se describen a continuación. En esta sección se incluyen además los p-valores y los p-valores ajustados, así como el q-valor y el procedimiento de Benjamini y Hochberg (1995) para controlar el FDR a un nivel α .

1.3.1 Tasas de errores de tipo I

Consideramos el problema de contrastar simultáneamente N hipótesis nulas paramétricas H_i , $i = 1, \dots, N$. La situación se puede resumir en la tabla 1.2 propuesta por Benjamini y Hochberg (1995), donde N es el número de hipótesis que se contrastan y por tanto un número conocido, mientras que N_0 y N_1 son, respectivamente, el número de hipótesis ciertas y falsas, y son parámetros desconocidos; U , V , T y S son variables aleatorias no observables, siendo V el número de falsos positivos y T el número de falsos negativos; W y R son, respectivamente, el número de hipótesis

aceptadas y rechazadas y son variables aleatorias observables.

A continuación se describen las principales generalizaciones del error de tipo I a la situación de los contrastes de hipótesis múltiples tal como se muestran en Shaffer (1995).

- El *per-comparison error rate* (*PCER*) se define como la proporción del número esperado de errores de tipo I, es decir,

$$PCER = E[V]/N.$$

- El *per-family error rate* (*PFER*) se define como el número esperado de errores de tipo I, por tanto,

$$PFER = E[V].$$

- El *family-wise error rate* (*FWER*), que se define como la probabilidad de cometer al menos un error de tipo I, es decir,

$$FWER = Pr(V \geq 1).$$

Se dice que un procedimiento para contrastar múltiples hipótesis controla un determinado tipo de error PCER, FWER o PFER a un determinado nivel α si ese error es menor o igual que α cuando se aplica dicho procedimiento. En general, $PCER \leq FWER \leq PFER$. Así, para un nivel α previamente fijado, los procedimientos que controlan el PFER son más conservadores que los otros, es decir, conducen a menos rechazos que los que controlan el FWER o el PCER y los procedimientos que controlan el FWER son más conservadores que los que controlan el PCER.

Sobre qué tipo de error debería utilizarse en los problemas de contrastes múltiples, señalamos a continuación algunas opiniones. Tukey (1953) sugiere en general el control del FWER, rechazando el control del PCER debido a la filosofía subyacente de

éste, ya que los errores pueden incrementarse en proporción al número de hipótesis; y una de las razones por las que prefiere el control del FWER al del PFER, es que para familias infinitas el FWER se puede controlar pero no el PFER. Spjøtvoll (1972a) recomienda, para familias finitas, el control del PFER por ser técnicamente más fácil trabajar con él que con el FWER y, puesto que $\text{FWER} \leq \text{PFER}$, controlando el PFER también se controla el FWER, además es más instructivo pensar en términos del número esperado de falsos positivos, que en términos de la probabilidad de obtener al menos un falso positivo. Para una discusión más detallada ver Hochberg y Tamhane (1987). Dudoit et al. (2003) señalan que el FWER y el PFER se incrementan bruscamente con el número de hipótesis, mientras que el PCER permanece constante, si bien los procedimientos que controlan el PCER tienden a ignorar la multiplicidad del problema en su conjunto.

El false discovery rate (FDR)

Benjamini y Hochberg (1995) sostienen que en algunas situaciones en las que se puede estar dispuesto a tolerar algunos errores de tipo I, siempre que su número sea pequeño en comparación con el número de hipótesis rechazadas, el control del FWER puede dar lugar a procedimientos muy conservadores. Por este motivo, Benjamini y Hochberg (1995) introdujeron el *false discovery rate (FDR)*, otro tipo de tasa de errores que conduce a procedimientos menos conservadores.

El FDR se define como la proporción esperada de errores de tipo I entre las hipótesis rechazadas. Para definir este tipo de error debe tenerse en cuenta que la proporción de errores de tipo I entre las hipótesis rechazadas, V/R , es cero cuando el número de hipótesis que se rechazan, R , es cero. Por tanto, se define el *false discovery rate* como

$$FDR = E[V/R | R > 0] Pr(R > 0).$$

Otro aspecto importante a la hora de controlar los errores de tipo I es el *tipo de control del error*, que puede ser débil o fuerte. El control fuerte se refiere al control del error bajo alguna combinación de hipótesis nulas verdaderas. Por el contrario, el control débil se refiere al control del error sólo cuando todas las hipótesis son ciertas, es decir, bajo la hipótesis nula completa (en general, este último tipo de contraste es menos realista). El control fuerte garantiza que el error se controla bajo la verdadera distribución que genera los datos. En el contexto de los biochips de ADN, donde parece poco probable que ningún gen se exprese de manera diferencial, es especialmente importante tener un control fuerte del error.

En general, el FDR es menor o igual que el FWER, dándose la igualdad bajo la hipótesis nula completa. Por tanto, la propuesta de Benjamini y Hochberg controla el FWER en sentido débil y es menos conservador que éste en sentido fuerte.

Benjamini y Hochberg (1995) comparan también la potencia de los procedimientos que controlan el FDR y los que controlan el FWER, concluyendo que los métodos que controlan el FDR son generalmente más potentes que los que controlan el FWER en sentido fuerte.

En el contexto de los experimentos con biochips de ADN, los contrastes simultáneos se utilizan en muchas situaciones como un primer paso exploratorio en el proceso de identificación de subgrupos de genes, que pueden estar implicados en importantes procesos biológicos y ser candidatos potenciales para una investigación posterior más a fondo. De esta manera, obtener falsos positivos no distorsionaría las conclusiones en esta etapa del análisis, siempre y cuando su proporción sea pequeña. En este sentido, se puede estar más interesado en reducir el error de tipo II en lugar de minimizar el error de tipo I, es decir, se puede estar dispuesto a tolerar un mayor número de falsos positivos con el fin de obtener el mayor número posible de genes

de interés. Por esta razón, en el análisis de experimentos con biochips de ADN, el tipo de error más utilizado en el enfoque frecuentista es el FDR.

1.3.2 P-valores y p-valores ajustados

En el caso de una sola hipótesis nula H_i , una medida frecuentista ampliamente utilizada de la evidencia contra la hipótesis nula es el p-valor $p_i(t_i)$, que se puede considerar como el nivel más pequeño del test en el que la hipótesis H_i sería rechazada, dado el valor de un estadístico $T_i = t_i$,

$$p_i(t_i) = \sup_{\theta|H_i} Pr(T_i \in \Gamma_{t_i}|H_i).$$

Cuanto más pequeño sea el p-valor, $p_i(t_i)$, más fuerte será la evidencia contra la hipótesis nula H_i .

Fijado un cierto nivel de significación, α , rechazar H_i cuando $p_i \leq \alpha$ supone que el error de tipo I queda controlado a un nivel α .

Para la situación de los contrastes de hipótesis múltiples, se puede definir también un procedimiento en términos de los p-valores de las hipótesis individuales, rechazando H_i si $p_i \leq \alpha_i$, $i = 1, \dots, N$, donde p_i y α_i son, respectivamente, los p-valores y los niveles de significación individuales, estos últimos se eligen para controlar el error correspondiente (PCER, FDR, FWER o PFER) a un nivel α previamente fijado.

Alternativamente, se pueden considerar procedimientos de contrastes múltiples basados en los denominados *p-valores ajustados*. Dado un procedimiento de contraste de hipótesis múltiples, el p-valor ajustado correspondiente a una hipótesis H_i se define como el nivel nominal más pequeño del procedimiento entero en el que la hipótesis H_i se rechazaría, dados los valores de todos los estadísticos. Por ejemplo,

Yekutieli y Benjamini (1999) definen el p-valor ajustado para controlar el FDR como

$$\tilde{p}_i = \inf\{\alpha \in [0, 1] : H_i \text{ es rechazada en FDR nominal} = \alpha\}.$$

De forma similar se definen los p-valores ajustados para controlar otros tipos de error.

Dados los p-valores ajustados, se rechazarían aquellas hipótesis correspondientes a los p-valores ajustados menores o iguales que α . Como sucede en el caso de una sola hipótesis, una de las ventajas de utilizar los p-valores ajustados es que no se necesita fijar previamente el nivel del test.

Existen distintos procedimientos para controlar PCER, PFER, FWER y FDR a un nivel α . Una revisión general se puede ver en Dudoit et al. (2003). Benjamini y Hochberg (1995) demuestran que el siguiente procedimiento, bajo independencia de los estadísticos, controla el FDR en sentido fuerte a un nivel α , es decir, bajo alguna combinación de hipótesis nulas ciertas.

Sean $p_{(1)}, \dots, p_{(N)}$ los p-valores ordenados correspondientes a cada hipótesis nula. El procedimiento consiste en rechazar las hipótesis correspondientes a los p-valores $p_{(1)}, \dots, p_{(i^*)}$, donde

$$i^* = \max\{i : p_{(i)} \leq (i/N)\alpha\},$$

lo que es equivalente a rechazar las hipótesis correspondientes a los p-valores ajustados $\tilde{p}_{(1)}, \dots, \tilde{p}_{(i^*)}$, donde

$$i^* = \max\{i : \tilde{p}_{(i)} \leq \alpha\}$$

y

$$\tilde{p}_{(i)} = \min_{j=i, \dots, N} \left\{ \min \left(\frac{N}{j} p_{(j)}, 1 \right) \right\}$$

es el p-valor ajustado para este procedimiento. Si i^* no existe, no se rechaza ninguna hipótesis.

Benjamini y Yekutieli (2001) prueban que este procedimiento también controla el FDR bajo ciertas estructuras de dependencia de los estadísticos de los contrastes. También proponen una modificación del procedimiento que controla el FDR para estructuras de dependencia arbitrarias.

El positive false discovery rate (pFDR)

Cuando se controla el FDR a un nivel α , y se han rechazado algunas hipótesis, la proporción esperada de falsos positivos entre las hipótesis rechazadas se controla realmente a un nivel $\alpha/P(R > 0)$, por lo que Storey (2003) sugiere un tipo de error alternativo al FDR, el *positive false discovery rate* (*pFDR*) que es una modificación del anterior y se define como

$$pFDR = E[V/R|R > 0].$$

El término *positivo* se añade para reflejar el hecho de que se está condicionando a que se hayan producido resultados positivos. Además bajo la hipótesis nula completa resulta $pFDR = 1$, y no se podrían aplicar en éste caso los procedimientos basados en los p-valores. Los autores sugieren otro procedimiento que consiste en estimar directamente el pFDR y el FDR para una región de rechazo previamente fijada, al contrario de lo que se hace cuando se aplican procedimientos basados en los p-valores.

Storey (2003) da una interpretación bayesiana del pFDR. Concretamente demuestra que, cuando se efectúan N contrastes de hipótesis idénticos con estadísticos T_1, \dots, T_N , y supuesto que (T_i, H_i) son variables aleatorias independientes e idénticamente distribuidas, mediante una mixtura de las distribuciones bajo las hipótesis nula y alternativa, y para una región de rechazo Γ , el pFDR se puede escribir como una probabilidad final,

$$pFDR = Pr(H = 0|T \in \Gamma),$$

siendo $H_i = 0$ si la hipótesis i es cierta y $H_i = 1$ si es falsa, y tomando como distribución inicial para $H_i \sim \text{Bernoulli}(\pi)$ para $i = 1, \dots, N$ (la razón por la que en la expresión anterior del pFDR no se tiene en cuenta el índice es porque $Pr(H_i = 0|T_i \in \Gamma)$ es igual para cada $i = 1, \dots, N$). Asimismo demuestra que, si los estadísticos son dependientes y bajo ciertas condiciones, el pFDR se aproxima a la probabilidad final que se obtiene bajo independencia cuando el número de contrastes es grande.

El q-valor

Storey (2003) también introduce el denominado q-valor, que define, para un estadístico observado $T = t$, como

$$q - \text{valor}(t) = \inf_{\{\Gamma_\alpha: t \in \Gamma_\alpha\}} pFDR(\Gamma_\alpha)$$

y demuestra que, bajo las hipótesis anteriores, a partir de las cuales se obtiene que $pFDR = Pr(H = 0|T \in \Gamma)$,

$$q - \text{valor}(t) = \inf_{\{\Gamma_\alpha: t \in \Gamma_\alpha\}} Pr(H = 0|T \in \Gamma_\alpha),$$

por lo que este valor por un lado tiene una interpretación bayesiana al estar definido como una probabilidad final y por otro frecuentista, pues esta cantidad es, en términos del pFDR, análoga al p-valor.

Un problema del pFDR es que depende de la densidad inicial. De la Horra (2007) realiza un análisis de robustez bayesiana, considerando la clase de distribuciones iniciales ε -contaminadas, y demuestra que el pFDR no es robusto cuando la distribución inicial básica concentra la probabilidad en la hipótesis nula.

Recientemente, se han publicado un gran número de artículos relativos a contraste de hipótesis múltiples desde un punto de vista frecuentista y, sobre todo,

relacionados con el control del FDR. A continuación mencionamos algunos de estos trabajos.

Meinshausen y Rice (2006) consideran el problema de estimar el número de hipótesis nulas falsas, de entre un gran número de contrastes de hipótesis independientes, centrándose en el caso en que la proporción de hipótesis nulas falsas es muy pequeña. Establecen un límite inferior para esta proporción, basado en la distribución empírica de los p-valores de los estadísticos de los contrastes y para un determinado nivel de confianza $1 - \alpha$.

Storey (2007) propone un procedimiento para maximizar el número esperado de resultados verdaderos positivos, para cada número esperado de resultados falsos positivos fijo. Muestra que la optimización que se obtiene por este procedimiento está relacionada con la optimización en términos del FDR. Este procedimiento se puede considerar como una extensión, para los contrastes de hipótesis múltiples, del procedimiento de Neyman-Pearson para contrastar una sola hipótesis.

Jin y Cai (2007) consideran el problema de la estimación de la distribución normal nula y la estimación de la proporción de efectos no nulos. Desarrollan una aproximación basada en la función característica empírica.

Dudoit y van der Laan (2008) resumen en este libro sus trabajos más recientes sobre contrastes de hipótesis múltiples. El libro se divide en dos partes. En la primera presentan los fundamentos teóricos de diversas aproximaciones para contrastes múltiples: introducen un marco general para los procedimientos de contrastes múltiples, discuten la distribución de los estadísticos de los contrastes bajo la hipótesis nula, presentan una perspectiva general de los procedimientos de contrastes múltiples para el control de diferentes tipos de tasas de errores de tipo I, proponen distintos procedimientos para controlar estas tasas de errores de tipo I e

introducen dos nuevas aproximaciones de carácter general para contrastes múltiples. En la segunda parte aplican estos métodos a datos simulados y a datos reales procedentes de diferentes campos de la investigación biomédica y genómica. Asimismo se presenta software para este tipo de procedimientos de contrastes múltiples.

Gavrilov et al. (2009) estudian un procedimiento por pasos para controlar el FDR y lo comparan numéricamente con otros procedimientos que controlan también el FDR, mostrando que el procedimiento por pasos propuesto es el segundo más potente bajo independencia de los estadísticos, y también controla el FDR bajo cierta estructura de dependencia, aunque en este caso es mucho más conservador.

1.4 Enfoque bayesiano

Desde la perspectiva bayesiana, decidir entre la hipótesis nula y la alternativa, cuando se considera una sola hipótesis, consiste en calcular las probabilidades finales de ambas y decidir en consecuencia.

Cuando se contrastan N de hipótesis,

$$H_{0i} : \theta_i \in \Theta_{0i} \quad \text{frente a} \quad H_{1i} : \theta_i \in \Theta - \Theta_{0i}, \quad i = 1, \dots, N,$$

la distribución final es

$$\pi(\boldsymbol{\theta}|\mathbf{x}) \propto f(\mathbf{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}),$$

donde $\boldsymbol{\theta} = (\theta_1, \dots, \theta_N)$ y $\mathbf{x} = (x_1, \dots, x_N)$, siendo f y $\pi(\boldsymbol{\theta})$, la función de verosimilitud y la distribución inicial de $\boldsymbol{\theta}$, respectivamente.

La probabilidad final de cada hipótesis nula, $Pr(\theta_i \in \Theta_{0i}|\mathbf{x})$, se obtiene a partir de la marginal correspondiente a cada θ_i . Normalmente se requieren técnicas de

integración numérica que actualmente no suponen ningún problema a nivel computacional.

En esta sección se hace una introducción a la inferencia bayesiana, así como a los métodos Empírico Bayes y a la teoría de la decisión bayesiana.

1.4.1 Inferencia bayesiana

Para efectuar un análisis bayesiano sobre unos parámetros desconocidos se necesita modelizar, por un lado, las observaciones mediante una determinada distribución de probabilidad y, por otro, la incertidumbre sobre dichos parámetros, considerando una distribución inicial o a priori, para posteriormente obtener la distribución final con la que se realizarán todas las inferencias.

En general, no es fácil obtener una expresión analítica para la distribución final, pero la inferencia bayesiana se puede realizar utilizando los métodos de Monte Carlo basados en cadenas de Markov (MCMC). Bajo algunas condiciones, dado un valor inicial $\theta^{(0)}$, la aproximación MCMC produce una cadena de Markov $\{\theta^{(j)} : j = 1, \dots, M\}$, donde $\theta^{(j)} = \{\theta_1^{(j)}, \dots, \theta_N^{(j)}\}$, que tiene como distribución de equilibrio la distribución final, $\pi(\boldsymbol{\theta}|\mathbf{x})$, más detalles pueden verse en Robert y Casella (2004). El desarrollo de estos métodos ha contribuido notablemente a que, en los últimos años, haya habido un importante aumento del empleo de la estadística bayesiana en la investigación. Un ejemplo son los contrastes múltiples que, como apunta Berger(2006), se encuadran de forma automática en la metodología bayesiana y su tratamiento no requiere de grandes tamaños muestrales.

Si las observaciones son condicionalmente independientes, con parámetros dis-

tintos, y se considera que $\theta_1, \dots, \theta_N$ son independientes a priori, entonces

$$\pi(\theta_1, \dots, \theta_N | x_1, \dots, x_N) \propto \pi(\theta_1 | x_1) \dots \pi(\theta_N | x_N)$$

y $\theta_1, \dots, \theta_N$ son también independientes a posteriori, por tanto, la distribución final de θ_i solo depende de x_i y las demás observaciones son irrelevantes a la hora de calcular esta distribución, por lo que el procedimiento no tendría en cuenta la multiplicidad del problema. En este sentido, el enfoque bayesiano es similar a los procedimientos frecuentistas que controlan el PCER, tal y como se refleja en Berry y Hochberg (1999). No parece sin embargo, que sea muy razonable considerar, en general, distribuciones iniciales independientes.

Si no se considera independencia a priori, la probabilidad final $Pr(\theta_i \in \Theta_{0i} | \mathbf{x})$, para cada hipótesis, depende de las observaciones de todos los contrastes y por tanto el procedimiento bayesiano sí tendría en cuenta la multiplicidad del problema. Además, en el contexto de los experimentos con biochips de ADN, esta última cuestión es una ventaja de los modelos bayesianos sobre los métodos frecuentistas, pues los métodos bayesianos se adaptan bien al tamaño pequeño de las muestras que resultan en este tipo de experimentos, ya que utilizan la información de todos los genes para estimar los parámetros del modelo, y además se pueden modelizar explícitamente todas las fuentes de variabilidad.

La distribución inicial no siempre se puede precisar completamente. Por ejemplo, se sabe que pertenece a una clase de distribuciones pero no se conocen alguno o algunos de los parámetros de los que depende. Desde un punto de vista bayesiano, la solución sería introducir un nivel más en la jerarquía, incluyendo una distribución inicial para los hiperparámetros de la distribución inicial de los parámetros del modelo, dando lugar, en algunas ocasiones, a modelos jerárquicos muy complejos.

Cabe desatacar, en este sentido, y relacionados con el problema de los contrastes

múltiples en el contexto de los experimentos con biochips de ADN, los siguientes trabajos.

Baldi y Long (2001) establecen un marco probabilístico bayesiano general para el análisis de datos en los experimentos con biochips de ADN, mediante un modelo con distribución normal para las observaciones y distribuciones iniciales conjugadas.

Ibrahim et al. (2002) proponen como distribución para las observaciones una mezcla de dos componentes, una discreta y otra continua, esta última con distribución lognormal y un algoritmo de selección de modelos basado en la L medida.

Do et al. (2005) proponen una aproximación bayesiana no paramétrica para el estudio de genes con expresión diferencial. El modelo de probabilidad que consideran es una mezcla de distribuciones normales. Los resultados que obtienen son similares a los de la aproximación Empírico Bayes propuesta por Efron et al. (2001) para el mismo problema; además demuestran, mediante un estudio con datos simulados y datos procedentes de biochips de ADN, que la aproximación que proponen reduce el sesgo que se produce en la aproximación Empírico Bayes, al sobreestimar esta última la proporción de genes que no se expresan diferencialmente.

Lönnstedt y Britton (2005) presentan dos modelos bayesianos completos para detectar genes con expresión diferencial; también comparan los resultados obtenidos con estos dos modelos y otros métodos Empírico Bayes. La principal conclusión que obtienen es que, para los datos analizados, los modelos bayesianos completos no mejoran los métodos Empírico Bayes, por lo que este estudio da cierta confianza a favor del uso de métodos Empírico Bayes para datos procedentes de biochips de ADN.

Gottardo et al. (2006) introducen un modelo bayesiano jerárquico, que es robusto

respecto a valores extremos y potente incluso con un pequeño número de replicas, motivados por el hecho de que los datos procedentes de experimentos con biochips de ADN, a menudo contienen valores extremos, debido a los muchos pasos implicados en el proceso experimental.

Scott y Berger (2006) consideran un modelo bayesiano completo. Estudian cuestiones relacionadas con el problema de la elección de las distribuciones iniciales y realizan un análisis de sensibilidad a la elección de la distribución inicial de la probabilidad asignada a la hipótesis nula, obteniendo gran sensibilidad a las especificaciones iniciales.

Cabras (2010) muestra que los procedimientos de contrastes múltiples, basados en los p-valores, no son robustos cuando éstos no son uniformes. Propone que la distribución de los p-valores, bajo la hipótesis nula, no dependa del parámetro de interés. Para ello utiliza un procedimiento basado en el p-valor predictivo a posteriori parcial, que permite extender el control del FDR a la clase general de hipótesis nulas compuestas. También aplica el procedimiento, en el contexto de los experimentos con biochips de ADN, para detectar genes con expresión diferencial, considerando para el nivel de expresión un modelo gamma.

Ausín et al. (2010) proponen un modelo bayesiano jerárquico basado en una mixtura de las distribuciones bajo las hipótesis nula y alternativa. Aplican el procedimiento al modelo normal considerando distribuciones iniciales conjugadas y emplean el muestreo de Gibbs. También muestran, mediante un ejemplo de simulación, que este procedimiento mejora el de Benjamini y Hochberg (1995) para controlar el FDR, en el sentido de que reduce el porcentaje de falsos negativos manteniendo un aceptable porcentaje de falsos positivos. Asimismo, realizan un análisis de sensibilidad a la elección de los parámetros de las distribuciones iniciales, obteniendo

que el procedimiento es robusto, excepto a la elección del parámetro asociado a la varianza inicial de las medias del modelo.

1.4.2 Aproximación Empírico Bayes

Una alternativa, cuando los hiperparámetros son desconocidos es estimarlos, habitualmente por máxima verosimilitud, y después continuar con el análisis bayesiano, es decir, calcular la probabilidad final usando los valores estimados de los hiperparámetros. Este enfoque es conocido como método Empírico Bayes y fue introducido por Robbins (1955) con una formulación no paramétrica. Posteriormente, los trabajos de Efron y Morris (1972, 1973, 1975) y Morris (1983) han sido decisivos para el desarrollo del análisis Empírico Bayes actual.

Efron et al. (2001) y Efron (2004) tratan el problema de los contrastes múltiples, para detectar genes con expresión diferencial en los experimentos con biochips de ADN, desde un punto de vista Empírico Bayes no paramétrico. Consideran un modelo de mixturas de dos componentes, y para cada gen, basan la inferencia en la estimación de una versión local del FDR, el *local false discovery rate*, $fdr(z) = p_0 f_0(z)/f(z)$, estimando p_0 y la proporción $f_0(z)/f(z)$, donde p_0 es la probabilidad inicial de que un gen no presente expresión diferencial, f_0 la densidad empírica nula y $f(z)$ la mixtura de las densidades de las hipótesis nula y alternativa, estimada de los datos.

Efron (2008), bajo modelos de mixturas de dos componentes, analiza la interacción entre la ideas frecuentistas, centrándose en el control del FDR, y las ideas bayesianas. Discute el modelo de mixturas de dos componentes en un entorno bayesiano y la necesidad, en muchas ocasiones, de los métodos Empírico Bayes

para llevar a cabo estas ideas en la práctica. Estudia la elección de la hipótesis nula y cuestiones como la potencia y los efectos de la correlación, entre otras.

Con un enfoque Empírico Bayes paramétrico cabe señalar, entre otros, el trabajo de Casella (2001), que establece un marco Empírico Bayes genérico usando el muestreo de Gibbs y considerando un modelo jerárquico general. Además, en el contexto de los experimentos con biochips de ADN para detectar genes con expresión diferencial, Newton et al. (2001) realizan un análisis Empírico Bayes usando un modelo jerárquico gamma-gamma y considerando dos condiciones distintas para cada gen. Kendziorski et al. (2003) amplian el trabajo anterior a múltiples condiciones y consideran dos parametrizaciones distintas, el modelo gamma-gamma y el lognormal-normal. Lönnstedt y Speed (2002) utilizan, igualmente, el modelo lognormal-normal, y Newton et al. (2004) establecen dos métodos de estimación según el tratamiento, paramétrico y no paramétrico, de la distribución inicial. Ausín et al. (2010) consideran una aproximación Empírico Bayes para estimar el parámetro asociado a la varianza inicial de las medias del modelo, obteniendo buenos resultados y resolviendo, de este modo, el problema de sensibilidad que existe a la elección de este parámetro.

1.4.3 Decisión bayesiana

Todos los métodos de inferencia estadística implican la toma de una decisión. En el caso de los contrastes múltiples, la decisión consiste en ver qué hipótesis se rechazan y qué hipótesis se aceptan. La teoría de la decisión es el marco teórico que permite estudiar globalmente los problemas de inferencia estadística como un único tipo de problema: la toma de una decisión. Asimismo, la teoría de la decisión proporciona un gran refuerzo para la lógica del enfoque bayesiano. Para una revisión completa

ver Berger (1985).

Los elementos necesarios para plantear un problema de inferencia estadística desde la perspectiva de la teoría de la decisión son:

- Una cantidad desconocida θ sobre la cual se quiere realizar el proceso de decisión, comunmente se le denomina estado de la naturaleza, y el conjunto de todos los posibles estados de la naturaleza se denota por Θ . Normalmente, cuando se realiza un experimento para obtener información sobre θ , éste se diseña de manera que las observaciones tengan como posibles distribuciones de probabilidad $\mathcal{P} = \{f_\theta : \theta \in \Theta\}$, donde f_θ es una función de densidad o una función de masa de probabilidad, es decir, una familia paramétrica de distribuciones de probabilidad que dependen del parámetro desconocido θ . En esta situación, θ es el parámetro, Θ el espacio paramétrico y \mathcal{X} el espacio muestral.

- El conjunto de todas las *acciones posibles*, que habitualmente se denota por \mathcal{A} y una acción particular por a .

- La *función de pérdida* L ,

$$\begin{aligned} L : \Theta \times A &\longrightarrow \mathbf{R} \\ (\theta, a) &\longrightarrow L(\theta, a) \end{aligned}$$

para valorar el coste de tomar la decisión a cuando el verdadero estado de la naturaleza es θ .

- Una *regla de decisión* (no aleatorizada), se define como la siguiente función medible,

$$\begin{aligned} \delta : \mathcal{X} &\longrightarrow A \\ \mathbf{x} &\longrightarrow \delta(\mathbf{x}) = a \end{aligned}$$

e indica, para cada \mathbf{x} de \mathcal{X} , la acción $\delta(\mathbf{x}) = a$ que se debe tomar cuando se ha

observado \mathbf{x} , donde \mathbf{x} es un vector n dimensional. El conjunto de reglas de decisión aceptables en un problema se denota por \mathcal{D} .

- La *función de riesgo* de una regla de decisión $\delta(\mathbf{x})$, que se define como

$$R(\theta, \delta) = E_{f(\mathbf{x}|\theta)}[L(\theta, \delta(\mathbf{x}))]$$

y mide la pérdida esperada si se usa la regla δ y el verdadero estado de la naturaleza es θ .

Como θ es desconocido, es deseable usar una regla de decisión que tenga valores bajos de la función de riesgo para todos los posibles valores $\theta \in \Theta$. En general, seleccionar una regla de decisión mediante la función de riesgo es difícil, ya que puede ocurrir que una regla de decisión tenga menor riesgo que otra para determinados valores de θ y ocurra lo contrario para otros valores de θ . En teoría de la decisión existen distintos principios para seleccionar reglas de decisión óptimas, uno de ellos es el principio del riesgo bayes:

Se define el *riesgo bayes* de la regla de decisión δ con respecto a una distribución inicial π en Θ como

$$r(\pi, \delta) = E_{\pi}[R(\theta, \delta)].$$

Puesto que el riesgo bayes de una regla de decisión resume en un solo número el comportamiento de dicha regla sobre todo el espacio paramétrico, se puede seleccionar la regla de decisión que minimice el riesgo bayes.

Principio del riesgo bayes: una regla de decisión δ_1 es preferible a una regla δ_2 si

$$r(\pi, \delta_1) < r(\pi, \delta_2).$$

La regla de decisión que minimiza $r(\pi, \delta)$ se llama regla bayes y se denota por δ^{π} ,

es decir, δ^π es tal que

$$r(\pi, \delta^\pi) = \min_{\delta \in \mathcal{D}} r(\pi, \delta).$$

La regla bayes puede no existir, o bien puede no ser única.

Desde una perspectiva bayesiana, θ es una variable aleatoria y el método más natural de proceder es observar la pérdida esperada de una acción con respecto a la distribución de θ en el momento de tomar la decisión; es decir, después de observar \mathbf{x} , esta distribución es la distribución final de θ , $\pi(\theta|\mathbf{x})$. A continuación se definen estos conceptos.

Se define la *pérdida esperada a posteriori o final* de una acción a como

$$\rho(\pi(\theta|\mathbf{x}), a) = E_{\pi(\theta|\mathbf{x})}[L(\theta, a)] = \int_{\Theta} L(\theta, a)\pi(\theta|\mathbf{x})d\theta. \quad (1.1)$$

La *acción bayes a posteriori*, $a^*(\mathbf{x})$, es la acción que minimiza $\rho(\pi(\theta|\mathbf{x}), a)$, o equivalentemente la que minimiza

$$\int_{\Theta} L(\theta, a)f(\mathbf{x}|\theta)\pi(\theta)d\theta. \quad (1.2)$$

Obtener la regla bayes, δ^π , que minimiza el riesgo bayes, $r(\pi, \delta)$, es más complicado que obtener la acción bayes que minimiza la pérdida esperada final, $\rho(\pi(\theta|\mathbf{x}), a)$. Sin embargo, los dos problemas son equivalentes ya que la regla bayes, δ^π , se puede obtener eligiendo la acción bayes, $a^*(\mathbf{x})$, para cada \mathbf{x} con $m(\mathbf{x}) > 0$, siendo $m(\mathbf{x})$ la densidad marginal de \mathbf{X} , es decir, la acción que minimice la pérdida esperada final (1.1) o equivalentemente (1.2), y de forma arbitraria cuando sea $m(\mathbf{x}) = 0$.

Cuando se contrasta una sola hipótesis,

$$H_0 : \theta \in \Theta_0 \quad \text{frente a} \quad H_1 : \theta \in \Theta_1 = \Theta - \Theta_0,$$

el conjunto de posibles estados de la naturaleza y el espacio de acciones constan solo de dos elementos. Si se denota por $\{H_0 = 0\}$ si H_0 es cierta y por $\{H_0 = 1\}$

si H_0 es falsa o, equivalentemente, si H_1 es cierta, el espacio paramétrico es $\Theta = \{\{H_0 = 0\}, \{H_0 = 1\}\}$ y el espacio de acciones es $A = \{a_0, a_1\}$, donde a_0 es la acción que lleva a aceptar H_0 y a_1 la que lleva a rechazar H_0 o, equivalentemente, aceptar H_1 .

La función de pérdida más utilizada, en este contexto, es la función de pérdida 0-1 generalizada,

$$L(\theta, a_i) = \begin{cases} 0 & \text{si } \theta \in \Theta_i \\ C_i & \text{si } \theta \in \Theta_j \quad \text{con } i, j \in \{0, 1\} (j \neq i) \end{cases}$$

entonces, las pérdidas esperadas finales de a_0 y a_1 son, respectivamente, $C_0 Pr(\Theta_1|\mathbf{x})$ y $C_1 Pr(\Theta_0|\mathbf{x})$.

La decisión, desde una perspectiva bayesiana, será elegir la acción con menor pérdida esperada final.

Cuando se contrastan simultáneamente N hipótesis,

$$H_{0i} : \theta_i \in \Theta_{0i} \quad \text{frente a} \quad H_{1i} : \theta_i \in \Theta_{1i} = \Theta - \Theta_{0i}, \quad i = 1, \dots, N,$$

podemos representar el conjunto de todos los posibles estados de la naturaleza por

$$\Theta = \bigcup_{j=1}^{2^N} \Theta_j,$$

donde

$$\Theta_j = \bigcap_{i=1}^N \{H_{0i} = \varepsilon_{ij}\}, \quad j = 1, \dots, 2^N$$

y siendo $\varepsilon_{ij} = 0$ si H_{0i} es cierta y $\varepsilon_{ij} = 1$ si H_{0i} es falsa o, equivalentemente, si H_{1i} es cierta.

De forma análoga, el espacio de acciones se puede expresar del siguiente modo

$$A = \bigcup_{j=1}^{2^N} A_j,$$

donde

$$A_j = \bigcap_{i=1}^N a_{e_{ij}}$$

y siendo $e_{ij} = 0$ si se acepta H_{0i} y $e_{ij} = 1$ si se rechaza H_{0i} o, equivalentemente, si se acepta H_{1i} . El conjunto de posibles estados de la naturaleza y el espacio de acciones, en este caso, constan de 2^N elementos.

Lehmann (1957a,b) fue el primer autor que consideró el problema de los contrastes múltiples desde el punto de vista de la teoría de la decisión bajo una perspectiva frecuentista. Desarrolló un procedimiento óptimo considerando una función de pérdida aditiva; así, el problema de minimizar la función de riesgo o el riesgo bayes se puede resolver minimizando separadamente cada una de las componentes del problema; esto lleva a un procedimiento que es óptimo para cada comparación, sin tener en cuenta las demás comparaciones. Posteriormente, Spjøtvoll (1972a), siguiendo la aproximación frecuentista de Neyman-Pearson y considerando el *per-family error rate* (PFER), deduce el procedimiento de contrastes múltiples más potente maximizando la potencia para cada contraste individual. El resultado principal de los trabajos de Lehmann y Spjøtvoll es que si se usa un procedimiento óptimo para cada componente del problema, el procedimiento producto resultante es óptimo para el problema de contrastes múltiples en su conjunto.

Desde un punto de vista de la teoría de la decisión bayesiana, Duncan (1961, 1965) desarrolló un procedimiento óptimo para la comparación de todos los pares de medias en un diseño equilibrado de una vía. Utilizando distribuciones iniciales normales e independientes, dedujo la regla bayes bajo la estructura de una función de pérdida aditiva que depende de las diferencias de medias reales. Esta regla no depende del número de comparaciones por lo que, en este sentido, el procedimiento bayesiano para este problema tiene la naturaleza de los procedimientos que con-

trolan el PCER, es decir, ignora la multiplicidad del problema en su conjunto; sin embargo, el procedimiento sí depende de la heterogeneidad entre las medias de los tratamientos. Hochberg y Tamhane (1987) presentan una descripción detallada de los trabajos de Lehmann, Spjøtvoll y Duncan.

Shaffer (1999) modifica el procedimiento de Duncan para controlar el FWER al nivel 0.05 en sentido débil. Mediante un estudio de simulación compara el método modificado con distintos procedimientos frecuentistas, obteniendo resultados muy similares a los obtenidos por el procedimiento de Benjamini y Hochberg (1995) para controlar el FDR.

Lewis y Thayer (2004), siguiendo los trabajos de Shaffer (1999) y Duncan (1965), aplican la teoría de la decisión bayesiana al problema de contrastes múltiples en un diseño con efectos aleatorios. Consideran la función de pérdida 0-1 y demuestran que la regla bayes controla el FDR, proporcionando un apoyo teórico a las conclusiones de Shaffer (1999) basadas en un estudio de simulación. Los métodos de Lewis y Thayer (2004) y Shaffer (1999) se pueden considerar como procedimientos semi-bayesianos, ya que pretenden controlar alguna medida frecuentista del error.

Scott y Berger (2006), en el contexto de los experimentos con biochips de ADN, tratan el problema de contrastes múltiples desde una perspectiva de la teoría de la decisión bayesiana. Uno de los objetivos principales, en este tipo de experimentos, es encontrar los genes que están activos, por lo que, para cada gen, hay dos posibles acciones: clasificar el gen como activo o clasificar el gen como inactivo. Consideran la pérdida, por clasificar erróneamente un gen activo como inactivo, proporcional a la distancia a cero del nivel medio de expresión correspondiente a cada gen, y en una unidad si un gen inactivo se clasifica como activo. Así, para los genes activos clasificados como inactivos, la pérdida será mayor cuanto mayor sea el nivel medio

de expresión del gen, es decir, cuanto más activo esté. Para especificar esta función de pérdida sólo se necesita precisar una cantidad (la constante de proporcionalidad). La regla de decisión será elegir, para cada gen, la acción con menor pérdida esperada final.

Este planteamiento con esta notación va a ser utilizado en el capítulo siguiente, donde se propone un nuevo criterio de decisión bayesiano y se describe la regla bayes cuando se considera, para cada acción conjunta, una función de pérdida aditiva y, para cada acción individual, la función de pérdida 0-1 generalizada.

Capítulo 2

Hipótesis simples

En este capítulo se plantea, desde una perspectiva bayesiana, el problema de los contrastes múltiples cuando tanto la hipótesis nula como la alternativa son simples e idénticas para todas las hipótesis. Los ejemplos en los que ambas hipótesis son simples resultan, en general, poco realistas. Sin embargo, sí son interesantes desde un punto de vista teórico, porque además de estar asociado a ellos uno de los resultados fundamentales en la cuestión de los contrastes de hipótesis, como es el Lema de Neyman-Pearson, sirven como introducción sencilla para problemas más generales.

El objetivo, en este capítulo, es proporcionar un marco general para el problema de los contrastes múltiples cuando ambas hipótesis son simples, con la finalidad de obtener la probabilidad final de cada hipótesis nula, así como la de elegir un valor de corte, de manera que todas las hipótesis nulas con probabilidad final menor o igual que este valor serían rechazadas, aceptando el resto.

Este capítulo consta de cinco secciones. En la sección 2.1, se formula el problema de manera general, asumiendo que las observaciones correspondientes a cada

hipótesis son muestras de variables aleatorias i.i.d., mediante una mixtura de las densidades bajo las hipótesis nula y alternativa. En esta sección, con objeto de aproximar la probabilidad final de cada hipótesis nula, cuando ésta no se pueda obtener de forma analítica, se utiliza el método de Monte Carlo, simulando directamente de la distribución inicial.

En la sección 2.2, se describe la regla bayes cuando se considera, para cada acción conjunta una función de pérdida aditiva y para cada acción individual la función de pérdida 0-1 generalizada. Además, en esta sección, se propone un criterio de decisión bayesiano basado en la estimación del número de hipótesis nulas falsas.

En la sección 2.3, se aplica la metodología descrita en las secciones anteriores, considerando las medias como parámetros de interés y asumiendo, para las observaciones, distribución normal con varianza conocida; asimismo, se ilustra el procedimiento mediante un ejemplo de simulación, realizando también un análisis de sensibilidad a la elección de los parámetros de la distribución inicial; finalmente, se comparan los resultados con los obtenidos mediante el procedimiento de Benjamini y Hochberg (1995) para controlar el FDR.

En la sección 2.4 se considera, para las observaciones, distribución normal con varianza desconocida siendo las medias, al igual que en la sección 2.3, los parámetros de interés. En esta sección se describe brevemente cómo hacer inferencia utilizando los métodos MCMC; posteriormente se aplica uno de estos métodos, concretamente el muestreo de Gibbs; también se ilustra la metodología mediante un ejemplo de simulación, realizando igualmente un análisis de sensibilidad a la elección de los parámetros de la distribución inicial; por último, se comparan los resultados con los obtenidos mediante el procedimiento de Benjamini y Hochberg (1995) para controlar el FDR. Todos los cálculos de este capítulo se han llevado a cabo utilizando

MATLAB. Para finalizar, en la sección 2.5, se exponen las conclusiones del capítulo.

2.1 Formulación del problema

En esta sección se aborda el problema de contrastar simultáneamente N hipótesis simples e idénticas para todas las hipótesis. Problema tratado también, aunque con un planteamiento diferente, por Storey y Tibshirani (2001), Storey (2003) para el modelo normal y De la Horra (2007), entre otros autores.

Consideramos el siguiente problema de contraste de hipótesis múltiples,

$$H_{0i} : \theta_i = \theta_0 \quad \text{frente a} \quad H_{1i} : \theta_i = \theta_1 \quad (\theta_0 \neq \theta_1), \quad i = 1, \dots, N,$$

donde θ_0 y θ_1 son valores conocidos y comunes para todas las hipótesis.

Desde una perspectiva bayesiana el objetivo es obtener la probabilidad final de cada hipótesis nula y decidir, en función de la misma, qué hipótesis nulas se aceptan y cuáles se rechazan.

Sea $T_i = (X_{i1}, \dots, X_{in})$ el vector de observaciones para cada hipótesis y suponemos que las T_i , para $i = 1, \dots, N$, son variables aleatorias independientes e idénticamente distribuidas con densidad $f_0(t_i)$ bajo H_{0i} y $f_1(t_i)$ bajo H_{1i} , por lo que resulta natural considerar las observaciones T_i procedentes de una mezcla de ambas densidades, véase Efron et al. (2001), Storey y Tibshirani (2001), Storey (2003), Kendzioriski et al. (2003), Efron (2004) y Do et al. (2005), entre otros.

Suponemos igualmente que hay una probabilidad inicial común p de que $\theta_i = \theta_0$, para todo i . Entonces también podemos considerar que p es la proporción desconocida de hipótesis nulas ciertas. Así, denotando por $H_{0i} = 0$ si H_{0i} es cierta y por

$H_{0i} = 1$ si H_{0i} es falsa, la distribución para las observaciones es la siguiente mixtura de las distribuciones bajo las hipótesis nula y alternativa,

$$\begin{aligned} f(t_i|p) &= Pr(H_{0i} = 0|p)f(t_i|H_{0i} = 0) + Pr(H_{0i} = 1|p)f(t_i|H_{0i} = 1) \\ &= pf_0(t_i) + (1-p)f_1(t_i), \end{aligned}$$

donde $p = Pr(H_{0i} = 0|p)$ y $t_i = (x_{i1}, \dots, x_{in})$.

La verosimilitud se puede escribir como

$$f(\mathbf{t}|p) = \prod_{i=1}^N f(t_i|p) = \prod_{i=1}^N [pf_0(t_i) + (1-p)f_1(t_i)],$$

siendo $\mathbf{t} = (t_1, \dots, t_N)$.

Con el fin de realizar la inferencia bayesiana es necesario definir una distribución inicial para el parámetro p del modelo. Se puede considerar una distribución $Beta(\alpha, \beta)$ por su versatilidad en el intervalo $[0, 1]$. Entonces, la probabilidad final de cada hipótesis nula tiene la siguiente expresión,

$$\begin{aligned} Pr(H_{0i} = 0|\mathbf{t}, \alpha, \beta) &= \frac{\int_0^1 pf_0(t_i) \prod_{j \neq i} [pf_0(t_j) + (1-p)f_1(t_j)] \pi(p|\alpha, \beta) dp}{\int_0^1 f(\mathbf{t}|p) \pi(p|\alpha, \beta) dp} \\ &= \frac{\int_0^1 g(t_i|p) f(\mathbf{t}|p) \pi(p|\alpha, \beta) dp}{\int_0^1 f(\mathbf{t}|p) \pi(p|\alpha, \beta) dp} \tag{2.1} \\ &= E_{\pi(p|\mathbf{t}, \alpha, \beta)} [g(t_i|p)], \end{aligned}$$

donde $g(t_i|p) = \frac{pf_0(t_i)}{pf_0(t_i) + (1-p)f_1(t_i)}$.

La probabilidad final de cada hipótesis nula, $Pr(H_{0i} = 0|\mathbf{t}, \alpha, \beta)$, dada por (2.1), no siempre se puede obtener de forma analítica, ni tampoco se puede aproximar

siempre mediante integración numérica por el método de Monte Carlo, ya que la distribución final de p , $\pi(p|\mathbf{t}, \alpha, \beta)$, puede no ser conocida. Una posibilidad es utilizar los métodos de cadenas de Markov Monte Carlo (MCMC), sin embargo es posible aproximar la probabilidad final de cada hipótesis nula de una forma más sencilla, ya que la expresión (2.1) se puede escribir como un cociente de esperanzas con respecto a la distribución inicial de p de la forma siguiente,

$$Pr(H_{0i} = 0|\mathbf{t}, \alpha, \beta) = \frac{E_{\pi(p|\alpha, \beta)} [g(t_i|p)f(\mathbf{t}|p)]}{E_{\pi(p|\alpha, \beta)} [f(\mathbf{t}|p)]}.$$

Por tanto, como la distribución inicial de p es conocida, la probabilidad final de cada hipótesis nula se puede aproximar aplicando el método de Monte Carlo al numerador y al denominador simulando una muestra directamente de la distribución inicial.

De manera que, si se simula una muestra aleatoria, p_1, \dots, p_M , de la distribución inicial $Beta(\alpha, \beta)$, la probabilidad final de cada hipótesis nula se puede aproximar como se indica a continuación,

$$\hat{Pr}(H_{0i} = 0|\mathbf{t}, \alpha, \beta) = \frac{\sum_{j=1}^M g(t_i|p_j)f(\mathbf{t}|p_j)}{\sum_{j=1}^M f(\mathbf{t}|p_j)}. \quad (2.2)$$

Una vez estimada la probabilidad final de cada hipótesis nula, $\hat{Pr}(H_{0i} = 0|\mathbf{t}, \alpha, \beta)$, mediante (2.2), es preciso decidir qué hipótesis nulas se rechazan y qué hipótesis nulas se aceptan. Es decir, hay que elegir un valor P de corte de manera que todas las hipótesis nulas, H_{0i} , tales que $\hat{Pr}(H_{0i} = 0|\mathbf{t}, \alpha, \beta) \leq P$ se rechazarán, aceptando el resto. El problema ahora es la elección de ese valor crítico P . En la sección siguiente se trata esta cuestión.

2.2 Criterios de decisión bayesiana

En esta sección consideramos dos criterios de decisión desde el punto de vista de la teoría de la decisión bayesiana. El primero es un criterio ya clásico, véase Duncan (1965) y Lewis y Thayer (2004), en el que se muestra la regla bayes cuando se considera, para el problema general de contraste de hipótesis múltiples, una función de pérdida aditiva y, para cada contraste individual, la función de pérdida 0-1 generalizada. El segundo es el que proponemos en este trabajo basado en la estimación del número de hipótesis nulas falsas.

- *Regla bayes*

Desde una perspectiva de la teoría de la decisión bayesiana, se trata de elegir la acción con menor pérdida esperada final.

En primer lugar definimos el espacio paramétrico y el espacio de acciones que, utilizando la misma notación que en el subsección 1.4.3, se pueden representar como

$$\Theta = \bigcup_{j=1}^{2^N} \Theta_j \quad \text{donde} \quad \Theta_j = \bigcap_{i=1}^N \{H_{0i} = \varepsilon_{ij}\}, \quad j = 1, \dots, 2^N,$$

siendo $\varepsilon_{ij} = 0$ si H_{0i} es cierta y $\varepsilon_{ij} = 1$ si H_{0i} es falsa.

$$A = \bigcup_{j=1}^{2^N} A_j \quad \text{donde} \quad A_j = \bigcap_{i=1}^N a_{e_{ij}i}, \quad j = 1, \dots, 2^N,$$

y $e_{ij} = 0$ si se acepta H_{0i} y $e_{ij} = 1$ si se rechaza H_{0i} ; es decir, $a_{e_{ij}i}$ es la acción individual que se toma con respecto a la hipótesis H_{0i} dentro de la acción conjunta A_j .

Podemos considerar una función de pérdida aditiva, de manera que el problema de minimizar la pérdida esperada final se puede resolver minimizando cada una de las componentes. Por tanto, la función de pérdida, cuando se toma la acción

conjunta A_j y el verdadero valor del parámetro es Θ_l , se puede escribir como

$$L(\Theta_l, A_j) = \sum_{i=1}^N L_i(H_{0i} = \varepsilon_{il}, a_{e_{ij}i}),$$

donde

$$L_i(H_{0i} = \varepsilon_{il}, a_{e_{ij}i}) = \begin{cases} 0 & \text{si } e_{ij} = \varepsilon_{il} \\ C_{e_{ij}i} & \text{si } e_{ij} \neq \varepsilon_{il} \end{cases}$$

es la función de pérdida 0-1 generalizada y representa el coste individual cuando, para la hipótesis H_{0i} , se toma la acción $a_{e_{ij}i}$ siendo $H_{0i} = \varepsilon_{il}$.

Las pérdidas esperadas finales individuales de a_{0i} y a_{1i} son, respectivamente, $C_{0i} Pr(H_{0i} = 1|\mathbf{t})$ y $C_{1i} Pr(H_{0i} = 0|\mathbf{t})$. Entonces, para cada hipótesis H_{0i} , la acción a_{1i} es preferible a la acción a_{0i} si

$$C_{1i} Pr(H_{0i} = 0|\mathbf{t}) \leq C_{0i} Pr(H_{0i} = 1|\mathbf{t}),$$

de donde se deduce la regla bayes: para cada \mathbf{t} , rechazar todas las hipótesis nulas H_{0i} tales que

$$Pr(H_{0i} = 0|\mathbf{t}) \leq \frac{C_{0i}}{C_{0i} + C_{1i}} \quad (2.3)$$

y aceptar el resto, dados los costes C_{0i} y C_{1i} , cuya especificación puede no ser un problema sencillo.

- *Criterio basado en la estimación del número de hipótesis nulas falsas (EHNF)*

Proponemos un criterio de decisión basado en la estimación del número de hipótesis nulas falsas, que denotamos por N_1 como en la tabla 1.2, y que consistiría en rechazar las \widehat{N}_1 hipótesis nulas con menor probabilidad final, siendo \widehat{N}_1 un estimador de N_1 . El objetivo ahora es obtener un estimador bayesiano de N_1 .

Si $p = Pr(H_{0i} = 0|p)$ y por tanto $1 - p = Pr(H_{0i} = 1|p)$, $i = 1, \dots, N$, y suponemos que las N hipótesis son independientes, entonces $H_{0i}|p \sim \text{Bernoulli}(1 - p)$

y como $N_1 = \sum_{i=1}^N H_{0i}$, se tiene que $N_1|p \sim \text{Binomial}(N, 1-p)$ y si bien p es un parámetro desconocido, podemos estimarlo utilizando la distribución final de p , $\pi(p|\mathbf{t}, \alpha, \beta)$, por ejemplo con la media de dicha distribución. Entonces,

$$\begin{aligned} \hat{p} &= E_{\pi(p|\mathbf{t}, \alpha, \beta)}[p] \\ &= \frac{\int_0^1 p f(\mathbf{t}|p) \pi(p|\alpha, \beta) dp}{\int_0^1 f(\mathbf{t}|p) \pi(p|\alpha, \beta) dp} = \frac{E_{\pi(p|\alpha, \beta)} [p f(\mathbf{t}|p)]}{E_{\pi(p|\alpha, \beta)} [f(\mathbf{t}|p)]}, \end{aligned} \quad (2.4)$$

es decir, \hat{p} , al igual que la probabilidad final de cada hipótesis nula, se puede expresar como un cociente de esperanzas con respecto a la distribución inicial de p , por lo que se puede aproximar, de manera similar a (2.2), por el método de Monte Carlo de la forma siguiente,

$$\hat{p} \approx \frac{\sum_{i=1}^M p_i f(\mathbf{t}|p_i)}{\sum_{i=1}^M f(\mathbf{t}|p_i)}, \quad (2.5)$$

donde p_1, \dots, p_M es una muestra aleatoria de la distribución inicial $\text{Beta}(\alpha, \beta)$ para p .

Así, $N_1|\hat{p} \sim \text{Binomial}(N, 1-\hat{p})$ y podemos estimar N_1 mediante $\widehat{N}_1 = E[N_1|\hat{p}] = N(1 - \hat{p})$.

De esta manera, se considera como criterio de decisión rechazar las \widehat{N}_1 hipótesis nulas con menor probabilidad final y aceptar el resto, salvando así el problema de la elección de las constantes C_{0i} y C_{1i} de la regla bayes definida en (2.3).

2.3 Modelo Gaussiano. Varianza conocida

Consideramos el siguiente problema de contrastar simultáneamente N hipótesis simples,

$$H_{0i} : \mu_i = \mu_0 \quad \text{frente a} \quad H_{1i} : \mu_i = \mu_1 \quad (\mu_0 \neq \mu_1), \quad i = 1, \dots, N,$$

donde μ_0 y μ_1 son valores conocidos y comunes para todas las hipótesis.

Sea $T_i = (X_{i1}, \dots, X_{in})$ el vector de observaciones para cada hipótesis y suponemos que las T_i , para $i = 1, \dots, N$, son variables aleatorias independientes e idénticamente distribuidas $N(\mu_0, 1/\sqrt{\phi})$ bajo H_{0i} y $N(\mu_1, 1/\sqrt{\phi})$ bajo H_{1i} , siendo ϕ la precisión que suponemos conocida y la misma, por simplicidad, para todas las hipótesis. Entonces, podemos considerar las observaciones T_i procedentes de una mezcla de ambas densidades,

$$\begin{aligned} f(t_i|p) &= pf_0(t_i) + (1-p)f_1(t_i) \\ &= p \prod_{j=1}^n f_0(x_{ij}) + (1-p) \prod_{j=1}^n f_1(x_{ij}), \end{aligned}$$

donde $t_i = (x_{i1}, \dots, x_{in})$, p es la probabilidad inicial de la hipótesis nula y $f_0(t_i)$ y $f_1(t_i)$ son las densidades bajo la hipótesis nula y bajo la hipótesis alternativa, respectivamente, para $i = 1, \dots, N$. Así, podemos expresar $f(t_i|p)$ de la siguiente forma,

$$f(t_i|p) = k_i(p u_{0i} + (1-p)u_{1i}), \tag{2.6}$$

donde

$$\begin{aligned} k_i &= (2\pi)^{-n/2} \phi^{n/2} e^{-(\phi/2) \sum_{j=1}^n x_{ij}^2}, \\ u_{0i} &= e^{-(\phi n/2)(\mu_0^2 - 2\mu_0 \bar{x}_i)} \quad \text{y} \\ u_{1i} &= e^{-(\phi n/2)(\mu_1^2 - 2\mu_1 \bar{x}_i)}. \end{aligned} \tag{2.7}$$

La verosimilitud se puede escribir como

$$f(\mathbf{t}|p) = \prod_{i=1}^N f(t_i|p) = \prod_{i=1}^N k_i(p u_{0i} + (1-p)u_{1i}). \quad (2.8)$$

Entonces, considerando para la distribución inicial del parámetro p una distribución $Beta(\alpha, \beta)$, podemos aproximar por el método de Monte Carlo, como en (2.2), la probabilidad final de cada hipótesis nula mediante la expresión

$$\hat{Pr}(H_{0i} = 0|\mathbf{t}, \alpha, \beta) = \frac{\sum_{k=1}^M \left(1 + \frac{1-p_k}{p_k} \frac{u_{1i}}{u_{0i}}\right)^{-1} \prod_{j=1}^N (p_k u_{0j} + (1-p_k)u_{1j})}{\sum_{k=1}^M \prod_{j=1}^N (p_k u_{0j} + (1-p_k)u_{1j})}, \quad (2.9)$$

siendo p_1, \dots, p_M una muestra aleatoria de la distribución inicial $Beta(\alpha, \beta)$.

De forma análoga se obtiene una aproximación para \hat{p} mediante el método de Monte Carlo, como en (2.5),

$$\hat{p} \approx \frac{\sum_{k=1}^M p_k \prod_{i=1}^N (p_k u_{0i} + (1-p_k)u_{1i})}{\sum_{k=1}^M \prod_{i=1}^N (p_k u_{0i} + (1-p_k)u_{1i})}, \quad (2.10)$$

donde p_1, \dots, p_M es una muestra aleatoria de la distribución inicial $Beta(\alpha, \beta)$.

Una vez se han obtenido, para cada i , una estimación de la probabilidad final de cada hipótesis nula, $\hat{Pr}(H_{0i} = 0|\mathbf{t}, \alpha, \beta)$, y una estimación de p , \hat{p} , dadas en (2.9) y (2.10), respectivamente, podemos aplicar los criterios de decisión descritos en la sección anterior. Así, se rechazarían todas las hipótesis nulas H_{0i} tales que $\hat{Pr}(H_{0i} = 0|\mathbf{t}, \alpha, \beta) \leq C_{0i}/(C_{0i} + C_{1i})$, aceptando el resto, si se aplica la regla bayes (2.3), o se rechazarían las $\widehat{N}_1 = N(1 - \hat{p})$ hipótesis nulas con menor probabilidad final estimada, aceptando el resto, si se aplica el criterio EHNF.

2.3.1 Ejemplo

Consideramos el siguiente problema de contrastes múltiples,

$$H_{0i} : \mu_i = 0 \quad \text{frente a} \quad H_{1i} : \mu_i = 1, \quad i = 1, \dots, N.$$

Sea $T_i = (X_{i1}, \dots, X_{in})$ el vector de observaciones para cada hipótesis y supongamos que las T_i , para $i = 1, \dots, N$, son variables aleatorias independientes con distribución $N(0, 1)$ bajo H_{0i} y $N(1, 1)$ bajo H_{1i} .

Entonces, considerando para la distribución inicial del parámetro p una distribución $Beta(\alpha, \beta)$ podemos aproximar, mediante el método de Monte Carlo, la probabilidad final de cada hipótesis nula y la probabilidad inicial p sustituyendo en (2.9) y (2.10), respectivamente, las expresiones definidas en (2.7) para $\mu_0 = 0$, $\mu_1 = 1$ y $\phi = 1$, de la forma siguiente,

$$\hat{Pr}(H_{0i} = 0 | \mathbf{t}, \alpha, \beta) = \frac{\sum_{k=1}^M \left(1 + \frac{1-p_k}{p_k} e^{n(\bar{x}_i-1/2)} \right)^{-1} \prod_{j=1}^N (p_k + (1-p_k)e^{n(\bar{x}_j-1/2)})}{\sum_{k=1}^M \prod_{j=1}^N (p_k + (1-p_k)e^{n(\bar{x}_j-1/2)}} \quad (2.11)$$

$$\hat{p} \approx \frac{\sum_{k=1}^M p_k \prod_{i=1}^N (p_k + (1-p_k)e^{n(\bar{x}_i-1/2)})}{\sum_{k=1}^M \prod_{i=1}^N (p_k + (1-p_k)e^{n(\bar{x}_i-1/2)}), \quad (2.12)$$

donde p_1, \dots, p_M es una muestra aleatoria de la distribución inicial $Beta(\alpha, \beta)$.

2.3.2 Simulación: resultados y análisis de sensibilidad

En esta subsección, basándonos en el ejemplo anterior, desarrollamos un ejemplo de simulación para estudiar el comportamiento de nuestra aproximación.

Se realizaron dos simulaciones. En la primera se simularon cuatro conjuntos de datos con $N = 500, 1000, 5000$ y 10000 hipótesis. Cada conjunto de datos se generó de una mezcla de $N(0, 1)$ con probabilidad $p = 0.9$ y de $N(1, 1)$ con probabilidad $1 - p = 0.1$, con $n = 5$ observaciones para cada hipótesis. En la segunda se simularon otros cuatro conjuntos de datos, con el mismo número de hipótesis y el mismo número de observaciones por hipótesis que en la primera simulación, pero los datos se generaron de una mezcla de $N(0, 1)$ con probabilidad $p = 0.7$ y de $N(1, 1)$ con probabilidad $1 - p = 0.3$.

Para la distribución inicial del parámetro p se ha considerado una distribución $Beta(\alpha, \beta)$. Los valores utilizados para los parámetros (α, β) fueron: $(1, 0.5)$, $(1, 1)$, $(0.5, 1)$, $(0.5, 0.5)$, $(1.5, 1.5)$ y $(1, 25)$. Se eligieron estos valores para los parámetros con el fin de realizar un análisis de sensibilidad, ya que las distribuciones correspondientes son muy diferentes. La distribución correspondiente a los parámetros $(1, 0.5)$ da más densidad para los valores de p próximos a uno; mientras que la distribución correspondiente a $(0.5, 1)$ da más densidad para los valores de p próximos a cero, igual que para $(1, 25)$, si bien esta última presenta un sesgo mucho más fuerte; las distribuciones correspondientes a $(0.5, 0.5)$ y a $(1.5, 1.5)$ son simétricas respecto a $p = 0.5$, siendo una la inversa de la otra y la distribución correspondiente a $(1, 1)$ es la uniforme en el intervalo $[0, 1]$.

Con el fin de realizar un análisis de sensibilidad para cada conjunto de datos simulados y para los distintos valores considerados de los parámetros (α, β) , aplicamos

Tabla 2.1: Estimación final de p para distintos valores de los parámetros de la distribución inicial de $p \sim Beta(\alpha, \beta)$, utilizando los datos generados mediante una mixtura de $N(0, 1)$ con probabilidad $p = 0.9$ y de $N(1, 1)$ con probabilidad $1 - p = 0.1$.

(α, β)	(1, 0.5)	(1, 1)	(0.5, 1)	(0.5, 0.5)	(1.5, 1.5)	(1, 25)
$N = 500$	0.8909	0.8900	0.8900	0.8916	0.8879	0.3030
$N = 1000$	0.8791	0.8769	0.8776	0.8789	0.8760	0.3437
$N = 5000$	0.8978	0.8980	0.8986	0.8984	0.8974	0.3636
$N = 10000$	0.9059	0.9056	0.9058	0.9057	0.9053	0.2882

el método de Monte Carlo para aproximar la probabilidad final de cada hipótesis nula y estimar p , como en (2.11) y (2.12) respectivamente. Este procedimiento es muy rápido computacionalmente, empleando entre dos y treinta y siete segundos, aproximadamente, dependiendo del número de hipótesis y simulando muestras de la distribución $Beta(\alpha, \beta)$ de tamaño 10000.

En la tabla 2.1 se muestran las estimaciones de p obtenidas para los distintos valores de los parámetros, utilizando los datos generados mediante la mixtura de

Tabla 2.2: Estimación final de p para distintos valores de los parámetros de la distribución inicial de $p \sim Beta(\alpha, \beta)$, utilizando los datos generados mediante una mixtura de $N(0, 1)$ con probabilidad $p = 0.7$ y de $N(1, 1)$ con probabilidad $1 - p = 0.3$.

(α, β)	(1, 0.5)	(1, 1)	(0.5, 1)	(0.5, 0.5)	(1.5, 1.5)	(1, 25)
$N = 500$	0.6924	0.6923	0.6910	0.6928	0.6912	0.3340
$N = 1000$	0.6985	0.6979	0.6978	0.6983	0.6975	0.3186
$N = 5000$	0.6768	0.6769	0.6772	0.6770	0.6762	0.3726
$N = 10000$	0.7018	0.7018	0.7011	0.7021	0.7020	0.3354

$N(0, 1)$ con probabilidad $p = 0.9$ y de $N(1, 1)$ con probabilidad $1 - p = 0.1$. Mientras que en la tabla 2.2 se muestran las estimaciones de p para los datos generados con la mixtura de $N(0, 1)$ con probabilidad $p = 0.7$ y de $N(1, 1)$ con probabilidad $1 - p = 0.3$.

Como puede observarse en ambos casos, el procedimiento es robusto respecto a las distribuciones consideradas, excepto para la distribución $Beta(1, 25)$, ya que no se observan cambios importantes en las estimaciones, a pesar de que las distribuciones iniciales correspondientes son muy diferentes. Por ejemplo, la distribución $Beta(1, 0.5)$ debería ser, entre las distribuciones iniciales consideradas, la más adecuada puesto que da más densidad para los valores de p próximos a uno. No obstante se obtienen resultados similares con la distribución $Beta(0.5, 1)$ que da más densidad para los valores próximos a cero. Por otro lado hay que destacar que las estimaciones que se obtienen están muy próximas a las probabilidades de las mixturas con las que se generaron los datos ($p = 0.9$ y $p = 0.7$), a excepción de las obtenidas con la distribución inicial $Beta(1, 25)$ como ya se ha comentado.

Para decidir qué hipótesis nulas se rechazan aplicamos, en primer lugar, la regla bayes definida en (2.3) considerando $C_{0i} = C_{1i}$ para $i = 1, \dots, N$, por lo que se rechazarán todas las hipótesis nulas tales que $\hat{P}r(H_{0i} = 0 | \mathbf{t}, \alpha, \beta) \leq 0.5$.

Utilizando los datos generados mediante la mixtura de $N(0, 1)$ con probabilidad $p = 0.9$ y de $N(1, 1)$ con probabilidad $1 - p = 0.1$, y para los distintos valores de los parámetros de la distribución inicial de p , en la tabla 2.3 se muestran los resultados obtenidos aplicando la regla bayes; es decir, el porcentaje de hipótesis nulas rechazadas ($R_C\%$), y los porcentajes de falsos positivos ($FP_C\%$) y falsos negativos ($FN_C\%$) calculados de la forma siguiente,

Tabla 2.3: Porcentaje de hipótesis nulas rechazadas ($R_C\%$) y porcentaje de falsos positivos ($FP_C\%$) y falsos negativos ($FN_C\%$), aplicando la regla bayes, para distintos valores de los parámetros de la distribución inicial de $p \sim Beta(\alpha, \beta)$, utilizando los datos generados mediante una mixtura de $N(0, 1)$ con probabilidad $p = 0.9$ y de $N(1, 1)$ con probabilidad $1 - p = 0.1$.

	(α, β)	(1, 0.5)	(1, 1)	(0.5, 1)	(0.5, 0.5)	(1.5, 1.5)	(1, 25)
$N = 500$	$R_C\%$	7.20	7.20	7.20	7.20	7.20	30.20
	$FP_C\%$	2.4176	2.4176	2.4176	2.4176	2.4176	24.1758
	$FN_C\%$	44.4444	44.4444	44.4444	44.4444	44.4444	8.8889
$N = 1000$	$R_C\%$	9.50	9.50	9.50	9.50	9.60	28.10
	$FP_C\%$	2.3516	2.3516	2.3516	2.3516	2.3516	20.4927
	$FN_C\%$	30.8411	30.8411	30.8411	30.8411	29.9065	8.4112
$N = 5000$	$R_C\%$	7.04	7.02	6.94	6.96	7.08	27.70
	$FP_C\%$	1.8961	1.8737	1.8514	1.8514	1.9184	20.0981
	$FN_C\%$	48.3559	48.3559	48.9362	48.7427	48.1625	6.3830
$N = 10000$	$R_C\%$	6.52	6.55	6.52	6.54	6.55	30.71
	$FP_C\%$	1.5459	1.5570	1.5459	1.5570	1.5570	24.0283
	$FN_C\%$	45.7627	45.5508	45.7627	45.6568	45.5508	5.1907

$$\begin{aligned}
 FP_C\% &= \frac{\sum_{i=1}^N I(\hat{Pr}(H_{0i} = 0 | \mathbf{t}, \alpha, \beta) \leq 0.5) I(\mu_i = 0)}{\sum_{i=1}^N I(\mu_i = 0)} \times 100, \\
 FN_C\% &= \frac{\sum_{i=1}^N I(\hat{Pr}(H_{0i} = 0 | \mathbf{t}, \alpha, \beta) > 0.5) I(\mu_i = 1)}{\sum_{i=1}^N I(\mu_i = 1)} \times 100. \quad (2.13)
 \end{aligned}$$

De igual forma, en la tabla 2.4 se presentan los mismos resultados que en la tabla 2.3 pero para los datos generados mediante la mixtura de $N(0, 1)$ con probabi-

Tabla 2.4: Porcentaje de hipótesis nulas rechazadas ($R_C\%$) y porcentaje de falsos positivos ($FP_C\%$) y falsos negativos ($FN_C\%$), aplicando la regla bayes, para distintos valores de los parámetros de la distribución inicial de $p \sim \text{Beta}(\alpha, \beta)$, utilizando los datos generados mediante una mixtura de $N(0, 1)$ con probabilidad $p = 0.7$ y de $N(1, 1)$ con probabilidad $1 - p = 0.3$.

	(α, β)	(1, 0.5)	(1, 1)	(0.5, 1)	(0.5, 0.5)	(1.5, 1.5)	(1, 25)
$N = 500$	$R_C\%$	27.60	27.60	27.80	27.60	27.80	45.60
	$FP_C\%$	5.4598	5.4598	5.4598	5.4598	5.4598	24.4253
	$FN_C\%$	21.7105	21.7105	21.0526	21.7105	21.0526	5.9211
$N = 1000$	$R_C\%$	28.40	28.40	28.40	28.40	28.50	41.60
	$FP_C\%$	6.7047	6.7047	6.7047	6.7047	6.7047	18.9729
	$FN_C\%$	20.7358	20.7358	20.7358	20.7358	20.4013	5.3512
$N = 5000$	$R_C\%$	30.48	30.48	30.48	30.48	30.52	41.76
	$FP_C\%$	8.1574	8.1574	8.1574	8.1574	8.1574	19.6124
	$FN_C\%$	19.5075	19.5075	19.5075	19.5075	19.3778	8.6196
$N = 10000$	$R_C\%$	27.68	27.68	27.69	27.68	27.68	42.28
	$FP_C\%$	6.4571	6.4571	6.4571	6.4571	6.4571	20.4571
	$FN_C\%$	22.8000	22.8000	22.7667	22.8000	22.8000	6.8000

lidad $p = 0.7$ y de $N(1, 1)$ con probabilidad $1 - p = 0.3$. Se puede observar que el procedimiento es robusto respecto a los parámetros considerados excepto para $(\alpha, \beta) = (1, 25)$.

Por último, aplicamos el criterio EHNF rechazando las $\widehat{N}_1 = N(1 - \hat{p})$ hipótesis nulas con menor probabilidad final estimada. Utilizando los datos generados mediante la mixtura de $N(0, 1)$ con probabilidad $p = 0.9$ y de $N(1, 1)$ con probabilidad $1 - p = 0.1$, y para los distintos valores de los parámetros de la distribución inicial de p , en la tabla 2.5 se presentan los resultados obtenidos aplicando el criterio de

decisión EHNF, es decir, el porcentaje estimado de hipótesis nulas falsas ($\widehat{N}_1\%$), la probabilidad final estimada más alta con la que se rechaza cada hipótesis nula: $prob_{\widehat{N}_1} = \hat{P}r(H_{(0\widehat{N}_1)} = 0|\mathbf{t}, \alpha, \beta)$, siendo $\hat{P}r(H_{(0i)} = 0|\mathbf{t}, \alpha, \beta)$ las probabilidades finales ordenadas, y los porcentajes de falsos positivos ($FP_{EHNF}\%$) y falsos negativos ($FN_{EHNF}\%$) calculados como se indica a continuación,

$$FP_{EHNF}\% = \frac{\sum_{i=1}^N I(\hat{P}r(H_{0i} = 0|\mathbf{t}, \alpha, \beta) \leq prob_{\widehat{N}_1}) I(\mu_i = 0)}{\sum_{i=1}^N I(\mu_i = 0)} \times 100,$$

$$FN_{EHNF}\% = \frac{\sum_{i=1}^N I(\hat{P}r(H_{0i} = 0|\mathbf{t}, \alpha, \beta) > prob_{\widehat{N}_1}) I(\mu_i = 1)}{\sum_{i=1}^N I(\mu_i = 1)} \times 100. \quad (2.14)$$

En la tabla 2.6 se muestran los mismos resultados que en la tabla 2.5 pero para los datos generados mediante la mezcla de $N(0, 1)$ con probabilidad $p = 0.7$ y de $N(1, 1)$ con probabilidad $1 - p = 0.3$. Se puede ver que también el procedimiento es robusto respecto a los parámetros considerados excepto para $(\alpha, \beta) = (1, 25)$, estando el porcentaje estimado de hipótesis nulas falsas muy próximo a los porcentajes de hipótesis nulas falsas con los que se generaron los datos en ambas mezclas (10% y 30%).

Asimismo, cabe destacar que el procedimiento seguido aplicando la regla bayes es ligeramente más conservador que el que resulta aplicando el criterio de decisión EHNF, en el sentido de que con el primero se rechazan menos hipótesis nulas. Las hipótesis nulas adicionales que se rechazan con el criterio EHNF tienen una probabilidad final estimada de ser ciertas mayor que 0.5, debido a que con la regla bayes se rechazan las hipótesis nulas con probabilidad final estimada de ser ciertas menor o igual que 0.5, lo que implica que, con el criterio EHNF, se rechazan más hipótesis

Tabla 2.5: Porcentaje estimado de hipótesis nulas falsas ($\widehat{N}_1\%$), probabilidad final estimada más alta con la que se rechaza cada hipótesis nula ($prob_{\widehat{N}_1}$) y porcentaje de falsos positivos ($FP_{EHNF}\%$) y falsos negativos ($FN_{EHNF}\%$), aplicando el criterio EHNF, para distintos valores de los parámetros de la distribución inicial de $p \sim Beta(\alpha, \beta)$, utilizando los datos generados mediante una mezcla de $N(0, 1)$ con probabilidad $p = 0.9$ y de $N(1, 1)$ con probabilidad $1 - p = 0.1$.

	(α, β)	(1, 0.5)	(1, 1)	(0.5, 1)	(0.5, 0.5)	(1.5, 1.5)	(1, 25)
$N = 500$	$\widehat{N}_1\%$	11	11	11	10.80	11.20	69.80
	$prob_{\widehat{N}_1}$	0.6117	0.6092	0.6092	0.6124	0.6044	0.9329
	$FP_{EHNF}\%$	5.0549	5.0549	5.0549	4.8352	5.2747	67.0330
	$FN_{EHNF}\%$	28.8889	28.8889	28.8889	28.8889	28.8889	2.2222
$N = 1000$	$\widehat{N}_1\%$	12.10	12.30	12.20	12.10	12.40	65.60
	$prob_{\widehat{N}_1}$	0.6682	0.6643	0.6654	0.6679	0.6674	0.9218
	$FP_{EHNF}\%$	4.7032	4.9272	4.8152	4.7032	5.0392	61.4782
	$FN_{EHNF}\%$	26.1682	26.1682	26.1682	26.1682	26.1682	0
$N = 5000$	$\widehat{N}_1\%$	10.22	10.20	10.14	10.16	10.26	63.64
	$prob_{\widehat{N}_1}$	0.6683	0.6665	0.6651	0.6649	0.6679	0.9190
	$FP_{EHNF}\%$	3.8144	3.7921	3.7475	3.7475	3.8590	59.5806
	$FN_{EHNF}\%$	34.2360	34.2360	34.4294	34.2360	34.2360	1.1605
$N = 10000$	$\widehat{N}_1\%$	9.41	9.44	9.42	9.43	9.47	71.18
	$prob_{\widehat{N}_1}$	0.6719	0.6721	0.6722	0.6722	0.6735	0.9350
	$FP_{EHNF}\%$	3.3679	3.39	3.3679	3.3790	3.4121	68.2089
	$FN_{EHNF}\%$	32.6271	32.5212	32.5212	32.5212	32.4153	0.3178

nulas ciertas, aumentando de este modo el porcentaje de falsos positivos, si bien se mantiene en unos niveles aceptables ($< 10\%$).

Además, con el criterio EHNF se rechazan algunas de las hipótesis nulas falsas que no se detectan utilizando la regla bayes, por tener una probabilidad final estimada de ser ciertas mayor que 0.5, obteniéndose así un porcentaje de falsos ne-

Tabla 2.6: Porcentaje estimado de hipótesis nulas falsas ($\widehat{N}_1\%$), probabilidad final estimada más alta con la que se rechaza cada hipótesis nula ($prob_{\widehat{N}_1}$) y porcentaje de falsos positivos ($FP_{EHNF}\%$) y falsos negativos ($FN_{EHNF}\%$), aplicando el criterio EHNF, para distintos valores de los parámetros de la distribución inicial de $p \sim Beta(\alpha, \beta)$, utilizando los datos generados mediante una mixtura de $N(0, 1)$ con probabilidad $p = 0.7$ y de $N(1, 1)$ con probabilidad $1 - p = 0.3$.

	(α, β)	(1, 0.5)	(1, 1)	(0.5, 1)	(0.5, 0.5)	(1.5, 1.5)	(1, 25)
$N = 500$	$\widehat{N}_1\%$	30.80	30.80	31	30.80	30.80	66.60
	$prob_{\widehat{N}_1}$	0.5640	0.5639	0.5815	0.5646	0.5627	0.8896
	$FP_{EHNF}\%$	8.3333	8.3333	8.6207	8.3333	8.3333	52.5862
	$FN_{EHNF}\%$	17.7632	17.7632	17.7632	17.7632	17.7632	1.3158
$N = 1000$	$\widehat{N}_1\%$	30.20	30.20	30.20	30.20	30.30	68.10
	$prob_{\widehat{N}_1}$	0.5471	0.5463	0.5463	0.5469	0.5483	0.8978
	$FP_{EHNF}\%$	7.8459	7.8459	7.8459	7.8459	7.8459	54.7789
	$FN_{EHNF}\%$	17.3913	17.3913	17.3913	17.3913	17.0569	0.6689
$N = 5000$	$\widehat{N}_1\%$	32.32	32.30	32.28	32.30	32.38	62.74
	$prob_{\widehat{N}_1}$	0.5545	0.5545	0.5548	0.5546	0.5556	0.8500
	$FP_{EHNF}\%$	9.8640	9.8351	9.8062	9.8351	9.9219	46.7747
	$FN_{EHNF}\%$	17.3688	17.3688	17.3688	17.3688	17.3040	1.4906
$N = 10000$	$\widehat{N}_1\%$	29.82	29.82	29.89	29.79	29.80	66.46
	$prob_{\widehat{N}_1}$	0.5686	0.5685	0.5699	0.5672	0.5673	0.8764
	$FP_{EHNF}\%$	7.9286	7.9286	7.9857	7.9143	7.9286	52.6571
	$FN_{EHNF}\%$	19.1000	19.1000	19.0000	19.1667	19.1667	1.3333

gativos más bajo que el que se obtiene mediante la regla bayes. Por tanto, resulta más adecuado utilizar el criterio de decisión EHNF que la regla bayes (2.3), cuando se toma $C_{0i} = C_{1i}$ para $i = 1, \dots, N$, puesto que con el criterio EHNF se alcanza un equilibrio mayor entre los porcentajes de falsos positivos y falsos negativos, ya que se detectan más hipótesis nulas falsas a la vez que el porcentaje de falsos positivos

se mantiene en unos niveles aceptables que es, en definitiva, lo que se pretende.

Con la distribución $Beta(1, 25)$ sí se obtienen importantes cambios en los resultados, que pueden ser debidos a que esta distribución concentra la densidad en torno a valores de p próximos a cero, mientras que los datos se generaron de mixturas con probabilidad $p = 0.9$ y $p = 0.7$. No ocurre lo mismo con la distribución $Beta(0.5, 1)$ que, aunque también da más densidad para los valores próximos a cero, presenta un sesgo mucho menor. Por lo que podemos deducir que si se considera una distribución $Beta(\alpha, \beta)$ como distribución inicial del parámetro p del modelo, el procedimiento bayesiano resulta ser robusto respecto a la elección de los parámetros α y β , siempre y cuando la distribución correspondiente no suponga un sesgo demasiado radical.

Con objeto de comparar el procedimiento bayesiano con el procedimiento de Benjamini y Hochberg (1995) para controlar el FDR, en la tabla 2.7 se muestran los resultados obtenidos con los datos generados mediante una mixtura de $N(0, 1)$ con probabilidad $p = 0.9$ y de $N(1, 1)$ con probabilidad $1 - p = 0.1$, aplicando la regla bayes definida en (2.3), considerando $C_{0i} = C_{1i}$ para $i = 1, \dots, N$, y el criterio de decisión EHNF, utilizando en ambos casos como distribución inicial de $p \sim Beta(1, 1)$, y los resultados obtenidos mediante el procedimiento de Benjamini y Hochberg (1995) para controlar el FDR a los niveles $\alpha = 0.05$ y $\alpha = 0.1$, es decir, el porcentaje de hipótesis nulas rechazadas ($R_{BH}\%$), y los porcentajes de falsos positivos ($FP_{BH}\%$) y falsos negativos ($FN_{BH}\%$) calculados como se indica a continuación,

$$FP_{BH}\% = \frac{\sum_{i=1}^N I(\tilde{p}_i \leq \alpha) I(\mu_i = 0)}{\sum_{i=1}^N I(\mu_i = 0)} \times 100,$$

Tabla 2.7: Resultados aplicando el procedimiento de Benjamini y Hochberg (1995) (BH), la regla bayes y el criterio EHNF, con distribución inicial de $p \sim Beta(1, 1)$, utilizando los datos generados mediante una mixtura de $N(0, 1)$ con probabilidad $p = 0.9$ y de $N(1, 1)$ con probabilidad $1 - p = 0.1$.

	Método BH		Regla bayes		Criterio EHNF		
	$\alpha = 0.05$	$\alpha = 0.1$					
$N = 500$	$R_{BH}\%$	1.80	3.60	$R_C\%$	7.20	$\widehat{N}_1\%$	11
	$FP_{BH}\%$	0	0.4396	$FP_C\%$	2.4176	$FP_{EHNF}\%$	5.0549
	$FN_{BH}\%$	80	64.4444	$FN_C\%$	44.4444	$FN_{EHNF}\%$	28.8889
$N = 1000$	$R_{BH}\%$	3.60	5.90	$R_C\%$	9.50	$\widehat{N}_1\%$	12.30
	$FP_{BH}\%$	0.2240	0.5599	$FP_C\%$	2.3516	$FP_{EHNF}\%$	4.9272
	$FN_{BH}\%$	68.2243	49.5327	$FN_C\%$	30.8411	$FN_{EHNF}\%$	26.1682
$N = 5000$	$R_{BH}\%$	2.12	3.22	$R_C\%$	7.02	$\widehat{N}_1\%$	10.20
	$FP_{BH}\%$	0.1115	0.2900	$FP_C\%$	1.8737	$FP_{EHNF}\%$	3.7921
	$FN_{BH}\%$	80.4642	71.3733	$FN_C\%$	48.3559	$FN_{EHNF}\%$	34.2360
$N = 10000$	$R\%$	1.87	3.07	$R_C\%$	6.55	$\widehat{N}_1\%$	9.44
	$FP_{BH}\%$	0.0331	0.2319	$FP_C\%$	1.5570	$FP_{EHNF}\%$	3.39
	$FN_{BH}\%$	80.5085	69.7034	$FN_C\%$	45.5508	$FN_{EHNF}\%$	32.5212

$$FN_{BH}\% = \frac{\sum_{i=1}^N I(\tilde{p}_i > \alpha) I(\mu_i = 1)}{\sum_{i=1}^N I(\mu_i = 1)} \times 100, \quad (2.15)$$

siendo $\tilde{p}_i = \min_{j=i, \dots, N} \left\{ \min \left(\frac{N}{j} p_{(j)}, 1 \right) \right\}$, para $i = 1, \dots, N$, los p-valores ajustados para este procedimiento.

Se puede observar que el procedimiento de Benjamini y Hochberg (1995), con los niveles usualmente utilizados en la literatura para controlar el FDR, es muy

conservador y con un alto porcentaje de falsos negativos; mientras que con el procedimiento bayesiano, tanto si se aplica la regla bayes como si se aplica el criterio de decisión EHNF, se rechazan más hipótesis nulas obteniéndose unos porcentajes de hipótesis nulas rechazadas, principalmente con el criterio EHNF, muy próximos al porcentaje de hipótesis nulas falsas con el que se generaron los datos (10%).

Además, mediante el procedimiento bayesiano con cualquiera de los dos criterios de decisión, sobre todo con el criterio EHNF, se detectan más hipótesis nulas falsas, por lo que se obtienen unos porcentajes de falsos negativos significativamente más bajos que los que se obtienen utilizando el procedimiento frecuentista, a la vez que los porcentajes de falsos positivos se mantienen en unos niveles aceptables. De hecho, para obtener, con los datos simulados, resultados similares con el procedimiento frecuentista y con el procedimiento bayesiano aplicando la regla bayes (más conservador y con mayor porcentaje de falsos negativos que aplicando el criterio de decisión EHNF) es necesario controlar el FDR a un nivel $\alpha > 0.2$, nivel demasiado elevado para los utilizados habitualmente en la literatura para controlar el FDR ($\alpha \leq 0.1$).

Así, podemos concluir que el procedimiento bayesiano con cualquiera de los dos criterios de decisión, principalmente con el criterio EHNF que nosotros proponemos, es más apropiado que el procedimiento frecuentista de Benjamini y Hochberg (1995) para controlar el FDR a los niveles habitualmente utilizados en la literatura, ya que detecta más hipótesis nulas falsas, manteniendo el porcentaje de falsos positivos en unos niveles admisibles.

2.4 Modelo Gaussiano. Varianza desconocida

Consideramos nuevamente el problema del contraste de hipótesis múltiples expuesto en la sección anterior,

$$H_{0i} : \mu_i = \mu_0 \quad \text{frente a} \quad H_{1i} : \mu_i = \mu_1 \quad (\mu_0 \neq \mu_1), \quad i = 1, \dots, N,$$

donde μ_0 y μ_1 son valores conocidos y comunes para todas las hipótesis.

Suponemos que el vector de observaciones $T_i = (X_{i1}, \dots, X_{in})$, para cada hipótesis, sigue el mismo modelo que en dicha sección, es decir, las T_i , para $i = 1, \dots, N$, son variables aleatorias independientes e idénticamente distribuidas $N(\mu_0, 1/\sqrt{\phi})$ bajo H_{0i} y $N(\mu_1, 1/\sqrt{\phi})$ bajo H_{1i} pero en este caso la precisión ϕ es desconocida. Entonces, podemos considerar la observaciones T_i procedentes de una mixtura de ambas densidades,

$$\begin{aligned} f(t_i|\boldsymbol{\theta}) &= pf_0(t_i|\phi) + (1-p)f_1(t_i|\phi) \\ &= p \prod_{j=1}^n f_0(x_{ij}|\phi) + (1-p) \prod_{j=1}^n f_1(x_{ij}|\phi), \end{aligned} \tag{2.16}$$

donde $t_i = (x_{i1}, \dots, x_{in})$, $\boldsymbol{\theta} = (p, \phi)$, siendo p la probabilidad inicial de cada hipótesis nula y $f_0(t_i|\phi)$ y $f_1(t_i|\phi)$ las densidades bajo la hipótesis nula y bajo la hipótesis alternativa, respectivamente, para $i = 1, \dots, N$. La densidad $f(t_i|\boldsymbol{\theta})$ y la verosimilitud, $f(\mathbf{t}|\boldsymbol{\theta})$, se pueden expresar como en (2.6) y (2.8), respectivamente.

Con el fin de llevar a cabo la inferencia bayesiana debemos definir una distribución inicial conjunta $\pi(\boldsymbol{\theta})$, para lo que consideramos dos distribuciones independientes. Para el parámetro p se considera como distribución inicial, igual que en la sección anterior, una distribución $Beta(\alpha, \beta)$, y para el parámetro ϕ una distribución $Gamma(a/2, b/2)$ por simplicidad en los cálculos, ya que esta distribución

es la distribución inicial natural conjugada. Por tanto,

$$\pi(\boldsymbol{\theta}|\alpha, \beta, a, b) = \pi(p|\alpha, \beta) \pi(\phi|a, b),$$

con

$$\begin{aligned} p &\sim \text{Beta}(\alpha, \beta), \\ \phi &\sim \text{Gamma}(a/2, b/2). \end{aligned} \tag{2.17}$$

Si bien, se pueden considerar otras posibilidades.

La probabilidad final de cada hipótesis nula, $Pr(H_{0i} = 0|\mathbf{t}, \alpha, \beta, a, b)$, se puede expresar como un cociente de esperanzas con respecto a la distribución inicial de $\boldsymbol{\theta}$,

$$\begin{aligned} Pr(H_{0i} = 0|\mathbf{t}, \alpha, \beta, a, b) &= \frac{\int g(t_i|\boldsymbol{\theta}) f(\mathbf{t}|\boldsymbol{\theta}) \pi(\boldsymbol{\theta}|\alpha, \beta, a, b) d\boldsymbol{\theta}}{\int f(\mathbf{t}|\boldsymbol{\theta}) \pi(\boldsymbol{\theta}|\alpha, \beta, a, b) d\boldsymbol{\theta}} \\ &= \frac{\int_0^\infty \int_0^1 g(t_i|\boldsymbol{\theta}) f(\mathbf{t}|\boldsymbol{\theta}) \pi(p|\alpha, \beta) \pi(\phi|a, b) dp d\phi}{\int_0^\infty \int_0^1 f(\mathbf{t}|\boldsymbol{\theta}) \pi(p|\alpha, \beta) \pi(\phi|a, b) dp d\phi}, \\ &= \frac{E_{\pi(\boldsymbol{\theta}|\alpha, \beta, a, b)} [g(t_i|\boldsymbol{\theta}) f(\mathbf{t}|\boldsymbol{\theta})]}{E_{\pi(\boldsymbol{\theta}|\alpha, \beta, a, b)} [f(\mathbf{t}|\boldsymbol{\theta})]}, \end{aligned}$$

donde $g(t_i|\boldsymbol{\theta}) = \frac{pf_0(t_i|\phi)}{pf_0(t_i|\phi) + (1-p)f_1(t_i|\phi)}$.

Entonces aplicando el método de Monte Carlo, de manera similar a (2.2), podemos aproximar la probabilidad final de cada hipótesis nula de la forma siguiente,

$$\hat{Pr}(H_{0i} = 0|\mathbf{t}, \alpha, \beta, a, b) = \frac{\sum_{k=1}^M g(t_i|\boldsymbol{\theta}_k) f(\mathbf{t}|\boldsymbol{\theta}_k)}{\sum_{k=1}^M f(\mathbf{t}|\boldsymbol{\theta}_k)},$$

donde $\boldsymbol{\theta}_k = (p_k, \phi_k)$, siendo p_1, \dots, p_M y ϕ_1, \dots, ϕ_M muestras aleatorias de las distribuciones iniciales $Beta(\alpha, \beta)$ y $Gamma(a/2, b/2)$, respectivamente.

De forma análoga a como se hizo en (2.4), podemos también estimar p y ϕ mediante las correspondientes medias marginales finales,

$$\begin{aligned} \hat{p} &= \int_0^1 \int_0^\infty p \pi(\boldsymbol{\theta}|\mathbf{t}, \alpha, \beta, a, b) d\phi dp \\ &= \frac{\int_0^1 \int_0^\infty p f(\mathbf{t}|\boldsymbol{\theta}) \pi(p|\alpha, \beta) \pi(\phi|a, b) d\phi dp}{\int_0^\infty \int_0^1 f(\mathbf{t}|\boldsymbol{\theta}) \pi(p|\alpha, \beta) \pi(\phi|a, b) dp d\phi}, \end{aligned}$$

y

$$\begin{aligned} \hat{\phi} &= \int_0^\infty \int_0^1 \phi \pi(\boldsymbol{\theta}|\mathbf{t}, \alpha, \beta, a, b) dp d\phi \\ &= \frac{\int_0^\infty \int_0^1 \phi f(\mathbf{t}|\boldsymbol{\theta}) \pi(p|\alpha, \beta) \pi(\phi|a, b) dp d\phi}{\int_0^\infty \int_0^1 f(\mathbf{t}|\boldsymbol{\theta}) \pi(p|\alpha, \beta) \pi(\phi|a, b) dp d\phi}, \end{aligned}$$

es decir, \hat{p} y $\hat{\phi}$ se pueden escribir como un cociente de esperanzas con respecto a las distribuciones iniciales de p y ϕ y, por tanto, se pueden aproximar por el método de Monte Carlo de manera similar a (2.5); esto es,

$$\hat{p} \approx \frac{\sum_{i=1}^M p_i f(\mathbf{t}|\boldsymbol{\theta}_i)}{\sum_{i=1}^M f(\mathbf{t}|\boldsymbol{\theta}_i)}, \quad \hat{\phi} \approx \frac{\sum_{i=1}^M \phi_i f(\mathbf{t}|\boldsymbol{\theta}_i)}{\sum_{i=1}^M f(\mathbf{t}|\boldsymbol{\theta}_i)},$$

donde $\boldsymbol{\theta}_i = (p_i, \phi_i)$, siendo p_1, \dots, p_M y ϕ_1, \dots, ϕ_M muestras aleatorias de las distribuciones iniciales $Beta(\alpha, \beta)$ y $Gamma(a/2, b/2)$, respectivamente.

Si bien el procedimiento resulta sencillo de aplicar, presenta el inconveniente de que no se puede utilizar en el caso de que la distribución inicial para ϕ sea la

distribución $Gamma(0, 0)$, debido a que al aplicar el método de Monte Carlo no podemos simular directamente de esta distribución por ser impropia. Sin embargo, se puede efectuar la inferencia bayesiana utilizando los métodos basados en cadenas de Markov Monte Carlo (MCMC), que detallamos en la siguiente subsección.

2.4.1 Inferencia basada en los métodos MCMC. Muestreo de Gibbs

En un contexto bayesiano, los métodos MCMC permiten aproximar integrales que constituyen la esperanza final de una determinada función, $g(\boldsymbol{\theta})$, de los parámetros, donde $\boldsymbol{\theta} = (\theta_1, \dots, \theta_k)$,

$$E[g(\boldsymbol{\theta})|\mathbf{t}] = \int_{\Theta} g(\boldsymbol{\theta})\pi(\boldsymbol{\theta}|\mathbf{t}) d\boldsymbol{\theta}. \quad (2.18)$$

Ver Robert y Casella (2004) para una revisión de estos métodos. La estrategia consiste en construir una cadena de Markov $\{\boldsymbol{\theta}^{(j)} : j = 1, \dots, M\}$, donde $\boldsymbol{\theta}^{(j)} = (\theta_1^{(j)}, \dots, \theta_k^{(j)})$, con distribución estacionaria $\pi(\boldsymbol{\theta}|\mathbf{t})$, de modo que

$$\frac{1}{M} \sum_{j=1}^M g(\boldsymbol{\theta}^{(j)}) \xrightarrow{c.s.} E_{\pi}[g(\boldsymbol{\theta})], \quad \text{cuando } M \longrightarrow \infty.$$

Así, se pueden aproximar integrales del tipo (2.18) con la media muestral obtenida a partir de una muestra de la cadena de Markov en equilibrio.

El problema en los métodos MCMC es encontrar las probabilidades de transición de la cadena que permitan que la distribución estacionaria sea la distribución de interés. Uno de los métodos más comunes es el muestreo de Gibbs, véase Gelfand y Smith (1990), que puede ser utilizado cuando es posible generar valores de la distribución final $\pi(\theta_i|\boldsymbol{\theta}_{-i}, \mathbf{t})$ de cada parámetro, θ_i , condicionada al resto de parámetros, $\boldsymbol{\theta}_{-i} = (\theta_1, \dots, \theta_{i-1}, \theta_{i+1}, \dots, \theta_k)$. Para generar valores de una cadena

de Markov construida según un muestreo de Gibbs se comienza con unos valores iniciales arbitrarios, $\boldsymbol{\theta}^{(0)}$, actualizando cada parámetro, $\theta_i^{(j)}$, mediante un valor $\theta_i^{(j+1)}$ generado de la distribución

$$\pi(\theta_i | \theta_1^{(j+1)}, \dots, \theta_{i-1}^{(j+1)}, \theta_{i+1}^{(j)}, \dots, \theta_k^{(j)}, \mathbf{t}),$$

hasta que la cadena alcanza la convergencia. Los siguientes valores generados por la cadena, a partir de ese momento, se toman como muestra de la distribución final, ya que la cadena construida de esta manera tiene por distribución estacionaria la distribución final, $\pi(\boldsymbol{\theta} | \mathbf{t})$, véase Smith y Roberts (1993).

Por tanto, dada una muestra de la cadena de Markov en equilibrio, se puede aproximar la esperanza final de una determinada función con la media muestral obtenida a partir de dicha muestra. En particular, se pueden estimar los parámetros del modelo aproximando las medias finales de los mismos con las correspondientes medias muestrales obtenidas de la muestra de la cadena de Markov en equilibrio,

$$\bar{\theta}_r = \frac{1}{M} \sum_{j=1}^M \theta_r^{(j)} \xrightarrow{c.s.} E_{\pi}[\theta_r]. \quad (2.19)$$

El uso de los métodos MCMC en la estadística bayesiana ha aumentado notablemente en los últimos años, debido a que permiten salvar la dificultad que se presenta, en muchas ocasiones, en la inferencia bayesiana cuando se pretenden calcular determinadas cantidades finales de interés. El muestreo de Gibbs constituye uno de estos métodos que es aplicable a una amplia clase de problemas bayesianos y ha sido utilizado, en el contexto de los contrastes de hipótesis múltiples, por algunos autores como Casella (2001), Ibrahim et al. (2002), Do et al. (2005), Lönnstedt y Britton (2005) y Gottardo et al. (2006), entre otros.

Consideramos de nuevo el problema del contraste de hipótesis múltiples expuesto

al principio de la sección 2.4,

$$H_{0i} : \mu_i = \mu_0 \quad \text{frente a} \quad H_{1i} : \mu_i = \mu_1 \quad (\mu_0 \neq \mu_1), \quad i = 1, \dots, N,$$

donde μ_0 y μ_1 son valores conocidos y comunes para todas las hipótesis.

Suponemos que los vectores de observaciones $T_i = (X_{i1}, \dots, X_{in})$ son, para todo i , variables aleatorias independientes e idénticamente distribuidas $N(\mu_0, 1/\sqrt{\phi})$ bajo H_{0i} y $N(\mu_1, 1/\sqrt{\phi})$ bajo H_{1i} , por lo que podemos considerar la observaciones T_i procedentes de la mixtura (2.16),

$$\begin{aligned} f(t_i|\boldsymbol{\theta}) &= pf_0(t_i|\phi) + (1-p)f_1(t_i|\phi) \\ &= p \prod_{j=1}^n f_0(x_{ij}|\phi) + (1-p) \prod_{j=1}^n f_1(x_{ij}|\phi), \end{aligned}$$

donde $t_i = (x_{i1}, \dots, x_{in})$, $\boldsymbol{\theta} = (p, \phi)$, siendo p la probabilidad inicial de la hipótesis nula, ϕ la precisión que suponemos desconocida y $f_0(t_i|\phi)$ y $f_1(t_i|\phi)$ las densidades bajo la hipótesis nula y bajo la hipótesis alternativa respectivamente, para $i = 1, \dots, N$.

A continuación aplicamos el muestreo de Gibbs para aproximar \hat{p} , $\hat{\phi}$ y la probabilidad final de cada hipótesis nula, ya que las distribuciones finales condicionadas son, como se verá después, explícitas y propias.

En primer lugar, con objeto de simplificar la verosimilitud como habitualmente se hace en mixturas, véase Diebolt y Robert (1994), consideramos N variables latentes i.i.d., Z_1, \dots, Z_N , cada una de ellas asociada a un vector de observaciones T_i , indicando la componente de la mixtura en (2.16) a la que pertenece el correspondiente vector de observaciones.

Por otro lado, como la probabilidad con la que cada vector de observaciones procede de cada una de las dos componentes de la mixtura viene dada por p y $1-p$,

resulta natural considerar como distribución inicial para Z_i ,

$$Pr(Z_i = 0|p) = p \quad \text{y} \quad Pr(Z_i = 1|p) = 1 - p, \quad i = 1, \dots, N,$$

donde $Z_i = 0$ indica que T_i procede de la primera componente de la mixtura y $Z_i = 1$ indica que T_i procede de la segunda componente.

De este modo, una vez observado el valor de las variables latentes, Z_i , las variables T_i son variables aleatorias independientes distribuidas según la densidad de la componente de la mixtura correspondiente,

$$T_i|Z_i = 0 \sim N(\mu_0, 1/\sqrt{\phi}) \quad \text{y} \quad T_i|Z_i = 1 \sim N(\mu_1, 1/\sqrt{\phi}), \quad i = 1, \dots, N,$$

es decir, $f(t_i|z_i = 0, \phi) = f_0(t_i|\phi)$ y $f(t_i|z_i = 1, \phi) = f_1(t_i|\phi)$.

De esta manera la verosimilitud se puede escribir como

$$\begin{aligned} f(\mathbf{t}, \mathbf{z}|\boldsymbol{\theta}) &= \prod_{i=1}^N f(t_i, z_i|\boldsymbol{\theta}) \\ &= \prod_{i:z_i=0} f(t_i, z_i = 0|\boldsymbol{\theta}) \prod_{i:z_i=1} f(t_i, z_i = 1|\boldsymbol{\theta}) \\ &= \prod_{i:z_i=0} Pr(z_i = 0|p) f(t_i|z_i = 0, \phi) \prod_{i:z_i=1} Pr(z_i = 1|p) f(t_i|z_i = 1, \phi) \\ &= \prod_{i:z_i=0} p f_0(t_i|\phi) \prod_{i:z_i=1} (1-p) f_1(t_i|\phi) \\ &= p^{n_0} (1-p)^{n_1} \prod_{i:z_i=0} \prod_{j=1}^n f_0(x_{ij}|\phi) \prod_{i:z_i=1} \prod_{j=1}^n f_1(x_{ij}|\phi), \end{aligned}$$

donde $\mathbf{t} = (t_1, \dots, t_N)$, $\mathbf{z} = (z_1, \dots, z_N)$, $n_0 = \sum_i I(z_i = 0)$ es el número de observaciones de la primera componente de la mixtura y $n_1 = \sum_i I(z_i = 1)$ el número de observaciones de la segunda componente, siendo $N = n_0 + n_1$. Por tanto, dado que $f_0(t_i|\phi)$ es la densidad de la distribución $N(\mu_0, 1/\sqrt{\phi})$ y $f_1(t_i|\phi)$ es la densidad de la distribución $N(\mu_1, 1/\sqrt{\phi})$, la verosimilitud, $f(\mathbf{t}, \mathbf{z}|\boldsymbol{\theta})$, se puede expresar como

$$f(\mathbf{t}, \mathbf{z}|\boldsymbol{\theta}) \propto k \prod_{i:z_i=0} u_{0i} \prod_{i:z_i=1} u_{1i}, \quad (2.20)$$

donde $k = p^{n_0}(1-p)^{n_1}\phi^{\frac{nN}{2}}e^{-\frac{\phi}{2}\sum_{i=1}^N\sum_{j=1}^n x_{ij}^2}$, siendo u_{0i} y u_{1i} las expresiones definidas en (2.7).

Entonces, la distribución final de $(\boldsymbol{\theta}, \mathbf{z})$ viene dada por

$$\pi(\boldsymbol{\theta}, \mathbf{z}|\mathbf{t}) \propto f(\mathbf{t}, \mathbf{z}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}).$$

Obsérvese que, en un contexto bayesiano, las variables latentes representan un conjunto adicional de parámetros.

Teniendo en cuenta las distribuciones iniciales definidas en (2.17) y la verosimilitud (2.20), se obtienen las distribuciones finales de cada uno de los parámetros del modelo condicionadas al resto de parámetros. Distribuciones que nos permitirán generar valores de una cadena MCMC, construida según el muestreo de Gibbs, cuya distribución estacionaria es la distribución final, $\pi(\boldsymbol{\theta}, \mathbf{z}|\mathbf{t}, \alpha, \beta, a, b)$.

Se puede ver fácilmente que la probabilidad final condicionada de que $z_i = 0$ o, equivalentemente, de que $t_i = (x_{i1}, \dots, x_{in})$ se haya generado por la primera componente de la mixtura, dados los datos y el resto de parámetros, es

$$Pr(z_i = 0|t_i, \boldsymbol{\theta}) = \left(1 + \frac{(1-p)u_{1i}}{p u_{0i}}\right)^{-1}, \quad (2.21)$$

siendo u_{0i} y u_{1i} las expresiones definidas en (2.7). Por tanto, la probabilidad final condicionada de que $z_i = 1$ o de que $t_i = (x_{i1}, \dots, x_{in})$ se haya generado por la segunda componente de la mixtura, dados los datos y el resto de parámetros, es $Pr(z_i = 1|t_i, \boldsymbol{\theta}) = 1 - Pr(z_i = 0|t_i, \boldsymbol{\theta})$.

También se puede comprobar fácilmente que la distribución final condicionada de p , dados los datos y el resto de parámetros, es

$$p|\mathbf{t}, \mathbf{z} \sim Beta(n_0 + \alpha, n_1 + \beta) \quad (2.22)$$

y la distribución final condicionada de ϕ , dados los datos y el resto de parámetros, es

$$\phi|\mathbf{t}, \mathbf{z} \sim \text{Gamma}\left(\frac{a + Nn}{2}, \frac{R}{2}\right), \quad (2.23)$$

donde $R = b + \sum_{i=1}^N \sum_{j=1}^n x_{ij}^2 + n \sum_{i:z_i=0} (\mu_0^2 - 2\mu_0\bar{x}_i) + n \sum_{i:z_i=1} (\mu_1^2 - 2\mu_1\bar{x}_i)$. Obsérvese que para $a = 0$ y $b = 0$ la distribución $\text{Gamma}((a + Nn)/2, R/2)$ es propia, salvando así el problema que existía, al simular directamente de la distribución inicial, cuando se aplicaba el método de Monte Carlo.

En el apéndice 2A de este capítulo se muestra la obtención de las distribuciones (2.21), (2.22) y (2.23).

Con las distribuciones (2.21), (2.22) y (2.23) se puede construir una cadena de Markov cuya distribución estacionaria sea la distribución final de los parámetros $(\boldsymbol{\theta}, \mathbf{z})$. Así, podemos obtener muestras de la distribución final de los parámetros generando valores de la cadena de Markov en equilibrio. El algoritmo MCMC consiste en un muestreo de Gibbs en el que se han incluido las variables latentes. La estructura del algoritmo es la siguiente,

1. Fijar valores iniciales para $\boldsymbol{\theta}^{(0)} = (p^{(0)}, \phi^{(0)})$.
2. Actualizar las variables latentes generando de $\mathbf{z}^{(j+1)} \sim \mathbf{z}|\mathbf{t}, \boldsymbol{\theta}^{(j)}$.
3. Actualizar los parámetros $\boldsymbol{\theta}^{(j+1)} = (p^{(j+1)}, \phi^{(j+1)})$.
 - 3.1. Generar valores de $p^{(j+1)} \sim p|\mathbf{t}, \mathbf{z}^{(j+1)}$.
 - 3.2. Generar valores de $\phi^{(j+1)} \sim \phi|\mathbf{t}, \mathbf{z}^{(j+1)}$.
4. $j = j + 1$. Ir a 2.

Este algoritmo es fácil de implementar ya que todas las distribuciones condicionadas son explícitas. En el paso 1 se fijan valores iniciales para los parámetros del modelo. En el paso 2 se genera un valor para cada una de las variables latentes a partir de la distribución dada en (2.21) y finalmente, en el paso 3, se genera un valor para cada uno de los parámetros p y ϕ a partir de las distribuciones dadas en (2.22) y (2.23), respectivamente.

Dada $\{(\boldsymbol{\theta}^{(j)}, \mathbf{z}^{(j)}) : j = 1, \dots, M\}$ una muestra de la cadena de Markov en equilibrio, donde $\boldsymbol{\theta}^{(j)} = (p^{(j)}, \phi^{(j)})$ y $\mathbf{z}^{(j)} = (z_1^{(j)}, \dots, z_N^{(j)})$, se puede hacer inferencia sobre los parámetros. Por ejemplo, se puede estimar la media final de cada parámetro con la correspondiente media muestral, ya que las medias muestrales convergen a su esperanza bajo la distribución estacionaria de la cadena, $\pi(\boldsymbol{\theta}, \mathbf{z} | \mathbf{t}, \alpha, \beta, a, b)$, como se indica en (2.19). Por tanto podemos estimar p mediante la siguiente expresión,

$$\hat{p} = E[p | \mathbf{t}, \alpha, \beta, a, b] \approx \frac{1}{M} \sum_{j=1}^M p^{(j)}, \quad (2.24)$$

de forma análoga, podemos estimar ϕ mediante

$$\hat{\phi} = E[\phi | \mathbf{t}, \alpha, \beta, a, b] \approx \frac{1}{M} \sum_{j=1}^M \phi^{(j)}. \quad (2.25)$$

Por último, podemos aproximar la probabilidad final de cada hipótesis nula como

$$\begin{aligned} Pr(H_{0i} = 0 | \mathbf{t}, \alpha, \beta, a, b) &= 1 - Pr(H_{0i} = 1 | \mathbf{t}, \alpha, \beta, a, b) \\ &= 1 - Pr(\mu_i = \mu_1 | \mathbf{t}, \alpha, \beta, a, b) \\ &= 1 - Pr(z_i = 1 | \mathbf{t}, \alpha, \beta, a, b) \\ &\approx 1 - \frac{1}{M} \sum_{j=1}^M I(z_i^{(j)} = 1), \end{aligned} \quad (2.26)$$

para $i = 1, \dots, N$.

Una vez se han obtenido, para cada i , una estimación de la probabilidad final de cada hipótesis nula, $\hat{Pr}(H_{0i} = 0 | \mathbf{t}, \alpha, \beta, a, b)$, aplicando (2.26), y una estimación

de p , \hat{p} , mediante (2.24), podemos aplicar los criterios de decisión descritos en la sección 2.2. Así, se rechazarán todas las hipótesis nulas H_{0i} tales que $\hat{P}r(H_{0i} = 0 | \mathbf{t}, \alpha, \beta, a, b) \leq C_{0i}/(C_{0i} + C_{1i})$, aceptando el resto, si se aplica la regla bayes, o se rechazarán las $\widehat{N}_1 = N(1 - \hat{p})$ hipótesis nulas con menor probabilidad final estimada, aceptando el resto, si se aplica el criterio EHNF.

2.4.2 Ejemplo

Al igual que en la subsección 2.3.1, consideramos el siguiente problema de contrastes múltiples,

$$H_{0i} : \mu_i = 0 \quad \text{frente a} \quad H_{1i} : \mu_i = 1, \quad i = 1, \dots, N.$$

Sea $T_i = (X_{i1}, \dots, X_{in})$ el vector de observaciones para cada hipótesis, y supongamos que las T_i , para $i = 1, \dots, N$, son variables aleatorias independientes con distribución $N(0, 1/\sqrt{\phi})$ bajo H_{0i} y $N(1, 1/\sqrt{\phi})$ bajo H_{1i} , siendo ϕ la precisión desconocida.

Entonces, considerando las distribuciones iniciales definidas en (2.17) y las distribuciones (2.21), (2.22) y (2.23) para $\mu_0 = 0$ y $\mu_1 = 1$, podemos obtener, aplicando el algoritmo MCMC descrito en la subsección anterior, una muestra de la cadena de Markov cuya distribución estacionaria es la distribución final $\pi(\boldsymbol{\theta}, \mathbf{z} | \mathbf{t}, \alpha, \beta, a, b)$, para estimar p , ϕ y la probabilidad final de cada hipótesis nula como en (2.24), (2.25) y (2.26), respectivamente.

2.4.3 Simulación: resultados y análisis de sensibilidad

En esta subsección, utilizando el ejemplo anterior, realizamos un estudio de simulación con el fin de evaluar los resultados del procedimiento propuesto en esta sección.

Tabla 2.8: Estimación final de p para distintos valores de los parámetros de la distribución inicial de $p \sim Beta(\alpha, \beta)$ y de $\phi \sim Gamma(a/2, b/2)$, utilizando los datos generados mediante una mixtura de $N(0, 1)$ con probabilidad $p = 0.9$ y de $N(1, 1)$ con probabilidad $1 - p = 0.1$.

		(α, β)	(1, 0.5)	(1, 1)	(0.5, 1)
$N = 500$	(a, b)	(0,0)	0.8897	0.8881	0.8879
		(1,0.5)	0.8899	0.8885	0.8880
		(10,1)	0.8899	0.8876	0.8873
$N = 10000$	(a, b)	(0,0)	0.9054	0.9054	0.9054
		(1,0.5)	0.9056	0.9054	0.9055
		(10,1)	0.9054	0.9054	0.9054

Se consideran, al igual que en la subsección 2.3.2, dos simulaciones. En la primera se utilizan los datos generados en la primera simulación de dicha subsección, es decir, los datos generados de una mixtura de $N(0, 1)$ con probabilidad $p = 0.9$ y de $N(1, 1)$ con probabilidad $1 - p = 0.1$, pero sólo para $N = 500$ y $N = 10000$ hipótesis y con $n = 5$ observaciones para cada hipótesis. En la segunda se simulan dos nuevos conjuntos de datos, con el mismo número de hipótesis y el mismo número de observaciones por hipótesis que en la primera simulación, pero los datos son generados de una mixtura de $N(0, 1/\sqrt{0.5})$ con probabilidad $p = 0.9$ y de $N(1, 1/\sqrt{0.5})$ con probabilidad $1 - p = 0.1$.

Para la distribución inicial del parámetro p se consideró una distribución $Beta(\alpha, \beta)$ con los siguientes valores para los parámetros (α, β) : (1, 0.5), (1, 1), y (0.5, 1). Para ϕ se consideró una distribución $Gamma(a/2, b/2)$ tomando como valores para (a, b) : (0, 0), (1, 0.5) y (10, 1). Se eligieron estos valores para los parámetros de la distribución inicial de ϕ con el objetivo de realizar un análisis de sensibili-

Tabla 2.9: Estimación final de p para distintos valores de los parámetros de la distribución inicial de $p \sim Beta(\alpha, \beta)$ y de $\phi \sim Gamma(a/2, b/2)$, utilizando los datos generados mediante una mixtura de $N(0, 1/\sqrt{0.5})$ con probabilidad $p = 0.9$ y de $N(1, 1/\sqrt{0.5})$ con probabilidad $1 - p = 0.1$.

		(α, β)	(1, 0.5)	(1, 1)	(0.5, 1)
$N = 500$	(a, b)	(0,0)	0.8781	0.8751	0.8736
		(1,0.5)	0.8773	0.8750	0.8733
		(10,1)	0.8761	0.8745	0.8739
$N = 10000$	(a, b)	(0,0)	0.9026	0.9025	0.9019
		(1,0.5)	0.9023	0.9020	0.9023
		(10,1)	0.9024	0.9022	0.9023

dad, ya que las distribuciones correspondientes son muy distintas. La distribución $Gamma(1/2, 0.5/2)$ debería ser, entre las consideradas, la más apropiada, puesto que concentra la densidad en torno a los valores de ϕ con los que se generaron los datos en ambas mixturas ($\phi = 1$ y $\phi = 0.5$), mientras que uno de los rangos donde menor densidad da la distribución $Gamma(10/2, 1/2)$ es precisamente en torno a estos valores de ϕ , y la distribución $Gamma(0, 0)$ es la distribución no informativa.

Con objeto de realizar un análisis de sensibilidad, para cada conjunto de datos simulados y para los distintos valores considerados de los parámetros (α, β) y (a, b) , obtenemos una muestra de la cadena de Markov construida mediante un muestreo de Gibbs, aplicando el algoritmo descrito en la subsección 2.4.1 con las distribuciones (2.21), (2.22) y (2.23) para $\mu_0 = 0$ y $\mu_1 = 1$, con el fin de estimar p , ϕ y la probabilidad final de cada hipótesis nula como en (2.24), (2.25) y (2.26), respectivamente. Es importante señalar que el coste computacional es muy bajo, empleando entre nueve y cien segundos, aproximadamente, dependiendo del número de hipótesis y realizando

Tabla 2.10: Estimación final de ϕ para distintos valores de los parámetros de la distribución inicial de $p \sim \text{Beta}(\alpha, \beta)$ y de $\phi \sim \text{Gamma}(a/2, b/2)$, utilizando los datos generados mediante una mixtura de $N(0, 1)$ con probabilidad $p = 0.9$ y de $N(1, 1)$ con probabilidad $1 - p = 0.1$.

		(α, β)	(1, 0.5)	(1, 1)	(0.5, 1)
$N = 500$	(a, b)	(0,0)	1.0185	1.0185	1.0185
		(1,0.5)	1.0184	1.0187	1.0184
		(10,1)	1.0212	1.0224	1.0232
$N = 10000$	(a, b)	(0,0)	1.0004	1.0004	1.0005
		(1,0.5)	1.0003	1.0004	1.0005
		(10,1)	1.0006	1.0006	1.0006

20000 iteraciones (descartando las 10000 primeras como iteraciones iniciales).

Las estimaciones de p , para los distintos valores de los parámetros de las distribuciones iniciales de p y ϕ , se muestran en la tabla 2.8 con los datos generados mediante una mixtura de $N(0, 1)$ con probabilidad $p = 0.9$ y de $N(1, 1)$ con probabilidad $1 - p = 0.1$.

En la tabla 2.9 se muestran las estimaciones de p con los datos generados con la mixtura de $N(0, 1/\sqrt{0.5})$ con probabilidad $p = 0.9$ y de $N(1, 1/\sqrt{0.5})$ con probabilidad $1 - p = 0.1$ para los mismos valores de los parámetros de las distribuciones iniciales de p y ϕ . Como se puede observar, el procedimiento es robusto respecto a las distribuciones consideradas ya que no se observan cambios importantes en las estimaciones, siendo éstas muy próximas a la probabilidad de la mixtura con las que se generaron los datos ($p = 0.9$).

En la tabla 2.10 se presentan las estimaciones de ϕ obtenidas para los distintos

Tabla 2.11: Estimación final de ϕ para distintos valores de los parámetros de la distribución inicial de $p \sim Beta(\alpha, \beta)$ y de $\phi \sim Gamma(a/2, b/2)$, utilizando los datos generados mediante una mixtura de $N(0, 1/\sqrt{0.5})$ con probabilidad $p = 0.9$ y de $N(1, 1/\sqrt{0.5})$ con probabilidad $1 - p = 0.1$.

		(α, β)	(1, 0.5)	(1, 1)	(0.5, 1)
$N = 500$	(a, b)	(0,0)	0.5092	0.5094	0.5099
		(1,0.5)	0.5093	0.5093	0.5096
		(10,1)	0.5108	0.5116	0.5112
$N = 10000$	(a, b)	(0,0)	0.5041	0.5041	0.5042
		(1,0.5)	0.5041	0.5041	0.5041
		(10,1)	0.5041	0.5042	0.5042

valores de los parámetros (α, β) y (a, b) , utilizando los datos generados mediante una mixtura de $N(0, 1)$ con probabilidad $p = 0.9$ y de $N(1, 1)$ con probabilidad $1 - p = 0.1$, y en la tabla 2.11 para los datos generados con la mixtura de $N(0, 1/\sqrt{0.5})$ con probabilidad $p = 0.9$ y de $N(1, 1/\sqrt{0.5})$ con probabilidad $1 - p = 0.1$. Se puede observar que el procedimiento es robusto respecto a las distribuciones consideradas, obteniéndose unos valores estimados de ϕ muy próximos a los valores con los que se generaron los datos en ambas mixturas ($\phi = 1$ y $\phi = 0.5$).

Para decidir qué hipótesis nulas se rechazan aplicamos, en primer lugar, la regla bayes (2.3) considerando, como en la subsección 2.3.2, $C_{0i} = C_{1i}$ con $i = 1, \dots, N$, por lo que se rechazarán todas las hipótesis nulas tales que $\hat{Pr}(H_{0i} = 0 | \mathbf{t}, \alpha, \beta, a, b) \leq 0.5$. Por último, aplicamos el criterio de decisión EHNF rechazando las $\widehat{N}_1 = N(1 - \hat{p})$ hipótesis nulas con menor probabilidad final estimada.

En las tablas 2.12, 2.13, 2.14 y 2.15 se muestran los resultados obtenidos aplicando la regla bayes, es decir, el porcentaje de hipótesis nulas rechazadas ($R_C\%$) y

Tabla 2.12: Resultados aplicando la regla bayes y el criterio de decisión EHNF, para distintos valores de los parámetros de la distribución inicial de $p \sim Beta(\alpha, \beta)$ y de $\phi \sim Gamma(a/2, b/2)$, utilizando los datos generados mediante una mixtura de $N(0, 1)$ con probabilidad $p = 0.9$ y de $N(1, 1)$ con probabilidad $1 - p = 0.1$, para $N = 500$.

(a, b)	(α, β)	$R_C\%$	$FP_C\%$	$FN_C\%$	$\widehat{N}_1\%$	$prob_{\widehat{N}_1}$	$FP_{EHNF}\%$	$FN_{EHNF}\%$
(0,0)	(1, 0.5)	7.20	2.4176	44.4444	11	0.6006	5.0549	28.8889
	(1, 1)	7.20	2.4176	44.4444	11.20	0.6036	5.2747	28.8889
	(0.5, 1)	7.20	2.4176	44.4444	11.20	0.6032	5.2747	28.8889
(1,0.5)	(1, 0.5)	7.20	2.4176	44.4444	11	0.6007	5.0549	28.8889
	(1, 1)	7.20	2.4176	44.4444	11.20	0.5996	5.4945	28.8889
	(0.5, 1)	7.40	2.6374	44.4444	11.20	0.5981	5.2747	28.8889
(10,1)	(1, 0.5)	7.20	2.4176	44.4444	11	0.6029	5.0549	28.8889
	(1, 1)	7.40	2.6374	44.4444	11.20	0.5985	5.2747	28.8889
	(0.5, 1)	7.40	2.6374	44.4444	11.20	0.5996	5.2747	28.8889

los porcentajes de falsos positivos ($FP_C\%$) y falsos negativos ($FN_C\%$) calculados como en (2.13), así como los resultados obtenidos utilizando el criterio de decisión EHNF, esto es, el porcentaje estimado de hipótesis nulas falsas ($\widehat{N}_1\%$), la probabilidad final estimada más alta con la que se rechaza cada hipótesis nula, $prob_{\widehat{N}_1} = \hat{Pr}(H_{(0\widehat{N}_1)} = 0 | \mathbf{t}, \alpha, \beta, a, b)$, y los porcentajes de falsos positivos ($FP_{EHNF}\%$) y falsos negativos ($FN_{EHNF}\%$) calculados como en (2.14).

En las tablas 2.12 y 2.13 se utilizaron, para $N = 500$ y $N = 10000$, respectivamente, los datos generados mediante la mixtura de $N(0, 1)$ con probabilidad $p = 0.9$ y de $N(1, 1)$ con probabilidad $1 - p = 0.1$, obteniéndose resultados muy similares a los obtenidos en las tablas 2.3 y 2.5 para el caso de varianza conocida, es decir, se puede observar que el procedimiento es robusto respecto a los parámetros considerados.

Tabla 2.13: Resultados aplicando la regla bayes y el criterio de decisión EHNF, para distintos valores de los parámetros de la distribución inicial de $p \sim Beta(\alpha, \beta)$ y de $\phi \sim Gamma(a/2, b/2)$, utilizando los datos generados mediante una mixtura de $N(0, 1)$ con probabilidad $p = 0.9$ y de $N(1, 1)$ con probabilidad $1 - p = 0.1$, para $N = 10000$.

(a, b)	(α, β)	$R_C\%$	$FP_C\%$	$FN_C\%$	$\widehat{N}_1\%$	$prob_{\widehat{N}_1}$	$FP_{EHNF}\%$	$FN_{EHNF}\%$
(0,0)	(1, 0.5)	6.55	1.5349	45.3390	9.46	0.6721	3.4231	32.6271
	(1, 1)	6.57	1.5680	45.4449	9.46	0.6740	3.3900	32.3093
	(0.5, 1)	6.55	1.5459	45.4449	9.46	0.6720	3.4231	32.6271
(1,0.5)	(1, 0.5)	6.57	1.5680	45.4449	9.44	0.6748	3.3790	32.4153
	(1, 1)	6.53	1.5459	45.6568	9.46	0.6728	3.4121	32.5212
	(0.5, 1)	6.53	1.5349	45.5508	9.45	0.6710	3.4011	32.5212
(10,1)	(1, 0.5)	6.54	1.5570	45.6568	9.46	0.6722	3.4011	32.4153
	(1, 1)	6.53	1.5459	45.6568	9.46	0.6714	3.4121	32.5212
	(0.5, 1)	6.54	1.5459	45.5508	9.46	0.6700	3.3900	32.3093

Cabe destacar también que el procedimiento seguido aplicando la regla bayes es más conservador que aplicando el criterio EHNF, siendo el porcentaje estimado de hipótesis nulas falsas, obtenido con este último, muy próximo al porcentaje de hipótesis nulas falsas con los que se generaron los datos (10%).

Por otro lado, hay que tener en cuenta que las hipótesis nulas adicionales que se rechazan con el criterio EHNF tienen una probabilidad final estimada de ser ciertas mayor que 0.5, por lo que se rechazan más hipótesis nulas ciertas, aumentando así el porcentaje de falsos positivos, aunque se mantiene en unos niveles aceptables ($< 6\%$). También, con el criterio EHNF se rechazan algunas de las hipótesis nulas falsas que no se detectan con la regla bayes, por tener una probabilidad final estimada de ser ciertas mayor que 0.5, obteniéndose de esta manera un porcentaje de falsos negativos más bajo que el que se obtiene con la regla bayes.

Tabla 2.14: Resultados aplicando la regla bayes y el criterio de decisión EHNF, para distintos valores de los parámetros de la distribución inicial de $p \sim \text{Beta}(\alpha, \beta)$ y de $\phi \sim \text{Gamma}(a/2, b/2)$, utilizando los datos generados mediante una mixtura de $N(0, 1/\sqrt{0.5})$ con probabilidad $p = 0.9$ y de $N(1, 1/\sqrt{0.5})$ con probabilidad $1 - p = 0.1$, para $N = 500$.

(a, b)	(α, β)	$R_C\%$	$FP_C\%$	$FN_C\%$	$\widehat{N}_1\%$	$prob_{\widehat{N}_1}$	$FP_{EHNF}\%$	$FN_{EHNF}\%$
(0,0)	(1, 0.5)	5.40	2.6374	66.6667	12.20	0.6785	8.1319	46.6667
	(1, 1)	5.60	2.6374	64.4444	12.40	0.6799	8.1319	44.4444
	(0.5, 1)	5.80	2.8571	64.4444	12.60	0.6881	8.3516	44.4444
(1,0.5)	(1, 0.5)	5.60	2.6374	64.4444	12.20	0.6823	8.1319	46.6667
	(1, 1)	5.60	2.6374	64.4444	12.60	0.6908	8.5714	44.4444
	(0.5, 1)	5.80	2.8571	64.4444	12.60	0.6876	8.3516	44.4444
(10,1)	(1, 0.5)	5.40	2.4176	64.4444	12.40	0.6794	8.1319	44.4444
	(1, 1)	5.80	2.8571	64.4444	12.60	0.6903	8.3516	44.4444
	(0.5, 1)	5.80	2.8571	64.4444	12.60	0.6902	8.3516	44.4444

En las tablas 2.14 y 2.15, para $N = 500$ y para $N = 10000$, respectivamente, se utilizaron los datos generados mediante la mixtura de $N(0, 1/\sqrt{0.5})$ con probabilidad $p = 0.9$ y de $N(1, 1/\sqrt{0.5})$ con probabilidad $1 - p = 0.1$; es decir, los datos generados con mayor varianza. Se puede observar que el procedimiento es también robusto respecto a los parámetros considerados.

Asimismo, es importante señalar que se observa una notable reducción en el porcentaje de hipótesis nulas rechazadas mediante la regla bayes, con respecto al obtenido utilizando los datos generados por la mixtura con menor varianza, hecho que no ocurre con el porcentaje estimado de hipótesis nulas falsas obtenido con el criterio de decisión EHNF.

También se puede ver que con el criterio de decisión EHNF se reduce el porcentaje

Tabla 2.15: Resultados aplicando la regla bayes y el criterio de decisión EHNF, para distintos valores de los parámetros de la distribución inicial de $p \sim Beta(\alpha, \beta)$ y de $\phi \sim Gamma(a/2, b/2)$, utilizando los datos generados mediante una mixtura de $N(0, 1/\sqrt{0.5})$ con probabilidad $p = 0.9$ y de $N(1, 1/\sqrt{0.5})$ con probabilidad $1 - p = 0.1$, para $N = 10000$.

(a, b)	(α, β)	$R_C\%$	$FP_C\%$	$FN_C\%$	$\widehat{N}_1\%$	$prob_{\widehat{N}_1}$	$FP_{EHNF}\%$	$FN_{EHNF}\%$
(0,0)	(1, 0.5)	3.88	1.1868	71.4431	9.74	0.7154	5.2462	49.0854
	(1, 1)	3.87	1.1757	71.4431	9.75	0.7151	5.2684	49.1870
	(0.5, 1)	3.86	1.2090	71.8496	9.81	0.7148	5.3017	48.8821
(1,0.5)	(1, 0.5)	3.89	1.1979	71.4431	9.77	0.7161	5.2684	48.9837
	(1, 1)	3.89	1.1868	71.3415	9.80	0.7140	5.3128	49.0854
	(0.5, 1)	3.84	1.1757	71.7480	9.77	0.7149	5.2795	48.9837
(10,1)	(1, 0.5)	3.85	1.1757	71.6463	9.76	0.7154	5.3017	49.2886
	(1, 1)	3.87	1.1868	71.5447	9.78	0.7148	5.3239	49.3902
	(0.5, 1)	3.89	1.1979	71.4431	9.77	0.7152	5.3017	49.2886

de falsos negativos, manteniendo aceptable ($< 10\%$) el porcentaje de falsos positivos, si bien se observa un aumento del porcentaje de falsos negativos con respecto al obtenido utilizando los datos generados por la mixtura con menor varianza, siendo este aumento mayor para el porcentaje de falsos negativos obtenidos con la regla bayes.

Por tanto, con el criterio EHNF se detectan más hipótesis nulas falsas que con la regla bayes, obteniéndose un menor porcentaje de falsos negativos, principalmente cuando los datos se generan con mayor varianza, a la vez que el porcentaje de falsos positivos se mantiene en unos niveles aceptables, consiguiendo de este modo un equilibrio mayor entre los porcentajes de falsos positivos y falsos negativos que el que se alcanza con la regla bayes, por lo que podemos concluir que el criterio de decisión EHNF resulta más apropiado que la regla bayes, cuando para aplicar ésta

se elige $C_{0i} = C_{1i}$ para $i = 1, \dots, N$, siendo especialmente conveniente cuando los datos se generan con mayor varianza.

En esta subsección también se aplican, a los datos generados mediante una mezcla de $N(0, 1/\sqrt{0.5})$ con probabilidad $p = 0.9$ y de $N(1, 1/\sqrt{0.5})$ con probabilidad $1-p = 0.1$, el procedimiento de Benjamini y Hochberg (1995) para controlar el FDR, no rechazándose ninguna hipótesis nula, tanto para $N = 500$ como para $N = 10000$, a los niveles $\alpha = 0.05$ y $\alpha = 0.1$. Por tanto el procedimiento de Benjamini y Hochberg (1995), con los niveles habitualmente utilizados para controlar el FDR, es excesivamente conservador cuando los datos se generan con mayor varianza, mientras que con el procedimiento bayesiano utilizando cualquiera de los dos criterios de decisión, la regla bayes o el criterio EHNF, principalmente con este último, se obtienen resultados más satisfactorios.

Es importante señalar que para obtener, con los datos simulados, resultados similares con los procedimientos frecuentista y bayesiano aplicando la regla bayes, es necesario controlar el FDR a un nivel $\alpha > 0.4$, nivel que obviamente no es admisible. Además debe tenerse en cuenta que la regla bayes es menos adecuada que el criterio de decisión EHNF, sobre todo para los datos generados de la mezcla con mayor varianza.

2.5 Conclusiones

En primer lugar, debemos señalar que el procedimiento bayesiano es robusto respecto a la elección de los parámetros de la distribución inicial $Gamma(a/2, b/2)$ de ϕ . También es robusto con respecto a los parámetros de la distribución inicial $Beta(\alpha, \beta)$ de p , siempre y cuando ésta no suponga una información demasiado

sesgada, como en el ejemplo simulado en la subsección 2.3.2 con la distribución $Beta(1, 25)$ que presenta un sesgo muy fuerte hacia la izquierda, mientras que los datos se generaron de mixturas con probabilidad $p = 0.9$ y $p = 0.7$. Sin embargo, no parece que tenga gran influencia en qué zona se concentra la densidad. Además, las estimaciones que se obtienen de los parámetros p y ϕ del modelo están muy próximas a los valores de los parámetros con los que se generaron los datos.

Cabe destacar también que el criterio de decisión EHNF propuesto por nosotros es menos conservador que la regla bayes, cuando se toma $C_{0i} = C_{1i}$ para $i = 1, \dots, N$, en el sentido de que se rechazan más hipótesis nulas. Además, con el criterio EHNF se consigue un equilibrio mayor entre los porcentajes de falsos positivos y falsos negativos que el que se alcanza con la regla bayes, puesto que se obtiene un porcentaje más bajo de falsos negativos. Esto es debido a que con el criterio EHNF se detectan más hipótesis nulas falsas, mientras que el porcentaje de falsos positivos se mantiene en unos niveles aceptables ($< 10\%$). Esta tendencia se acentúa cuando los datos se generan con mayor varianza, por lo que podemos concluir que el criterio de decisión EHNF es más adecuado que la regla bayes cuando se toma $C_{0i} = C_{1i}$ para $i = 1, \dots, N$, siendo especialmente conveniente cuando los datos se generan con mayor varianza, evitando además la necesidad de fijar estas cantidades C_{0i} y C_{1i} .

Uno de los problemas de los procedimientos de contraste de hipótesis múltiples, desde una perspectiva frecuentista, es que tienden a detectar un bajo número de hipótesis nulas falsas. En este sentido, otra conclusión importante que se puede extraer de los resultados obtenidos con los datos simulados es que el procedimiento bayesiano, con ambos criterios de decisión, es menos conservador que el procedimiento de Benjamini y Hochberg (1995), siendo el porcentaje de hipótesis nulas rechazadas con el procedimiento bayesiano, principalmente con el criterio EHNF,

más ajustado a la realidad que con el procedimiento frecuentista.

Además, el procedimiento bayesiano, con cualquiera de los dos criterios de decisión, sobre todo con el criterio de decisión EHNF, detecta más hipótesis nulas falsas reduciendo significativamente el porcentaje de falsos negativos, mientras que los porcentajes de falsos positivos se mantienen en unos niveles admisibles. Estas características se acentúan cuando los datos se generan con mayor varianza.

Por otro lado, es importante señalar que el coste computacional del procedimiento bayesiano es muy bajo. Por tanto, podemos concluir que el procedimiento bayesiano, con cualquiera de los dos criterios de decisión, principalmente con el criterio EHNF, es más conveniente que el procedimiento frecuentista de Benjamini y Hochberg (1995), sobre todo cuando los datos se generan con mayor varianza.

Como se comentó al principio del capítulo, los ejemplos en los que tanto la hipótesis nula como la alternativa son simples resultan, en general, poco realistas pero sirven como introducción sencilla para problemas más generales. En el siguiente capítulo, se plantea el problema de los contrastes múltiples cuando la hipótesis nula es simple y la alternativa bilateral.

Apéndice 2A: Modelo Gaussiano. Varianza desconocida. Distribuciones finales condicionadas

Denotamos por $\mathbf{z}_{-i} = (z_1, \dots, z_{i-1}, z_{i+1}, \dots, z_N)$ y $\mathbf{t}_{-i} = (t_1, \dots, t_{i-1}, t_{i+1}, \dots, t_N)$, donde $t_i = (x_{i1}, \dots, x_{in})$.

- *Distribución final condicionada de $z_i = 0$, dados los datos y el resto de parámetros*

$$\begin{aligned}
Pr(z_i = 0 | \mathbf{t}, \mathbf{z}_{-i}, \boldsymbol{\theta}) &= \frac{\pi(\boldsymbol{\theta}) f(\mathbf{t}, z_i = 0, \mathbf{z}_{-i} | \boldsymbol{\theta})}{\sum_{k=0}^1 \pi(\boldsymbol{\theta}) f(\mathbf{t}, z_i = k, \mathbf{z}_{-i} | \boldsymbol{\theta})} \\
&= \frac{\pi(\boldsymbol{\theta}) Pr(z_i = 0 | p) f(t_i | z_i = 0, \phi) f(\mathbf{t}_{-i}, \mathbf{z}_{-i} | \boldsymbol{\theta})}{\sum_{k=0}^1 \pi(\boldsymbol{\theta}) Pr(z_i = k | p) f(t_i | z_i = k, \phi) f(\mathbf{t}_{-i}, \mathbf{z}_{-i} | \boldsymbol{\theta})} \\
&= \frac{Pr(z_i = 0 | p) f(t_i | z_i = 0, \phi)}{Pr(z_i = 0 | p) f(t_i | z_i = 0, \phi) + Pr(z_i = 1 | p) f(t_i | z_i = 1, \phi)} \\
&= \frac{Pr(z_i = 0 | p) \prod_{j=1}^n f_0(x_{ij} | \phi)}{Pr(z_i = 0 | p) \prod_{j=1}^n f_0(x_{ij} | \phi) + Pr(z_i = 1 | p) \prod_{j=1}^n f_1(x_{ij} | \phi)} \\
&= \frac{p \prod_{j=1}^n \left(\phi^{\frac{1}{2}} (2\pi)^{-\frac{1}{2}} e^{-\frac{\phi}{2}(x_{ij} - \mu_0)^2} \right)}{p \prod_{j=1}^n \left(\phi^{\frac{1}{2}} (2\pi)^{-\frac{1}{2}} e^{-\frac{\phi}{2}(x_{ij} - \mu_0)^2} \right) + (1-p) \prod_{j=1}^n \left(\phi^{\frac{1}{2}} (2\pi)^{-\frac{1}{2}} e^{-\frac{\phi}{2}(x_{ij} - \mu_1)^2} \right)} \\
&= \frac{p \phi^{\frac{n}{2}} (2\pi)^{-\frac{n}{2}} e^{-\frac{\phi}{2} \sum_{j=1}^n (x_{ij} - \mu_0)^2}}{\phi^{\frac{n}{2}} (2\pi)^{-\frac{n}{2}} \left(p e^{-\frac{\phi}{2} \sum_{j=1}^n (x_{ij} - \mu_0)^2} + (1-p) e^{-\frac{\phi}{2} \sum_{j=1}^n (x_{ij} - \mu_1)^2} \right)}
\end{aligned}$$

$$= \left(1 + \frac{(1-p)u_{1i}}{p u_{0i}} \right)^{-1},$$

donde $u_{0i} = e^{-(\phi n/2)(\mu_0^2 - 2\mu_0 \bar{x}_i)}$ y $u_{1i} = e^{-(\phi n/2)(\mu_1^2 - 2\mu_1 \bar{x}_i)}$ son las expresiones definidas en (2.7).

- *Distribución final condicionada de p , dados los datos y el resto de parámetros*

$$\begin{aligned} f(p|\mathbf{t}, \mathbf{z}, \phi) &= \frac{\pi(\boldsymbol{\theta})f(\mathbf{t}, \mathbf{z}|\boldsymbol{\theta})}{\int_0^1 \pi(\boldsymbol{\theta})f(\mathbf{t}, \mathbf{z}|\boldsymbol{\theta}) dp} \\ &\propto \pi(\boldsymbol{\theta}) p^{n_0} (1-p)^{n_1} \prod_{i:z_i=0} f(t_i|z_i=0, \phi) \prod_{i:z_i=1} f(t_i|z_i=1, \phi) \\ &= \pi(\boldsymbol{\theta}) p^{n_0} (1-p)^{n_1} \prod_{i:z_i=0} \prod_{j=1}^n f_0(x_{ij}|\phi) \prod_{i:z_i=1} \prod_{j=1}^n f_1(x_{ij}|\phi) \\ &\propto \pi(p|\alpha, \beta) p^{n_0} (1-p)^{n_1} \\ &\propto p^{\alpha-1} (1-p)^{\beta-1} p^{n_0} (1-p)^{n_1} \sim \text{Beta}(n_0 + \alpha, n_1 + \beta). \end{aligned}$$

- *Distribución final condicionada de ϕ , dados los datos y el resto de parámetros*

$$\begin{aligned} f(\phi|\mathbf{t}, \mathbf{z}, p) &= \frac{\pi(\boldsymbol{\theta})f(\mathbf{t}, \mathbf{z}|\boldsymbol{\theta})}{\int_0^\infty \pi(\boldsymbol{\theta})f(\mathbf{t}, \mathbf{z}|\boldsymbol{\theta}) d\phi} \\ &\propto \pi(\boldsymbol{\theta}) p^{n_0} (1-p)^{n_1} \prod_{i:z_i=0} f(t_i|z_i=0, \phi) \prod_{i:z_i=1} f(t_i|z_i=1, \phi) \\ &= \pi(\boldsymbol{\theta}) p^{n_0} (1-p)^{n_1} \prod_{i:z_i=0} \prod_{j=1}^n f_0(x_{ij}|\phi) \prod_{i:z_i=1} \prod_{j=1}^n f_1(x_{ij}|\phi) \\ &\propto \pi(\phi|a, b) \prod_{i:z_i=0} \prod_{j=1}^n f_0(x_{ij}|\phi) \prod_{i:z_i=1} \prod_{j=1}^n f_1(x_{ij}|\phi) \end{aligned}$$

$$\begin{aligned}
&\propto \phi^{\frac{a}{2}-1} e^{-\frac{b}{2}\phi} \prod_{i:z_i=0} \prod_{j=1}^n \left(\phi^{\frac{1}{2}} e^{-\frac{\phi}{2}(x_{ij}-\mu_0)^2} \right) \prod_{i:z_i=1} \left(\phi^{\frac{1}{2}} e^{-\frac{\phi}{2}(x_{ij}-\mu_1)^2} \right) \\
&= \phi^{\frac{a}{2}-1} e^{-\frac{b}{2}\phi} \prod_{i:z_i=0} \left(\phi^{\frac{n}{2}} e^{-\frac{\phi}{2} \sum_{j=1}^n (x_{ij}-\mu_0)^2} \right) \prod_{i:z_i=1} \left(\phi^{\frac{n}{2}} e^{-\frac{\phi}{2} \sum_{j=1}^n (x_{ij}-\mu_1)^2} \right) \\
&= \phi^{\frac{a+Nn}{2}-1} e^{-\frac{\phi}{2} \left(b + \sum_{i=1}^N \sum_{j=1}^n x_{ij}^2 + n \sum_{i:z_i=0} (\mu_0^2 - 2\mu_0 \bar{x}_i) + n \sum_{i:z_i=1} (\mu_1^2 - 2\mu_1 \bar{x}_i) \right)} \\
&\sim \text{Gamma} \left(\frac{a + Nn}{2}, \frac{R}{2} \right),
\end{aligned}$$

donde $R = b + \sum_{i=1}^N \sum_{j=1}^n x_{ij}^2 + n \sum_{i:z_i=0} (\mu_0^2 - 2\mu_0 \bar{x}_i) + n \sum_{i:z_i=1} (\mu_1^2 - 2\mu_1 \bar{x}_i)$.

Capítulo 3

Contraste de hipótesis bilaterales.

Modelo Empírico Bayes

En este capítulo se aborda, desde una perspectiva bayesiana, el problema de los contrastes múltiples cuando la hipótesis nula es simple y la alternativa es bilateral. Como hicimos en el capítulo 2, se considera como modelo de probabilidad una mezcla de las densidades bajo las hipótesis nula y alternativa. Sin embargo, en este capítulo, suponemos que para cada hipótesis se observa el valor de un estadístico, en lugar del vector de observaciones. La metodología que se utiliza es similar a la utilizada en el capítulo precedente, es decir, la inferencia bayesiana se lleva a cabo utilizando los métodos de cadenas de Markov Monte Carlo (MCMC); igualmente se aplican los criterios de decisión descritos en la sección 2.2 del capítulo anterior, esto es, la regla bayes definida en (2.3) y el criterio basado en la estimación del número de hipótesis nulas falsas (EHNF), que es el criterio que se propone en este trabajo. En este capítulo, además, se considera una aproximación Empírico Bayes para el modelo Gaussiano, con la finalidad de estimar el parámetro asociado a la varianza

inicial de las medias del modelo, con el fin de evitar así la sensibilidad existente a la elección de dicho parámetro.

3.1 Introducción

Los contrastes en los que la hipótesis nula es simple y la alternativa es bilateral resultan especialmente interesantes, ya que son muchas las situaciones, en cualquier campo de la investigación, en las que se plantea la necesidad de estudiar si existen diferencias significativas cuando se mide una determinada variable en dos condiciones distintas.

Un ejemplo muy actual al que se ajusta este tipo de contrastes lo constituyen los estudios realizados con biochips de ADN, en los que uno de los objetivos principales es la identificación de genes con expresión diferencial; es decir, estudiar si para cada gen existen diferencias significativas entre el nivel de expresión en tejido sano y el nivel de expresión en tejido enfermo. Esta cuestión puede plantearse como un problema de contrastes múltiples con hipótesis nulas simples y alternativas bilaterales, en el que cada test individual estaría referido a un gen, contrastándose la hipótesis nula de igualdad de medias entre tejido normal y tejido enfermo o, equivalentemente, que la diferencia de medias es cero frente a la hipótesis alternativa en la que la diferencia de medias es distinta de cero.

Algunos de los autores que tratan este tipo de contrastes múltiples en el contexto de los biochips de ADN son, entre otros, y desde un punto de vista frecuentista Dudoit et al. (2003), Storey et al. (2007) y Storey (2007), este último trabajo con aplicaciones en contextos más generales. Desde un punto de vista bayesiano Ibrahim et al. (2002), Lönnstedt y Speed (2002), Lönnstedt y Britton (2005), Gottardo et

al. (2006), Scott y Berger (2006) y Cabras (2010).

Aunque en general es obvio que es preferible utilizar el vector de observaciones a utilizar un estadístico, en este capítulo, con el fin de aplicar el procedimiento que se describe en el mismo a datos procedentes de experimentos con biochips de ADN, supondremos que para cada hipótesis se observa el valor de un estadístico; sin embargo, para otros problemas el procedimiento también se podría desarrollar, al igual que en el capítulo anterior, utilizando el vector de observaciones.

El objetivo en este capítulo es proporcionar un marco general para el problema de los tests de hipótesis múltiples cuando, para cada contraste, la hipótesis nula es simple y la alternativa bilateral, así como comprobar que también resulta más apropiado utilizar en este caso el criterio de decisión EHNF propuesto por nosotros, que utilizar la regla bayes (2.3) definida en la sección 2.2 del capítulo 2.

Este capítulo está organizado en seis secciones más la presente introducción. En la sección 3.2, se plantea el problema de manera general, asumiendo que los estadísticos correspondientes a cada hipótesis son variables aleatorias i.i.d., mediante una mixtura de las densidades bajo las hipótesis nula y alternativa.

En la sección 3.3, se aplica el procedimiento descrito en la sección precedente, considerando para los estadísticos una distribución normal con varianza desconocida y siendo las medias los parámetros de interés. En esta sección, mediante un ejemplo de simulación, también se realiza un análisis de sensibilidad a la elección de los parámetros de la distribución inicial.

En la sección 3.4, con el fin de evitar la sensibilidad a la elección del parámetro asociado a la varianza inicial de las medias del modelo, se propone un modelo jerárquico; asimismo se ilustra el procedimiento y se realiza un análisis de sensi-

bilidad utilizando los datos simulados en la sección anterior.

Como los resultados obtenidos en la sección 3.4 muestran que, con el modelo jerárquico, no se resuelve el problema de sensibilidad detectada respecto al parámetro asociado a la varianza de la distribución inicial, en la sección 3.5 proponemos una aproximación Empírico Bayes con el objetivo de estimar dicho parámetro; igualmente se ilustra el procedimiento utilizando los mismos datos que en las secciones precedentes. Los resultados obtenidos en la sección 3.5 muestran que la aproximación Empírico Bayes resuelve de manera adecuada el problema que existía de sensibilidad. Por último, se comparan los resultados con los obtenidos mediante el procedimiento de Benjamini y Hochberg (1995) para controlar el FDR.

En la sección 3.6, se aplica el procedimiento que se propone a datos reales procedentes de experimentos con biochips de ADN. Todos los cálculos de este capítulo se han realizado utilizando MATLAB. Finalmente, en la sección 3.7, se obtienen las conclusiones y las posibles extensiones del capítulo.

3.2 Formulación del problema

En esta sección se plantea, de un modo general, el problema de los contrastes de hipótesis múltiples cuando la hipótesis nula es simple e idéntica para todas las hipótesis y la alternativa es bilateral.

Consideramos el siguiente problema de contrastes múltiples,

$$H_{0i} : \theta_i = \theta_0 \quad \text{frente a} \quad H_{1i} : \theta_i \neq \theta_0, \quad i = 1, \dots, N, \quad (3.1)$$

donde θ_0 es un valor conocido y común para todas las hipótesis. Si los tests se

plantean con θ_0 distinto para cada hipótesis, es decir, si

$$H_{0i} : \theta_i = \theta_{0i} \quad \text{frente a} \quad H_{1i} : \theta_i \neq \theta_{0i}, \quad i = 1, \dots, N,$$

basta tomar $\theta'_i = \theta_i - \theta_{0i}$, para $i = 1, \dots, N$, y el problema queda

$$H_{0i} : \theta'_i = 0 \quad \text{frente a} \quad H_{1i} : \theta'_i \neq 0, \quad i = 1, \dots, N,$$

que equivale al propuesto en (3.1).

Aunque el tratamiento que se dará para este tipo de contrastes es análogo al utilizado en el caso de hipótesis simples, a diferencia de éste supondremos, como se comentó en la sección previa, que se observa un estadístico para cada hipótesis en lugar del vector de observaciones. Por otro lado hay que tener en cuenta que, para el problema definido en (3.1), el modelo es obviamente más complejo ya que depende de N parámetros más que en el caso de hipótesis simples, éstos son $\theta_1, \dots, \theta_N$, para los que será necesario definir una distribución inicial con objeto de realizar la inferencia bayesiana. A continuación se expone la metodología para el tratamiento del problema de contrastes múltiples (3.1).

Suponemos que se observa un estadístico $T_i = T_i(X_{i1}, \dots, X_{in})$ para cada hipótesis H_{0i} y que las T_i , para $i = 1, \dots, N$, son variables aleatorias independientes e idénticamente distribuidas con densidad $f_0(t_i|\phi)$ bajo la hipótesis nula y $f_1(t_i|\phi, \theta_i)$, para algún $\theta_i \neq \theta_0$, bajo la hipótesis alternativa, siendo θ_i el parámetro de interés y ϕ un parámetro perturbador que suponemos, por simplicidad, el mismo para todas las hipótesis.

De la misma forma que en el caso de hipótesis simples, suponemos que hay una probabilidad inicial común p de que $\theta_i = \theta_0$, para todo $i = 1, \dots, N$. De esta manera podemos considerar que p es la proporción desconocida de hipótesis nulas ciertas. Entonces, denotando por $H_{0i} = 0$ si H_{0i} es cierta y por $H_{0i} = 1$ si H_{0i} es

falsa, podemos considerar los estadísticos T_i procedentes de una mezcla de ambas densidades,

$$\begin{aligned} f(t_i|p, \phi, \theta_i) &= Pr(H_{0i} = 0|p)f(t_i|H_{0i} = 0) + Pr(H_{0i} = 1|p)f(t_i|H_{0i} = 1) \\ &= pf_0(t_i|\phi) + (1-p)f_1(t_i|\phi, \theta_i), \end{aligned} \quad (3.2)$$

siendo $p = Pr(H_{0i} = 0|p)$ y $t_i = T_i(x_{i1}, \dots, x_{in})$ una realización particular del estadístico T_i para $i = 1, \dots, N$.

La verosimilitud se puede escribir como

$$f(\mathbf{t}|\boldsymbol{\theta}) = \prod_{i=1}^N f(t_i|p, \phi, \theta_i) = \prod_{i=1}^N [pf_0(t_i|\phi) + (1-p)f_1(t_i|\phi, \theta_i)],$$

donde $\boldsymbol{\theta} = (p, \phi, \theta_1, \dots, \theta_N)$ es el vector de parámetros y $\mathbf{t} = (t_1, \dots, t_N)$ es el vector de estadísticos .

Con el fin de llevar a cabo la inferencia bayesiana, necesitamos también definir una distribución inicial, $\pi(\boldsymbol{\theta})$, para el conjunto de parámetros del modelo. Entonces, la probabilidad final de cada hipótesis nula tiene la siguiente expresión,

$$\begin{aligned} Pr(H_{0i} = 0|\mathbf{t}) &= \frac{\int pf_0(t_i|\phi) \prod_{j \neq i} [pf_0(t_j|\phi) + (1-p)f_1(t_j|\phi, \theta_j)] \pi(\boldsymbol{\theta}) d\boldsymbol{\theta}}{\int f(\mathbf{t}|\boldsymbol{\theta}) \pi(\boldsymbol{\theta}) d\boldsymbol{\theta}} \\ &= \frac{\int g(t_i|p, \phi, \theta_i) f(\mathbf{t}|\boldsymbol{\theta}) \pi(\boldsymbol{\theta}) d\boldsymbol{\theta}}{\int f(\mathbf{t}|\boldsymbol{\theta}) \pi(\boldsymbol{\theta}) d\boldsymbol{\theta}} \\ &= E_{\pi(\boldsymbol{\theta}|\mathbf{t})} [g(t_i|p, \phi, \theta_i)], \end{aligned}$$

donde $g(t_i|p, \phi, \theta_i) = \frac{pf_0(t_i|\phi)}{pf_0(t_i|\phi) + (1-p)f_1(t_i|\phi, \theta_i)}$.

Como se comentó en el capítulo anterior, esta integral no siempre se puede obtener de forma analítica, ni puede ser aproximada siempre por el método de

Monte Carlo, ya que la distribución final $\pi(\boldsymbol{\theta}|\mathbf{t})$ puede no ser conocida. Tampoco se puede aplicar el método de Monte Carlo cuando la distribución inicial es impropia y se considera, como en (2.2), la probabilidad final $Pr(H_{0i} = 0|\mathbf{t})$ como un cociente de esperanzas con respecto a la distribución inicial. Sin embargo, la inferencia bayesiana se puede llevar a cabo utilizando los métodos MCMC como se describe en la subsección 2.4.1 del capítulo anterior.

Primero, como suele hacerse en mixturas para simplificar la verosimilitud, consideramos, nuevamente, N variables latentes i.i.d., Z_1, \dots, Z_N , cada una de ellas asociada a un estadístico T_i , indicando la componente de la mixtura (3.2) a la que pertenece el correspondiente estadístico, y con distribución inicial

$$Pr(Z_i = 0|p) = p \quad \text{y} \quad Pr(Z_i = 1|p) = 1 - p, \quad i = 1, \dots, N,$$

donde $Z_i = 0$ indica que T_i procede de la primera componente de la mixtura y $Z_i = 1$ indica que T_i procede de la segunda componente, siendo Z_i y θ_j independientes para todo $i, j = 1, \dots, N$.

Así, una vez observado el valor de las variables latentes Z_i , las T_i son variables aleatorias independientes distribuidas según la densidad de la correspondiente componente de la mixtura, es decir, para $i = 1, \dots, N$, $f(t_i|z_i = 0, \phi, \theta_i) = f_0(t_i|\phi)$ y $f(t_i|z_i = 1, \phi, \theta_i) = f_1(t_i|\phi, \theta_i)$, para algún $\theta_i \neq \theta_0$.

De este modo la verosimilitud se puede escribir como

$$\begin{aligned} f(\mathbf{t}, \mathbf{z}|\boldsymbol{\theta}) &= \prod_{i=1}^N f(t_i, z_i|\boldsymbol{\theta}) \\ &= \prod_{i:z_i=0} Pr(z_i = 0|p) f(t_i|z_i = 0, \phi, \theta_i) \prod_{i:z_i=1} Pr(z_i = 1|p) f(t_i|z_i = 1, \phi, \theta_i) \\ &= \prod_{i:z_i=0} p f_0(t_i|\phi) \prod_{i:z_i=1} (1-p) f_1(t_i|\phi, \theta_i) \\ &= p^{n_0} (1-p)^{n_1} \prod_{i:z_i=0} f_0(t_i|\phi) \prod_{i:z_i=1} f_1(t_i|\phi, \theta_i), \end{aligned} \tag{3.3}$$

donde $\mathbf{t} = (t_1, \dots, t_N)$, $\mathbf{z} = (z_1, \dots, z_N)$, $n_0 = \sum_i I(z_i = 0)$ es el número de observaciones de la primera componente de la mezcla y $n_1 = \sum_i I(z_i = 1)$ el número de observaciones de la segunda componente, siendo $N = n_0 + n_1$.

La distribución final de $(\boldsymbol{\theta}, \mathbf{z})$ viene dada por $\pi(\boldsymbol{\theta}, \mathbf{z}|\mathbf{t}) \propto f(\mathbf{t}, \mathbf{z}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})$. Entonces, dada la distribución inicial $\pi(\boldsymbol{\theta})$ y la verosimilitud (3.3), se puede construir una cadena de Markov, utilizando los métodos MCMC, que tenga como distribución estacionaria $\pi(\boldsymbol{\theta}, \mathbf{z}|\mathbf{t})$.

De esta manera, dada una muestra de la cadena de Markov en equilibrio,

$$\{(\boldsymbol{\theta}^{(j)}, \mathbf{z}^{(j)}) : j = 1, \dots, M\},$$

donde $\boldsymbol{\theta}^{(j)} = (p^{(j)}, \phi^{(j)}, \theta_1^{(j)}, \dots, \theta_N^{(j)})$ y $\mathbf{z}^{(j)} = (z_1^{(j)}, \dots, z_N^{(j)})$, se puede estimar, como se indica en (2.19), la media final de cada parámetro a través de la correspondiente media muestral. Así, podemos estimar p y ϕ como en (2.24) y (2.25), es decir,

$$\hat{p} = E[p|\mathbf{t}] \approx \frac{1}{M} \sum_{j=1}^M p^{(j)} \quad (3.4)$$

y

$$\hat{\phi} = E[\phi|\mathbf{t}] \approx \frac{1}{M} \sum_{j=1}^M \phi^{(j)}. \quad (3.5)$$

También podemos estimar, para cada $i = 1, \dots, N$, el parámetro θ_i de la forma siguiente,

$$\hat{\theta}_i = E[\theta_i|\mathbf{t}] \approx \frac{1}{M} \sum_{j=1}^M \theta_i^{(j)}, \quad (3.6)$$

y de forma análoga a (2.26), podemos aproximar la probabilidad final de cada hipótesis nula como

$$\begin{aligned} Pr(H_{0i} = 0|\mathbf{t}) &= 1 - Pr(H_{0i} = 1|\mathbf{t}) \\ &= 1 - Pr(\theta_i \neq \theta_0|\mathbf{t}) \end{aligned}$$

$$\begin{aligned}
&= 1 - Pr(z_i = 1|\mathbf{t}) \\
&\approx 1 - \frac{1}{M} \sum_{j=1}^M I(z_i^{(j)} = 1),
\end{aligned} \tag{3.7}$$

para $i = 1, \dots, N$.

Una vez se ha obtenido \hat{p} y una estimación de la probabilidad final de cada hipótesis nula, $\hat{Pr}(H_{0i} = 0|\mathbf{t})$, como se indica en (3.4) y (3.7), respectivamente, podemos aplicar los criterios de decisión descritos en la sección 2.2 del capítulo anterior, esto es, la regla bayes definida en (2.3): para cada \mathbf{t} , rechazar todas las hipótesis nulas H_{0i} tales que

$$\hat{Pr}(H_{0i} = 0|\mathbf{t}) \leq \frac{C_{0i}}{C_{0i} + C_{1i}} \tag{3.8}$$

y aceptar el resto, dados los costes C_{0i} y C_{1i} , y el criterio de decisión EHNF: rechazar las $\widehat{N}_1 = N(1 - \hat{p})$ hipótesis nulas con menor probabilidad final estimada y aceptar el resto.

3.3 Modelo Gaussiano

En esta sección se aplica la metodología descrita en la sección previa, asumiendo para los estadísticos una distribución normal con varianza desconocida y considerando las medias como parámetros de interés.

Si bien el procedimiento es similar al utilizado en la subsección 2.4.1, para el caso de hipótesis simples bajo el modelo normal con varianza desconocida, en esta sección es necesario definir además una distribución inicial para las medias del modelo resultando, como se verá posteriormente en la subsección 3.3.2, que el procedimiento no es robusto respecto del parámetro asociado a la varianza de la distribución inicial considerada.

Este problema de falta de robustez se trata en la sección 3.4, mediante la utilización de un modelo jerárquico, y en la sección 3.5, mediante una aproximación Empírico Bayes, ya que el modelo jerárquico no resuelve el problema de sensibilidad a la elección de dicho parámetro.

Consideramos el siguiente problema de contraste de hipótesis múltiples,

$$H_{0i} : \mu_i = 0 \quad \text{frente a} \quad H_{1i} : \mu_i \neq 0 \quad i = 1, \dots, N. \quad (3.9)$$

Este tipo de contrastes múltiples, bajo el modelo normal, se utiliza con mucha frecuencia en el contexto de los experimentos con biochips de ADN, véase Dudoit et al. (2003), Lönnstedt y Speed (2002), Lönnstedt y Britton (2005), Scott y Berger (2006) y Storey et al. (2007) entre otros.

Suponemos que para cada hipótesis se observa el estadístico media muestral, $T_i = \sum_{j=1}^n X_{ij}/n$, y que las T_i , para $i = 1, \dots, N$, son variables aleatorias independientes e idénticamente distribuidas $N(0, 1/\sqrt{\phi})$ bajo la hipótesis nula y $N(\mu_i, 1/\sqrt{\phi})$, donde $\mu_i \neq 0$, bajo la hipótesis alternativa, siendo ϕ la precisión que suponemos desconocida y la misma, por simplicidad, para todas las hipótesis.

Consideramos para cada T_i una variable latente Z_i , definida como en la sección anterior. De modo que, bajo el modelo normal, la verosimilitud (3.3) para el parámetro $\boldsymbol{\theta} = (p, \phi, \mu_1, \dots, \mu_N)$ se puede expresar como

$$\begin{aligned} f(\mathbf{t}, \mathbf{z}|\boldsymbol{\theta}) &\propto p^{n_0}(1-p)^{n_1} \prod_{i:z_i=0} \left(\phi^{\frac{1}{2}} e^{-\frac{\phi}{2} t_i^2} \right) \prod_{i:z_i=1} \left(\phi^{\frac{1}{2}} e^{-\frac{\phi}{2} (t_i - \mu_i)^2} \right) \\ &= p^{n_0}(1-p)^{n_1} \phi^{\frac{N}{2}} e^{-\frac{\phi}{2} \sum_{i:z_i=0} t_i^2} e^{-\frac{\phi}{2} \sum_{i:z_i=1} (t_i - \mu_i)^2}. \end{aligned} \quad (3.10)$$

Para la distribución inicial $\pi(\boldsymbol{\theta})$ consideramos las siguientes distribuciones iniciales,

$$p \sim \text{Beta}(\alpha, \beta),$$

$$\begin{aligned}\phi &\sim \text{Gamma}(a/2, b/2), \\ \mu_i|\phi &\sim N\left(0, \frac{1}{\sqrt{c_i\phi}}\right), \quad i = 1, \dots, N,\end{aligned}\tag{3.11}$$

es decir, las distribuciones iniciales naturales conjugadas, véase Gelman et al. (2004). Además, asumir inicialmente que μ_i y ϕ son dependientes resulta muy adecuado en el contexto de los experimentos con biochips de ADN, ya que esta dependencia suele ser muy frecuente en este tipo de datos.

Entonces, teniendo en cuenta las distribuciones iniciales definidas en (3.11) y la verosimilitud (3.10), a continuación se describen las distribuciones finales de cada uno de los parámetros del modelo condicionadas al resto de parámetros. Estas distribuciones nos permitirán generar valores de una cadena MCMC, construida según el muestreo de Gibbs, cuya distribución estacionaria es la distribución final $\pi(\boldsymbol{\theta}, \mathbf{z}|\mathbf{t}, \alpha, \beta, a, b, \mathbf{c})$, donde $\mathbf{c} = (c_1, \dots, c_N)$.

Se puede comprobar fácilmente que la probabilidad final condicionada de que $z_i = 0$ o, equivalentemente, de que t_i se haya generado por la primera componente de la mixtura, dados los datos y el resto de parámetros, es

$$Pr(z_i = 0|t_i, p, \phi, \mu_i) = \left(1 + \frac{1-p}{p} e^{-\frac{\phi}{2}(\mu_i^2 - 2\mu_i t_i)}\right)^{-1}\tag{3.12}$$

y por tanto, $Pr(z_i = 1|t_i, p, \phi, \mu_i) = 1 - Pr(z_i = 0|t_i, p, \phi, \mu_i)$.

También se puede ver fácilmente que la distribución final condicionada de p , dados los datos y el resto de parámetros, es la distribución definida en (2.22), es decir,

$$p|\mathbf{t}, \mathbf{z} \sim \text{Beta}(n_0 + \alpha, n_1 + \beta)\tag{3.13}$$

y la distribución final condicionada de ϕ , dados los datos y el resto de parámetros, es

$$\phi|\mathbf{t}, \mathbf{z}, \mu_1, \dots, \mu_N \sim \text{Gamma}\left(\frac{a + 2N}{2}, \frac{K}{2}\right),\tag{3.14}$$

donde $K = b + \sum_{i=1}^N c_i \mu_i^2 + \sum_{i:z_i=0} t_i^2 + \sum_{i:z_i=1} (t_i - \mu_i)^2$.

Finalmente, para $i = 1, \dots, N$, la distribución final condicionada de μ_i , dados los datos y el resto de parámetros, depende de los valores de z_i y viene dada por

$$\mu_i | t_i, z_i = 0, \phi \sim N \left(0, \frac{1}{\sqrt{c_i \phi}} \right) \quad (3.15)$$

y

$$\mu_i | t_i, z_i = 1, \phi \sim N \left(\frac{t_i}{1 + c_i}, \frac{1}{\sqrt{(1 + c_i) \phi}} \right) \quad (3.16)$$

En el apéndice 3A.1 del capítulo se muestra la obtención de las distribuciones (3.12), (3.13), (3.14), (3.15) y (3.16).

3.3.1 Algoritmo MCMC: muestreo de Gibbs

El objetivo es generar valores de una cadena MCMC, cuya distribución estacionaria sea la distribución final $\pi(\boldsymbol{\theta}, \mathbf{z} | \mathbf{t}, \alpha, \beta, a, b, \mathbf{c})$ para estimar, mediante dicha muestra, los parámetros del modelo y la probabilidad final de cada hipótesis nula. Esta cadena MCMC se puede construir con las distribuciones dadas en (3.12), (3.13), (3.14), (3.15) y (3.16) mediante un algoritmo MCMC que, al igual que en la sección 2.4.1, consiste en un muestreo de Gibbs incluyendo las variables latentes. El esquema es el siguiente,

1. Especificar valores iniciales para $\boldsymbol{\theta}^{(0)} = (p^{(0)}, \phi^{(0)}, \mu_1^{(0)}, \dots, \mu_N^{(0)})$.
2. Actualizar las variables latentes generando de $\mathbf{z}^{(j+1)} \sim \mathbf{z} | \mathbf{t}, \boldsymbol{\theta}^{(j)}$.
3. Actualizar los parámetros $\boldsymbol{\theta}^{(j+1)} = (p^{(j+1)}, \phi^{(j+1)}, \mu_1^{(j+1)}, \dots, \mu_N^{(j+1)})$.
 - 3.1. Generar valores de $p^{(j+1)} \sim p | \mathbf{t}, \mathbf{z}^{(j+1)}$.

3.2. Generar valores de $\phi^{(j+1)} \sim \phi | \mathbf{t}, \mathbf{z}^{(j+1)}, \mu_1^{(j)}, \dots, \mu_N^{(j)}$.

3.3. Para $i = 1, \dots, N$, generar valores de

$$\begin{aligned} \mu_i^{(j+1)} &\sim \mu_i | t_i, z_i^{(j+1)} = 0, \phi^{(j+1)} & \text{si } z_i^{(j+1)} = 0 & \text{ y de} \\ \mu_i^{(j+1)} &\sim \mu_i | t_i, z_i^{(j+1)} = 1, \phi^{(j+1)} & \text{si } z_i^{(j+1)} = 1. \end{aligned}$$

4. $j = j + 1$. Ir a 2.

Este algoritmo se puede implementar fácilmente, puesto que todas las distribuciones condicionadas son explícitas. En el paso 1 se escogen valores iniciales para los parámetros del modelo. En el paso 2 se genera un valor para cada una de las variables latentes a partir de la distribución dada en (3.12) y por último, en el paso 3, se genera un valor para cada uno de los parámetros del modelo, utilizando las distribuciones dadas en (3.13) y (3.14) para p y ϕ , respectivamente y para μ_i , con $i = 1, \dots, N$, a partir de las distribuciones dadas en (3.15) y (3.16) dependiendo de si z_i es cero o uno, respectivamente.

Entonces, una vez obtenida una muestra de la cadena de Markov en equilibrio,

$$\left\{ \left(\boldsymbol{\theta}^{(j)}, \mathbf{z}^{(j)} \right) : j = 1, \dots, M \right\},$$

donde $\boldsymbol{\theta}^{(j)} = \left(p^{(j)}, \phi^{(j)}, \mu_1^{(j)}, \dots, \mu_N^{(j)} \right)$ y $\mathbf{z}^{(j)} = \left(z_1^{(j)}, \dots, z_N^{(j)} \right)$, se puede estimar la media final de cada parámetro mediante la correspondiente media muestral. Por tanto, podemos estimar p y ϕ aproximando las medias finales de los mismos, $E[p | \mathbf{t}, \alpha, \beta, a, b, \mathbf{c}]$ y $E[\phi | \mathbf{t}, \alpha, \beta, a, b, \mathbf{c}]$, como en (3.4) y (3.5), respectivamente, y de forma análoga a la expresión dada en (3.6), podemos estimar μ_i , para $i = 1, \dots, N$, mediante

$$\hat{\mu}_i = E[\mu_i | \mathbf{t}, \alpha, \beta, a, b, \mathbf{c}] \approx \frac{1}{M} \sum_{j=1}^M \mu_i^{(j)}. \quad (3.17)$$

Finalmente podemos aproximar la probabilidad final de cada hipótesis nula, $Pr(H_{0i} = 0 | \mathbf{t}, \alpha, \beta, a, b, \mathbf{c})$, como en (3.7).

Una vez se ha obtenido una estimación de p y una estimación de la probabilidad final de cada hipótesis nula, podemos aplicar la regla bayes (3.8) y el criterio de decisión EHNF.

3.3.2 Simulación: resultados y análisis de sensibilidad

En esta subsección, desarrollamos un ejemplo de simulación para ilustrar el procedimiento descrito en esta sección.

Se simuló un conjunto de datos con $N = 5000$ hipótesis y con $n = 5$ observaciones por hipótesis. Los datos se generaron de una mixtura de dos distribuciones normales, tales que $x_{ij} \sim N(0, 1)$ con probabilidad $p = 0.9$ y $x_{ij} \sim N(\mu_i, 1)$ con probabilidad $p = 0.1$, para $i = 1, \dots, 5000$ y $j = 1, \dots, 5$, eligiendo para μ_i , con $i = 1, \dots, 5000$, valores linealmente espaciados en el intervalo $[-4, 4]$. Posteriormente se calcularon las medias muestrales $t_i = \sum_{j=1}^5 x_{ij}/5$, para $i = 1, \dots, 5000$, de manera que los datos t_i quedan como si se hubiesen simulado de la mixtura $0.9 \times N(0, 1/\sqrt{5}) + 0.1 \times N(\mu_i, 1/\sqrt{5})$, para $i = 1, \dots, 5000$.

Con el fin de realizar un análisis de sensibilidad, se consideraron distintos valores para los parámetros de las distribuciones iniciales de p , ϕ y μ_i , para $i = 1, \dots, 5000$, dadas en (3.11), que se detallan a continuación.

Para los parámetros de la distribución inicial $Beta(\alpha, \beta)$ de p , se eligieron los valores $(1, 25)$, $(1, 1)$ y $(25, 1)$, debido a que las distribuciones correspondientes proporcionan un amplio abanico de posibilidades. La distribución $Beta(1, 1)$ es la distribución uniforme en el intervalo $[0, 1]$. La distribución $Beta(1, 25)$ presenta un

sesgo muy fuerte hacia la izquierda, por lo que es poco adecuada inicialmente para los datos generados de una mixtura con probabilidad $p = 0.9$. Por último, la distribución $Beta(25, 1)$ está muy sesgada hacia la derecha, siendo esta distribución la que debería ser, entre las distribuciones iniciales consideradas, la más apropiada para el conjunto de datos que se ha simulado.

Para los parámetros de la distribución inicial $Gamma(a/2, b/2)$ de ϕ , se consideraron los valores $(1, 10)$, $(0, 0)$ y $(10, 1)$, ya que las distribuciones que se obtienen con dichos parámetros presentan grandes diferencias. La distribución $Gamma(1/2, 10/2)$ está muy sesgada hacia la izquierda, concentrando la densidad para valores inferiores a uno, por lo que esta distribución resulta ser inicialmente poco adecuada, dado que los t_i , para $i = 1, \dots, 5000$, se generaron de una mixtura con precisión $\phi = 5$. La distribución $Gamma(10/2, 1/2)$ debería ser la más conveniente entre las distribuciones iniciales consideradas, puesto que concentra la densidad para valores de ϕ próximos a cinco, siendo éste el valor utilizado en la simulación y, finalmente, la distribución $Gamma(0, 0)$ es la distribución no informativa.

Por último, para el parámetro asociado a la varianza de la distribución inicial de μ_i se consideró $c = c_i$ para todo $i = 1, \dots, 5000$. Dado que los valores para μ_i , con $i = 1, \dots, 5000$, se eligieron en el intervalo $[-4, 4]$, la varianza inicial de μ_i en (3.11) debería ser un valor próximo a dos, y como el valor elegido en la simulación para ϕ es cinco, un valor apropiado inicialmente para c debería estar próximo a 0.1. Así, se eligieron para c los valores 0.00001, 0.0001, 0.001, 0.01, 0.1, 0.2 y 0.4, de manera que, para los valores de c menores que 0.1, se obtiene una varianza inicial para μ_i mayor que dos, y para los valores de c mayores que 0.1, se obtiene una varianza inicial menor que dos.

Dados los datos simulados y para los distintos valores considerados de los pará-

Tabla 3.1: Estimación final de p y ϕ para distintos valores de los parámetros de las distribuciones iniciales de p y μ_i , $i = 1, \dots, 5000$ y con $\phi \sim Gamma(0, 0)$.

(α, β)	c	0.00001	0.0001	0.001	0.01	0.1	0.2	0.4
(1,25)	\hat{p}	0.94	0.93	0.92	0.89	0.82	0.80	0.77
	$\hat{\phi}$	3.95	4.50	5.01	5.60	4.88	3.90	2.96
(1,1)	\hat{p}	0.95	0.94	0.92	0.90	0.83	0.81	0.79
	$\hat{\phi}$	3.89	4.46	4.99	5.54	4.81	3.86	2.93
(25,1)	\hat{p}	0.95	0.94	0.93	0.90	0.83	0.82	0.79
	$\hat{\phi}$	3.91	4.47	4.99	5.54	4.80	3.85	2.93

metros (α, β) , (a, b) y c , se obtuvo una muestra de la cadena de Markov construida según un muestreo de Gibbs, aplicando el algoritmo descrito en esta sección con las distribuciones (3.12), (3.13), (3.14), (3.15) y (3.16), para estimar p , ϕ y μ_i , con $i = 1, \dots, 5000$ y la probabilidad final de cada hipótesis nula como en (3.4), (3.5), (3.17) y (3.7) respectivamente. Cabe señalar que el coste computacional es muy bajo, empleando noventa segundos aproximadamente, con 20000 iteraciones (descartando las 10000 primeras como iteraciones iniciales).

En la tabla 3.1 se muestran las estimaciones obtenidas de p y ϕ para distintos valores de los parámetros (α, β) y c de las respectivas distribuciones iniciales de p y μ_i , para $i = 1, \dots, 5000$, y con distribución inicial $Gamma(0, 0)$ para ϕ . Se puede observar que, fijado el valor de c , el procedimiento es robusto respecto a los parámetros considerados de la distribución inicial de p . Cabe destacar que se obtienen resultados muy similares incluso con la distribución $Beta(1, 25)$, poco adecuada inicialmente para los datos generados de una mixtura con probabilidad $p = 0.9$, hecho que no sucedía en el caso de los contrastes múltiples con hipótesis simples tratado en el capítulo anterior.

Tabla 3.2: Estimación final de p y ϕ para distintos valores de los parámetros de las distribuciones iniciales de p y μ_i , $i = 1, \dots, 5000$ y con $\phi \sim \text{Gamma}(1/2, 10/2)$.

(α, β)	c	0.00001	0.0001	0.001	0.01	0.1	0.2	0.4
(1,25)	\hat{p}	0.94	0.93	0.92	0.89	0.82	0.80	0.77
	$\hat{\phi}$	3.86	4.44	4.94	5.50	4.81	3.86	2.94
(1,1)	\hat{p}	0.95	0.94	0.93	0.90	0.83	0.81	0.79
	$\hat{\phi}$	3.86	4.41	4.91	5.46	4.74	3.82	2.91
(25,1)	\hat{p}	0.95	0.94	0.93	0.90	0.84	0.82	0.79
	$\hat{\phi}$	3.82	4.41	4.91	5.45	4.73	3.81	2.90

Sin embargo, el parámetro c tiene gran influencia en los resultados, ya que se observan diferencias importantes en las estimaciones obtenidas para los distintos valores de c . Se puede inferir que, para estimar p , un valor adecuado para c debería estar próximo a 0.01, mientras que para estimar ϕ debería estar entre 0.001 y 0.1, puesto que para estos valores de c se obtienen las estimaciones de p y ϕ más próximas a los valores con los que se generaron los datos ($p = 0.9$ y $\phi = 5$).

En las tablas 3.2 y 3.3, para los valores $(a, b) = (1, 10)$ y $(a, b) = (10, 1)$ respectivamente, de la distribución inicial $\text{Gamma}(a/2, b/2)$ de ϕ , se muestran las estimaciones de p y ϕ para los mismos valores de los parámetros (α, β) y c considerados en la tabla 3.1. Como se puede observar se obtienen resultados muy similares a los obtenidos en la tabla 3.1, por lo que el procedimiento también es robusto respecto a los parámetros considerados de la distribución inicial de ϕ .

Por otro lado, para decidir qué hipótesis nulas se rechazan aplicamos, en primer lugar, la regla bayes (3.8), considerando $C_{0i} = C_{1i}$ para $i = 1, \dots, 5000$, por lo que se rechazarán todas las hipótesis nulas tales que $\hat{Pr}(H_{0i} = 0 | \mathbf{t}, \alpha, \beta, a, b, \mathbf{c}) \leq 0.5$. Por

Tabla 3.3: Estimación final de p y ϕ para distintos valores de los parámetros de las distribuciones iniciales de p y μ_i , $i = 1, \dots, 5000$ y con $\phi \sim \text{Gamma}(10/2, 1/2)$.

(α, β)	c	0.00001	0.0001	0.001	0.01	0.1	0.2	0.4
(1,25)	\hat{p}	0.94	0.93	0.92	0.89	0.82	0.80	0.77
	$\hat{\phi}$	3.93	4.52	5.03	5.61	4.88	3.91	2.97
(1,1)	\hat{p}	0.95	0.94	0.92	0.90	0.83	0.81	0.78
	$\hat{\phi}$	3.90	4.48	4.99	5.55	4.82	3.87	2.94
(25,1)	\hat{p}	0.95	0.94	0.92	0.90	0.83	0.81	0.79
	$\hat{\phi}$	3.91	4.48	4.99	5.55	4.81	3.86	2.93

último, aplicamos el criterio EHNF rechazando las $\widehat{N}_1 = N(1 - \hat{p})$ hipótesis nulas con menor probabilidad final estimada.

En la tabla 3.4, para la distribución inicial $\text{Gamma}(0, 0)$ de ϕ y para los distintos valores considerados de los parámetros (α, β) y c de las respectivas distribuciones iniciales de p y μ_i , para $i = 1, \dots, 5000$, se muestran los resultados obtenidos aplicando la regla bayes, esto es, el porcentaje de hipótesis nulas rechazadas ($R_C\%$) y los porcentajes de falsos positivos ($FP_C\%$) y falsos negativos ($FN_C\%$) calculados como en (2.13), es decir,

$$\begin{aligned}
 FP_C\% &= \frac{\sum_{i=1}^N I(\hat{Pr}(H_{0i} = 0 | \mathbf{t}, \alpha, \beta, a, b, \mathbf{c}) \leq 0.5) I(\mu_i = 0)}{\sum_{i=1}^N I(\mu_i = 0)} \times 100, \\
 FN_C\% &= \frac{\sum_{i=1}^N I(\hat{Pr}(H_{0i} = 0 | \mathbf{t}, \alpha, \beta, a, b, \mathbf{c}) > 0.5) I(\mu_i \neq 0)}{\sum_{i=1}^N I(\mu_i \neq 0)} \times 100. \quad (3.18)
 \end{aligned}$$

Como puede observarse, para cada valor de c , los valores de $R_C\%$, $FP_C\%$ y

Tabla 3.4: Porcentaje de hipótesis nulas rechazadas ($R_C\%$) y porcentajes de falsos positivos ($FP_C\%$) y falsos negativos ($FN_C\%$), aplicando la regla bayes para distintos valores de los parámetros de las distribuciones iniciales de p y μ_i , $i = 1, \dots, 5000$ y con $\phi \sim \text{Gamma}(0, 0)$.

(α, β)	c	0.00001	0.0001	0.001	0.01	0.1	0.2	0.4
(1,25)	$R_C\%$	5.28	6.22	6.88	7.92	9.02	8.22	7.70
	$FP_C\%$	0	0.02	0.07	0.36	1.08	0.58	0.22
	$FN_C\%$	51.91	43.53	37.89	30.78	26.59	29.87	31.69
(1,1)	$R_C\%$	5.20	6.16	6.88	7.78	8.76	8.04	7.56
	$FP_C\%$	0	0.02	0.07	0.27	0.90	0.47	0.16
	$FN_C\%$	52.64	44.08	37.89	31.33	27.50	30.60	32.42
(25,1)	$R_C\%$	5.24	6.22	6.86	7.80	8.70	8.00	7.54
	$FP_C\%$	0	0.02	0.07	0.29	0.88	0.45	0.16
	$FN_C\%$	52.28	43.53	38.07	31.33	27.87	30.78	32.60

$FN_C\%$ apenas sufren variación al modificar los parámetros de la distribución inicial de p . Con respecto al parámetro c sí se obtiene sensibilidad, siendo los valores de c en torno a 0.1 los más apropiados para los datos simulados, ya que los porcentajes de hipótesis nulas rechazadas que se obtienen para este valor de c son los más próximos al porcentaje de hipótesis nulas falsas con el que se generaron los datos (10%), a la vez que se obtienen los porcentajes más bajos de falsos negativos.

En las tablas 3.5 y 3.6, para los mismos valores de los parámetros (α, β) y c considerados en la tabla 3.4, se muestran los mismos resultados que en esta última pero para los valores $(a, b) = (1, 10)$ y $(a, b) = (10, 1)$, respectivamente, de la distribución inicial $\text{Gamma}(a/2, b/2)$ de ϕ . Se puede observar que el procedimiento también es robusto respecto a los parámetros considerados de la distribución inicial de ϕ ya que los resultados que se obtienen son muy similares a los obtenidos en la tabla 3.4.

Tabla 3.5: Porcentaje de hipótesis nulas rechazadas ($R_C\%$) y porcentajes de falsos positivos ($FP_C\%$) y falsos negativos ($FN_C\%$), aplicando la regla bayes para distintos valores de los parámetros de las distribuciones iniciales de p y μ_i , $i = 1, \dots, 5000$ y con $\phi \sim \text{Gamma}(1/2, 10/2)$.

(α, β)	c	0.00001	0.0001	0.001	0.01	0.1	0.2	0.4
(1,25)	$R_C\%$	5.18	6.20	6.84	7.82	8.86	8.14	7.66
	$FP_C\%$	0	0.02	0.07	0.29	0.99	0.49	0.18
	$FN_C\%$	52.82	43.72	38.25	31.15	27.32	29.87	31.69
(1,1)	$R_C\%$	5.18	6.16	6.80	7.76	8.68	8.02	7.60
	$FP_C\%$	0	0.02	0.07	0.27	0.88	0.47	0.18
	$FN_C\%$	52.82	44.08	38.62	31.51	28.05	30.78	32.24
(25,1)	$R_C\%$	5.14	6.18	6.76	7.74	8.58	7.98	7.48
	$FP_C\%$	0	0.02	0.04	0.27	0.79	0.43	0.16
	$FN_C\%$	53.19	43.90	38.80	31.69	28.23	30.78	33.15

Finalmente, en las tablas 3.7, 3.8 y 3.9, para los valores $(0, 0)$, $(1, 10)$ y $(10, 1)$, respectivamente, de los parámetros (a, b) de la distribución inicial de ϕ y para los distintos valores considerados de los parámetros (α, β) y c , se muestran los resultados obtenidos aplicando el criterio de decisión EHNF que nosotros proponemos, es decir, el porcentaje estimado de hipótesis nulas falsas ($\widehat{N}_1\%$), la probabilidad final estimada más alta, $prob_{\widehat{N}_1} = \hat{Pr}(H_{(0\widehat{N}_1)} = 0 | \mathbf{t}, \alpha, \beta, a, b, \mathbf{c})$, con la que se rechaza cada hipótesis nula, siendo $\hat{Pr}(H_{(0i)} = 0 | \mathbf{t}, \alpha, \beta, a, b, \mathbf{c})$ las probabilidades finales ordenadas, y el porcentaje de falsos positivos ($FP_{EHNF}\%$) y falsos negativos ($FN_{EHNF}\%$) calculados como en (2.14), esto es,

$$FP_{EHNF}\% = \frac{\sum_{i=1}^N I(\hat{Pr}(H_{0i} = 0 | \mathbf{t}, \alpha, \beta, a, b, \mathbf{c}) \leq prob_{\widehat{N}_1}) I(\mu_i = 0)}{\sum_{i=1}^N I(\mu_i = 0)} \times 100,$$

Tabla 3.6: Porcentaje de hipótesis nulas rechazadas ($R_C\%$) y porcentajes de falsos positivos ($FP_C\%$) y falsos negativos ($FN_C\%$), aplicando la regla bayes para distintos valores de los parámetros de las distribuciones iniciales de p y μ_i , $i = 1, \dots, 5000$ y con $\phi \sim \text{Gamma}(10/2, 1/2)$.

(α, β)	c	0.00001	0.0001	0.001	0.01	0.1	0.2	0.4
(1,25)	$R_C\%$	5.24	6.22	6.88	7.90	9.04	8.30	7.70
	$FP_C\%$	0	0.02	0.07	0.34	1.10	0.65	0.22
	$FN_C\%$	52.28	43.53	37.89	30.78	26.59	29.69	31.69
(1,1)	$R_C\%$	5.20	6.22	6.82	7.80	8.78	8.08	7.56
	$FP_C\%$	0	0.02	0.07	0.29	0.92	0.47	0.16
	$FN_C\%$	52.64	43.53	38.43	31.33	27.50	30.24	32.42
(25,1)	$R_C\%$	5.24	6.22	6.88	7.84	8.74	8.04	7.56
	$FP_C\%$	0	0.02	0.07	0.31	0.90	0.47	0.16
	$FN_C\%$	52.28	43.53	37.89	31.15	27.69	30.60	32.42

$$FN_{EHN\%} = \frac{\sum_{i=1}^N I(\hat{Pr}(H_{0i} = 0 | \mathbf{t}, \alpha, \beta, a, b, \mathbf{c}) > prob_{\hat{N}_1}) I(\mu_i \neq 0)}{\sum_{i=1}^N I(\mu_i \neq 0)} \times 100. \tag{3.19}$$

Igualmente se puede ver que, para cada valor de c , el procedimiento es robusto respecto a los parámetros considerados de las distribuciones iniciales de p y ϕ .

Sin embargo, los resultados muestran aún más sensibilidad respecto al parámetro c que cuando se aplica la regla bayes. Se puede observar que un valor adecuado para c debería estar entre 0.01 y 0.1, puesto que en este intervalo se obtienen los porcentajes más bajos de falsos negativos, manteniendo aceptables los porcentajes de falsos positivos ($< 10\%$).

Tabla 3.7: Porcentaje estimado de hipótesis nulas falsas ($\widehat{N}_1\%$), probabilidad final estimada más alta con la que se rechaza cada hipótesis nula ($prob_{\widehat{N}_1}$) y porcentajes de falsos positivos ($FP_{EHN\%}$) y falsos negativos ($FN_{EHN\%}$), aplicando el criterio de decisión EHN, para distintos valores de los parámetros de las distribuciones iniciales de p y μ_i , $i = 1, \dots, 5000$ y con $\phi \sim Gamma(0, 0)$.

(α, β)	c	0.00001	0.0001	0.001	0.01	0.1	0.2	0.4
(1,25)	$\widehat{N}_1\%$	5.90	6.82	7.98	10.74	17.76	20.02	23.06
	$prob_{\widehat{N}_1}$	0.8016	0.8520	0.8933	0.8647	0.8280	0.8235	0.8058
	$FP_{EHN\%}$	0.02	0.07	0.38	2.70	9.86	12.31	15.68
	$FN_{EHN\%}$	46.45	38.43	30.42	24.04	18.03	17.49	16.76
(1,1)	$\widehat{N}_1\%$	5.36	6.30	7.48	10.08	16.68	18.72	21.42
	$prob_{\widehat{N}_1}$	0.6118	0.6514	0.7612	0.8464	0.8288	0.8271	0.8145
	$FP_{EHN\%}$	0	0.02	0.18	2.00	8.65	10.87	13.88
	$FN_{EHN\%}$	51.18	42.8051	33.33	24.41	18.21	17.67	17.12
(25,1)	$\widehat{N}_1\%$	5.36	6.28	7.44	10.02	16.46	18.44	21.00
	$prob_{\widehat{N}_1}$	0.6276	0.6220	0.7395	0.8437	0.8287	0.8296	0.8169
	$FP_{EHN\%}$	0	0.02	0.16	1.98	8.40	10.56	13.43
	$FN_{EHN\%}$	51.18	42.99	33.51	24.59	18.21	17.67	17.12

Por tanto, podemos concluir que el procedimiento descrito en esta sección es robusto respecto a la elección de los parámetros de las distribuciones iniciales de p y ϕ . Sin embargo, existe sensibilidad respecto al parámetro c_i asociado a la varianza de la distribución inicial de μ_i , siendo esta sensibilidad más acusada con el criterio de decisión EHN que con la regla bayes.

Por último, también hemos podido comprobar que, para los datos simulados, un valor apropiado para c_i debería estar comprendido entre 0.01 y 0.1. Este último resultado es coherente con la idea, expuesta al principio de la subsección, de que un

Tabla 3.8: Porcentaje estimado de hipótesis nulas falsas ($\widehat{N}_1\%$), probabilidad final estimada más alta con la que se rechaza cada hipótesis nula ($prob_{\widehat{N}_1}$) y porcentajes de falsos positivos ($FP_{EHNF}\%$) y falsos negativos ($FN_{EHNF}\%$), aplicando el criterio de decisión EHNF, para distintos valores de los parámetros de las distribuciones iniciales de p y μ_i , $i = 1, \dots, 5000$ y con $\phi \sim Gamma(1/2, 10/2)$.

(α, β)	c	0.00001	0.0001	0.001	0.01	0.1	0.2	0.4
(1,25)	$\widehat{N}_1\%$	5.80	6.78	7.94	10.64	17.56	19.82	22.88
	$prob_{\widehat{N}_1}$	0.7860	0.8592	0.8877	0.8616	0.8298	0.8244	0.8069
	$FP_{EHNF}\%$	0.02	0.04	0.38	2.61	9.62	12.09	15.52
	$FN_{EHNF}\%$	47.36	38.62	30.78	24.23	18.03	17.49	17.12
(1,1)	$\widehat{N}_1\%$	5.34	6.28	7.42	10.00	16.48	18.58	21.28
	$prob_{\widehat{N}_1}$	0.6018	0.6656	0.7420	0.8497	0.8312	0.8284	0.8155
	$FP_{EHNF}\%$	0	0.02	0.16	1.93	8.42	10.74	13.68
	$FN_{EHNF}\%$	51.37	42.99	33.70	24.59	18.21	17.85	17.12
(25,1)	$\widehat{N}_1\%$	5.26	6.24	7.36	9.92	16.26	18.28	20.84
	$prob_{\widehat{N}_1}$	0.5909	0.6097	0.7491	0.8463	0.8304	0.8297	0.8186
	$FP_{EHNF}\%$	0	0.02	0.16	1.86	8.20	10.42	13.17
	$FN_{EHNF}\%$	52.09	43.35	34.24	24.59	18.40	18.03	16.94

valor adecuado inicialmente para c_i debería estar próximo a 0.1, para que la varianza inicial de μ_i en (3.11) sea un valor próximo a dos, puesto que los valores para μ_i , con $i = 1, \dots, 5000$, se eligieron en el intervalo $[-4, 4]$.

Por otro lado, decidir cuál de los dos criterios de decisión es más adecuado depende también del valor de c . Comparando los resultados obtenidos en las tablas 3.4, 3.5 y 3.6 aplicando la regla bayes, considerando $C_{0i} = C_{1i}$ para $i = 1, \dots, 5000$, con los obtenidos en las tablas 3.7, 3.8 y 3.9 utilizando el criterio EHNF, se puede observar que con el criterio EHNF se obtienen unos porcentajes más bajos de falsos

Tabla 3.9: Porcentaje estimado de hipótesis nulas falsas ($\widehat{N}_1\%$), probabilidad final estimada más alta con la que se rechaza cada hipótesis nula ($prob_{\widehat{N}_1}$) y porcentajes de falsos positivos ($FP_{EHNF}\%$) y falsos negativos ($FN_{EHNF}\%$), aplicando el criterio de decisión EHNF, para distintos valores de los parámetros de las distribuciones iniciales de p y μ_i , $i = 1, \dots, 5000$ y con $\phi \sim Gamma(10/2, 1/2)$.

(α, β)	c	0.00001	0.0001	0.001	0.01	0.1	0.2	0.4
(1,25)	$\widehat{N}_1\%$	5.86	6.84	8.00	10.76	17.78	20.04	23.02
	$prob_{\widehat{N}_1}$	0.7910	0.8440	0.8892	0.8600	0.8275	0.8222	0.8053
	$FP_{EHNF}\%$	0.02	0.07	0.45	2.67	9.91	12.36	15.61
	$FN_{EHNF}\%$	46.81	38.25	30.78	23.68	18.21	17.49	16.94
(1,1)	$\widehat{N}_1\%$	5.34	6.32	7.48	10.10	16.72	18.78	21.46
	$prob_{\widehat{N}_1}$	0.5860	0.6625	0.7668	0.8459	0.8289	0.8269	0.8144
	$FP_{EHNF}\%$	0	0.02	0.16	2.02	8.69	10.94	13.88
	$FN_{EHNF}\%$	51.37	42.62	33.15	24.41	18.21	17.67	17.12
(25,1)	$\widehat{N}_1\%$	5.34	6.28	7.44	10.02	16.50	18.48	21.02
	$prob_{\widehat{N}_1}$	0.5561	0.6260	0.7463	0.8436	0.8283	0.8274	0.8169
	$FP_{EHNF}\%$	0	0.02	0.16	1.93	8.47	10.63	13.39
	$FN_{EHNF}\%$	51.37	42.99	33.51	24.41	18.21	17.85	17.12

negativos que los que se obtienen con la regla bayes. Sin embargo, los porcentajes de falsos positivos son más altos que los que se obtienen con la regla bayes; si bien para los valores de c menores o iguales que 0.1 se mantienen en unos niveles aceptables ($< 10\%$), mientras que para los valores de c mayores que 0.1 son algo elevados ($> 10\%$) para los niveles habitualmente utilizados en la literatura.

Por tanto, podemos afirmar que, para los valores de c menores o iguales que 0.1, resulta más adecuado aplicar el criterio EHNF, ya que se consigue un equilibrio mayor entre los porcentajes de falsos positivos y falsos negativos; mientras que, para

los valores de c mayores que 0.1, resultaría más adecuado utilizar la regla bayes, con $C_{0i} = C_{1i}$ para $i = 1, \dots, 5000$.

Con el fin de resolver la falta de robustez observada respecto a la elección del parámetro c_i , en la siguiente sección se propone un modelo bayesiano jerárquico, considerando una única distribución inicial para el parámetro c_i , con $i = 1, \dots, 5000$.

3.4 Modelo Gaussiano jerárquico

En esta sección se propone un modelo bayesiano jerárquico, con la finalidad de corregir la sensibilidad observada a la elección del parámetro c_i de la distribución inicial de μ_i dada en (3.11). Consideramos una distribución inicial para dicho parámetro, y añadimos así un nivel más en la jerarquía del modelo planteado en la sección anterior.

Consideramos, nuevamente, el problema de contraste de hipótesis múltiples definido en (3.9),

$$H_{0i} : \mu_i = 0 \quad \text{frente a} \quad H_{1i} : \mu_i \neq 0, \quad i = 1, \dots, N.$$

Suponemos que para cada hipótesis se observa el estadístico media muestral, $T_i = \sum_{j=1}^n X_{ij}/n$, y que las T_i , para $i = 1, \dots, N$, son variables aleatorias independientes e idénticamente distribuidas $N(0, 1/\sqrt{\phi})$ bajo la hipótesis nula y $N(\mu_i, 1/\sqrt{\phi})$, con $\mu_i \neq 0$, bajo la hipótesis alternativa, siendo ϕ la precisión que suponemos desconocida y la misma para todas las hipótesis.

Consideramos para p , ϕ y para μ_i , con $i = 1, \dots, N$, las distribuciones iniciales definidas en (3.11). Para el parámetro c_i de la distribución de μ_i , con $i = 1, \dots, N$,

elegimos la distribución inicial

$$c_i \sim \text{Gamma}(e, f), \quad i = 1, \dots, N. \quad (3.20)$$

Esta distribución se ha seleccionado por conveniencia en los cálculos ya que es la distribución natural conjugada.

La verosimilitud $f(\mathbf{t}, \mathbf{z}|\boldsymbol{\theta})$ viene dada por la expresión (3.10), donde el conjunto de parámetros es ahora $\boldsymbol{\theta} = (p, \phi, \mu_1, \dots, \mu_N, \mathbf{c})$, siendo $\mathbf{c} = (c_1, \dots, c_N)$ y $\mathbf{z} = (z_1, \dots, z_N)$ las variables latentes definidas como en la sección 3.2 (pág. 87).

Entonces, dada la verosimilitud (3.10) y las distribuciones iniciales (3.11) y (3.20), se puede comprobar fácilmente que la probabilidad final condicionada de que $z_i = 0$ o, equivalentemente, de que t_i se haya generado por la primera componente de la mixtura, dados los datos y el resto de parámetros, tiene la expresión dada en (3.12) y las distribuciones finales condicionadas de p , ϕ y μ_i , para $i = 1, \dots, N$, dados los datos y el resto de parámetros son, respectivamente, las distribuciones dadas en (3.13), (3.14) y (3.15) si $z_i = 0$ o (3.16) si $z_i = 1$. En el apéndice 3A.2 del capítulo se prueba que la distribución final condicionada de c_i , dados los datos y el resto de parámetros, viene dada por

$$c_i|\mathbf{t}, \phi, \mu_i \sim \text{Gamma}\left(e + \frac{1}{2}, f + \frac{\phi}{2}\mu_i^2\right) \quad i = 1, \dots, N. \quad (3.21)$$

Así, se puede construir una cadena de Markov cuya distribución estacionaria es la distribución final $\pi(\boldsymbol{\theta}, \mathbf{z}|\alpha, \beta, a, b, e, f)$. El algoritmo MCMC consiste en un muestreo de Gibbs con la misma estructura que el algoritmo expuesto en la subsección 3.3.1, al que se le añade el paso siguiente,

3.4. Para $i = 1, \dots, N$, generar valores de $c_i^{(j+1)} \sim c_i|\mathbf{t}, \phi^{(j+1)}, \mu_i^{(j+1)}$ dada en (3.21).

donde $\boldsymbol{\theta}^{(j)} = (p^{(j)}, \phi^{(j)}, \mu_1^{(j)}, \dots, \mu_N^{(j)}, \mathbf{c}^{(j)})$, siendo $\mathbf{c}^{(j)} = (c_1^{(j)}, \dots, c_N^{(j)})$.

Una vez obtenida una muestra de la cadena de Markov en equilibrio, $\{(\boldsymbol{\theta}^{(j)}, \mathbf{z}^{(j)}) : j = 1, \dots, M\}$, podemos estimar los parámetros del modelo p , ϕ y μ_i , para $i = 1, \dots, N$, aproximando las medias finales de los mismos, $E[p|\mathbf{t}, \alpha, \beta, a, b, e, f]$, $E[\phi|\mathbf{t}, \alpha, \beta, a, b, e, f]$ y $E[\mu_i|\mathbf{t}, \alpha, \beta, a, b, e, f]$ como en (3.4), (3.5) y (3.17), respectivamente. De forma análoga, podemos también estimar c_i , para $i = 1, \dots, N$, mediante

$$\hat{c}_i = E[c_i|\mathbf{t}, \alpha, \beta, a, b, e, f] \approx \frac{1}{M} \sum_{j=1}^M c_i^{(j)}.$$

Finalmente, podemos aproximar la probabilidad final de cada hipótesis nula, $Pr(H_{0i} = 0|\mathbf{t}, \alpha, \beta, a, b, e, f)$ como en (3.7).

Por último, dada una estimación de la probabilidad final de cada hipótesis nula y una estimación de p podemos aplicar igualmente la regla bayes (3.8) y el criterio de decisión EHNF.

3.4.1 Simulación: resultados y análisis de sensibilidad

En esta subsección, se ilustra el modelo bayesiano jerárquico utilizando los datos simulados en la sección anterior.

En la sección 3.3 comprobamos que el procedimiento bayesiano es robusto respecto a la elección de los parámetros iniciales, excepto para el parámetro c_i de la distribución inicial de μ_i , que tiene gran influencia en los resultados. Por tanto, en esta sección, nos centramos en analizar la sensibilidad a la elección de los parámetros (e, f) de la distribución inicial de c_i dada en (3.20).

Con el fin de realizar un análisis de sensibilidad a la elección de estos parámetros,

Tabla 3.10: Estimación final de p y ϕ para distintos valores de los parámetros (e, f) de la distribución inicial de c_i , $i = 1, \dots, 5000$ y para $p \sim Beta(1, 1)$ y $\phi \sim Gamma(0, 0)$.

(e, f)	(1, 100000)	(1, 1000)	(1, 10)	(9, 10)
\hat{p}	0.95	0.93	0.83	0.70
$\hat{\phi}$	3.82	4.92	5.73	2.54

se consideraron los valores (1, 100000), (1, 1000), (1, 10) y (9, 10) para (e, f) , ya que las distribuciones que se obtienen son muy distintas. Inicialmente la distribución más apropiada para c_i debería ser la distribución $Gamma(1, 10)$, ya que concentra la densidad en los valores de c_i próximos a 0.1, siendo éste un valor adecuado inicialmente para c_i como se comentó en la sección anterior. Sin embargo, para los demás valores considerados de los parámetros, es precisamente en torno a este valor de c_i donde las distribuciones correspondientes dan menor densidad, por lo que estos valores de los parámetros deberían ser poco apropiados inicialmente. Para los parámetros (α, β) y (a, b) de las distribuciones iniciales de p y ϕ se utilizaron los valores (1, 1) y (0, 0), respectivamente.

Dados los datos simulados y para los valores considerados de los parámetros (α, β) , (a, b) y (e, f) , se obtuvo una muestra de la cadena de Markov construida según un muestreo de Gibbs, aplicando el algoritmo descrito en esta sección con las distribuciones dadas en (3.12), (3.13), (3.14), (3.15), (3.16) y (3.21), para estimar p , ϕ y μ_i , con $i = 1, \dots, 5000$ y la probabilidad final de cada hipótesis nula como en (3.4), (3.5), (3.17) y (3.7) respectivamente.

En la tabla 3.10 se muestran las estimaciones obtenidas de p y ϕ , utilizando como distribución inicial de p una distribución $Beta(1, 1)$ y de ϕ una distribución $Gamma(0, 0)$ y para distintos valores de los parámetros (e, f) de las distribución

Tabla 3.11: Resultados aplicando la regla bayes y el criterio de decisión EHNF, para distintos valores de los parámetros (e, f) de la distribución inicial de $c_i, i = 1, \dots, 5000$ y para $p \sim Beta(1, 1)$ y $\phi \sim Gamma(0, 0)$.

(e, f)	$R_C\%$	$FP_C\%$	$FN_C\%$	$\widehat{N}_1\%$	$prob_{\widehat{N}_1}$	$FP_{EHNF}\%$	$FN_{EHNF}\%$
(1, 100000)	5.16	0	53.00	5.24	0.5626	0	52.28
(1, 1000)	6.74	0.04	38.98	7.30	0.7626	0.16	34.79
(1, 10)	9.38	1.35	25.50	16.62	0.8197	8.58	18.21
(9, 10)	7.58	0.18	32.42	29.62	0.7435	22.78	14.57

inicial de $c_i, i = 1, \dots, 5000$. Como se puede observar hay sensibilidad a la elección de los parámetros (e, f) de la distribución inicial de c_i .

Para decidir qué hipótesis nulas se rechazan aplicamos, en primer lugar, la regla bayes (3.8), considerando $C_{0i} = C_{1i}$ para $i = 1, \dots, 5000$, rechazando así todas las hipótesis nulas tales que $\hat{P}r(H_{0i} = 0 | \mathbf{t}, \alpha, \beta, a, b, e, f) \leq 0.5$. Finalmente, aplicamos el criterio EHNF rechazando las $\widehat{N}_1 = N(1 - \hat{p})$ hipótesis nulas con menor probabilidad final estimada.

En la tabla 3.11 se muestran, para los mismos valores de los parámetros (α, β) , (a, b) y (e, f) utilizados en la tabla 3.10, los resultados obtenidos aplicando la regla bayes considerando $C_{0i} = C_{1i}$ para $i = 1, \dots, 5000$, es decir, el porcentaje de hipótesis nulas rechazadas ($R_C\%$) y los porcentajes de falsos positivos ($FP_C\%$) y falsos negativos ($FN_C\%$) calculados como en (3.18). En esta tabla también se muestran los resultados obtenidos utilizando el criterio de decisión EHNF, esto es, el porcentaje estimado de hipótesis nulas falsas ($\widehat{N}_1\%$), la probabilidad final estimada más alta con la que se rechaza cada hipótesis nula ($prob_{\widehat{N}_1}$) y los porcentajes de falsos positivos ($FP_{EHNF}\%$) y falsos negativos ($FN_{EHNF}\%$) calculados como

en (3.19). Se puede observar que el procedimiento no es robusto respecto a los parámetros (e, f) de la distribución inicial de c_i .

Así, podemos concluir que la sensibilidad que había, respecto a la elección del parámetro c_i asociado a la varianza de las medias del modelo, se ha trasladado a los parámetros (e, f) de la distribución inicial de dicho parámetro. Ello nos lleva a considerar, mediante una aproximación Empírico Bayes, la estimación del parámetro $c = c_i$, para $i = 1, \dots, N$. Ese es el objetivo de la siguiente sección.

3.5 Modelo Gaussiano. Aproximación Empírico Bayes

En esta sección consideramos el mismo modelo Gaussiano definido en la sección 3.3, donde $\boldsymbol{\theta} = (p, \phi, \mu_1, \dots, \mu_N)$ es el conjunto de parámetros del modelo con las distribuciones iniciales dadas en (3.11).

En la sección 3.3 se comprobó, mediante un ejemplo de simulación, que el procedimiento es robusto respecto a la elección de los parámetros iniciales, excepto para el parámetro c_i . En la sección 3.4, con la finalidad de evitar la sensibilidad a la elección de c_i se utilizó un modelo jerárquico, considerando una distribución inicial para este parámetro y comprobando, con los datos simulados, que la sensibilidad se trasladaba a la elección de los parámetros de la distribución inicial considerada.

Alternativamente, en esta sección se considera una aproximación Empírico Bayes para estimar el parámetro $c = c_i$, para $i = 1, \dots, N$. Como se propone en Casella (2001), una vez obtenido un valor estimado, \hat{c} , se pueden generar muestras de la distribución final $\pi(\boldsymbol{\theta}, \mathbf{z}|\mathbf{t}, \alpha, \beta, a, b, \hat{c})$ mediante un muestreo de Gibbs, utilizando el

algoritmo descrito en la sección 3.3. Por tanto, podemos estimar p , ϕ y μ_i , para $i = 1, \dots, N$, como en (3.4), (3.5) y (3.17), respectivamente, y la probabilidad final de cada hipótesis nula, $Pr(H_{0i}|\mathbf{t}, \alpha, \beta, a, b, \hat{c})$, como en (3.7).

A diferencia de Casella (2001) que estima un parámetro del modelo, nosotros estamos interesados en estimar un parámetro de la distribución inicial de las medias del modelo. Por otro lado, Casella (2001) considera el estimador de máxima verosimilitud bajo la distribución predictiva inicial, mientras que nosotros proponemos utilizar el estimador de máxima verosimilitud bajo la distribución predictiva final. Lo hacemos así porque, en este caso, no se puede obtener la distribución predictiva inicial de forma analítica, y tampoco se puede aproximar por el método de Monte Carlo cuando se utiliza la distribución inicial no informativa $Gamma(0, 0)$ de ϕ , ya que no podemos simular directamente de esta distribución por ser impropia, como se comentó en la sección 2.4.

Sin embargo, sí podemos aproximar la distribución predictiva final por el método de Monte Carlo, utilizando una muestra de la distribución final obtenida mediante un muestreo de Gibbs, ya que todas las distribuciones finales condicionadas que se utilizan son explícitas y propias.

Por tanto, en esta sección se propone utilizar como estimador de c ,

$$\hat{c} = \arg \max_c m(\mathbf{t}|\alpha, \beta, a, b, c),$$

donde

$$\begin{aligned} m(\mathbf{t}|\alpha, \beta, a, b, c) &= \int f(\mathbf{t}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}|\mathbf{t}, \alpha, \beta, a, b, c) d\boldsymbol{\theta} \\ &= \int \prod_{i=1}^N f(t_i|\boldsymbol{\theta})\pi(\boldsymbol{\theta}|\mathbf{t}, \alpha, \beta, a, b, c) d\boldsymbol{\theta}, \end{aligned} \quad (3.22)$$

siendo $f(t_i|\boldsymbol{\theta}) \propto p\phi^{-\frac{1}{2}}e^{-\frac{\phi}{2}t_i^2} + (1-p)\phi^{-\frac{1}{2}}e^{-\frac{\phi}{2}(t_i-\mu_i)^2}$.

Esta integral se puede aproximar por el método de Monte Carlo, utilizando una muestra de la distribución final obtenida mediante un muestreo de Gibbs, aplicando el algoritmo descrito en la sección 3.3. Es decir, para cada valor de c se puede obtener una muestra de la distribución final,

$$\{\boldsymbol{\theta}^{(j)}(c) : j = 1, \dots, M\},$$

donde $\boldsymbol{\theta}^{(j)}(c) = (p^{(j)}(c), \phi^{(j)}(c), \mu_1^{(j)}(c), \dots, \mu_N^{(j)}(c))$, y podemos aproximar la integral (3.22) por

$$\hat{m}(\mathbf{t}|\alpha, \beta, a, b, c) = \frac{1}{M} \sum_{j=1}^M \prod_{i=1}^N f(t_i|\boldsymbol{\theta}^{(j)}(c)). \quad (3.23)$$

Así, podemos obtener una estimación de c de la forma siguiente,

$$\hat{c} = \arg \max_c \frac{1}{M} \sum_{j=1}^M \prod_{i=1}^N f(t_i|\boldsymbol{\theta}^{(j)}(c)). \quad (3.24)$$

Finalmente, como se indicó al principio de la sección, con este valor estimado de c se genera de nuevo, utilizando el mismo algoritmo, una muestra de la distribución final $\pi(\boldsymbol{\theta}, \mathbf{z}|\mathbf{t}, \alpha, \beta, a, b, \hat{c})$, para estimar p , ϕ , μ_i y la probabilidad final de cada hipótesis nula, $Pr(H_{0i}|\mathbf{t}, \alpha, \beta, a, b, \hat{c})$, como en (3.4), (3.5), (3.17) y (3.7), respectivamente.

Por último se pueden aplicar la regla bayes y el criterio de decisión EHNF para decidir las hipótesis nulas que se rechazan y las hipótesis nulas que se aceptan.

3.5.1 Simulación: resultados y análisis de sensibilidad

En esta subsección, se ilustra el procedimiento Empírico Bayes utilizando los mismos datos simulados que en las secciones anteriores.

Para los parámetros (α, β) y (a, b) de las distribuciones iniciales de p y ϕ , dadas en (3.11), se consideraron los mismos valores que en la sección 3.3, es decir, para (α, β) los valores $(1, 25)$, $(1, 1)$ y $(25, 1)$, y para (a, b) los valores $(1, 10)$, $(0, 0)$ y $(10, 1)$.

En primer lugar, para los distintos valores considerados de los parámetros (α, β) y (a, b) , se obtuvo una estimación de c aplicando (3.24). Hay que tener en cuenta que en el proceso de maximización se requiere aplicar el muestreo de Gibbs para cada evaluación de la función (3.23). Sin embargo, esto no supone un problema importante en términos del coste computacional, empleando ochenta y cinco segundos aproximadamente, con 20000 iteraciones (descartando las 10000 primeras como iteraciones iniciales).

Posteriormente, para los distintos valores de los parámetros (α, β) y (a, b) de las distribuciones iniciales de p y ϕ y para las respectivas estimaciones de c , se obtuvo, aplicando de nuevo el algoritmo descrito en la sección 3.3, una muestra de la distribución final $\pi(\boldsymbol{\theta}, \mathbf{z}|\mathbf{t}, \alpha, \beta, a, b, \hat{c})$ para estimar p , ϕ , μ_i y la probabilidad final de cada hipótesis nula, $Pr(H_{0i}|\mathbf{t}, \alpha, \beta, a, b, \hat{c})$, como en (3.4), (3.5), (3.17) y (3.7), respectivamente.

Para decidir las hipótesis nulas que se rechazan, aplicamos, en primer lugar, la regla bayes (3.8) considerando $C_{0i} = C_{1i}$, para $i = 1, \dots, 5000$, por lo que se rechazarán las hipótesis nulas tales que $\hat{Pr}(H_{0i} = 0|\mathbf{t}, \alpha, \beta, a, b, \hat{c}) \leq 0.5$. Por último, aplicamos el criterio de decisión EHNF rechazando las $\hat{N}_1 = N(1 - \hat{p})$ hipótesis nulas con menor probabilidad final estimada.

En la tabla 3.12 se muestra la estimación final de c para distintos valores de los parámetros (α, β) y (a, b) de las respectivas distribuciones iniciales de p y ϕ , junto con las estimaciones finales de p y ϕ para los mismos valores de los parámetros

Tabla 3.12: Estimación final de c para distintos valores de los parámetros de las distribuciones iniciales de $p \sim Beta(\alpha, \beta)$ y $\phi \sim Gamma(a/2, b/2)$, y estimación final de p y ϕ para los mismos valores de los parámetros (α, β) y (a, b) y para las respectivas estimaciones finales de c .

		(a, b)								
		$(0, 0)$			$(1, 10)$			$(10, 1)$		
(α, β)		$(1, 25)$	$(1, 1)$	$(25, 1)$	$(1, 25)$	$(1, 1)$	$(25, 1)$	$(1, 25)$	$(1, 1)$	$(25, 1)$
\hat{c}		0.0860	0.0739	0.0911	0.0865	0.0673	0.0792	0.0865	0.0735	0.0705
\hat{p}		0.83	0.84	0.84	0.83	0.85	0.84	0.83	0.84	0.84
$\hat{\phi}$		5.06	5.15	4.91	4.97	5.15	4.99	5.06	5.17	5.19

(α, β) y (a, b) y para las respectivas estimaciones finales de c . Los valores estimados que se obtienen para c en todos los casos son muy similares y coherentes con los resultados obtenidos en la subsección 3.3.2, donde se pudo observar que un valor apropiado para c debería estar comprendido entre 0.01 y 0.1, por lo que parece que la aproximación Empírico Bayes resulta adecuada para seleccionar los valores óptimos para el parámetro c .

En las tablas 3.13 y 3.14, para los mismos valores de los parámetros considerados en la tabla 3.12, se muestran, respectivamente, los resultados obtenidos aplicando la regla bayes considerando $C_{0i} = C_{1i}$, para $i = 1, \dots, 5000$, es decir, el porcentaje de hipótesis nulas rechazadas ($R_C\%$) y los porcentajes de falsos positivos ($FP_C\%$) y falsos negativos ($FN_C\%$) calculados como en (3.18), y los resultados obtenidos aplicando el criterio de decisión EHNF, esto es, el porcentaje estimado de hipótesis nulas falsas ($\widehat{N}_1\%$), la probabilidad final estimada más alta con la que se rechaza cada hipótesis nula ($prob_{\widehat{N}_1}$) y los porcentajes de falsos positivos ($FP_{EHNF}\%$) y falsos negativos ($FN_{EHNF}\%$) calculados como en (3.19).

Tabla 3.13: Porcentaje de hipótesis nulas rechazadas ($R_C\%$) y porcentajes de falsos positivos ($FP_C\%$) y falsos negativos ($FN_C\%$), aplicando la regla bayes, para distintos valores de los parámetros de las distribuciones iniciales de $p \sim Beta(\alpha, \beta)$ y de $\phi \sim Gamma(a/2, b/2)$ y para las respectivas estimaciones finales de c .

(a, b)	(α, β)	\hat{c}	$R_C\%$	$FP_C\%$	$FN_C\%$
(0, 0)	(1, 25)	0.0860	9.14	1.19	26.41
	(1, 1)	0.0739	8.90	0.99	26.96
	(25, 1)	0.0911	8.74	0.90	27.69
(1, 10)	(1, 25)	0.0865	8.90	1.01	27.14
	(1, 1)	0.0673	8.74	0.90	27.69
	(25, 1)	0.0792	8.66	0.85	28.05
(10, 1)	(1, 25)	0.0865	9.16	1.19	26.23
	(1, 1)	0.0735	8.90	0.99	26.96
	(25, 1)	0.0705	8.88	0.97	26.96

Se puede observar que con el criterio de decisión EHNF se rechazan más hipótesis nulas que con la regla bayes. Como consecuencia se rechazan más hipótesis nulas ciertas, debido a que las hipótesis nulas adicionales, que se rechazan con el criterio EHNF, tienen una probabilidad final estimada de ser ciertas mayor que 0.5, llegando incluso, algunas hipótesis, a tener una probabilidad final mayor que 0.8 ($prob_{\hat{N}_1} > 0.8$), obteniéndose así un porcentaje de falsos positivos mayor que el que se obtiene con la regla bayes, aunque se mantiene en unos niveles aceptables ($< 10\%$).

Igualmente, con el criterio de decisión EHNF, se rechazan más hipótesis nulas falsas no detectadas con la regla bayes, por tener una probabilidad final estimada de ser ciertas mayor que 0.5, por lo que se obtiene un porcentaje más bajo de falsos negativos; es decir, con el criterio de decisión EHNF se detectan más hipótesis nulas falsas, a la vez que el porcentaje de falsos positivos se mantiene en unos niveles

Tabla 3.14: Porcentaje estimado de hipótesis nulas falsas ($\widehat{N}_1\%$), probabilidad final estimada más alta con la que se rechaza cada hipótesis nula ($prob_{\widehat{N}_1}$) y porcentajes de falsos positivos ($FP_{EHNF}\%$) y falsos negativos ($FN_{EHNF}\%$), aplicando el criterio de decisión EHNF, para distintos valores de los parámetros de las distribuciones iniciales de $p \sim Beta(\alpha, \beta)$ y de $\phi \sim Gamma(a/2, b/2)$ y para las respectivas estimaciones finales de c .

(a, b)	(α, β)	\hat{c}	$\widehat{N}_1\%$	$prob_{\widehat{N}_1}$	$FP_{EHNF}\%$	$FN_{EHNF}\%$
(0, 0)	(1, 25)	0.0860	17.30	0.8280	9.35	18.21
	(1, 1)	0.0739	15.76	0.8283	7.68	18.76
	(25, 1)	0.0911	16.20	0.8290	8.13	18.21
(1, 10)	(1, 25)	0.0865	17.06	0.8315	9.08	18.03
	(1, 1)	0.0673	15.26	0.8319	7.19	19.13
	(25, 1)	0.0792	15.58	0.8304	7.44	18.40
(10, 1)	(1, 25)	0.0865	17.34	0.8266	9.39	18.21
	(1, 1)	0.0735	15.80	0.8266	7.73	18.76
	(25, 1)	0.0705	15.46	0.8277	7.37	18.94

aceptables, consiguiendo de este modo un mayor equilibrio entre los porcentajes de falsos positivos y falsos negativos.

Por tanto, el criterio de decisión EHNF es más adecuado que la regla bayes, cuando se toma $C_{0i} = C_{1i}$ para $i = 1, \dots, N$, siendo especialmente conveniente en el contexto de los experimentos con biochips de ADN, donde los contrastes múltiples se utilizan, en muchas ocasiones, como un primer paso exploratorio con el fin de identificar aquellos grupos de genes que potencialmente se expresan de manera diferencial, para posteriormente realizar con ellos un estudio más detallado. Así, se puede estar dispuesto a admitir un mayor número de falsos positivos, siempre y cuando su proporción sea admisible, para obtener el mayor número posible de genes de interés.

Tabla 3.15: Resultados aplicando el procedimiento de Benjamini y Hochberg (1995) (BH) y la aproximación Empírico Bayes con la regla bayes y con el criterio de decisión EHNF, con los parámetros iniciales $(\alpha, \beta) = (1, 1)$ y $(a, b) = (0, 0)$ y para el correspondiente $\hat{c} = 0.0739$.

	Método B-H		Regla bayes		Criterio EHNF	
	$\alpha = 0.05$	$\alpha = 0.1$				
$R_{BH}\%$	2.74	4.72	$R_C\%$	8.90	$\widehat{N}_1\%$	15.76
$FP_{BH}\%$	0.13	0.40	$FP_C\%$	0.99	$FP_{EHNF}\%$	7.68
$FN_{BH}\%$	76.14	60.29	$FN_C\%$	26.96	$FN_{EHNF}\%$	18.76

Con el fin de comparar la aproximación Empírico Bayes con el procedimiento de Benjamini y Hochberg (1995) para controlar el FDR, en la tabla 3.15 se muestran los resultados obtenidos mediante la aproximación Empírico Bayes, aplicando la regla bayes y el criterio de decisión EHNF, para las distribuciones iniciales $Beta(1, 1)$ y $Gamma(0, 0)$ de p y ϕ , respectivamente, y para el correspondiente valor estimado de c . En la misma tabla también se muestran los resultados obtenidos mediante el procedimiento de Benjamini y Hochberg (1995) para controlar el FDR a los niveles $\alpha = 0.05$ y $\alpha = 0.1$, esto es, el porcentaje de hipótesis nulas rechazadas ($R_{BH}\%$) y los porcentajes de falsos positivos ($FP_{BH}\%$) y falsos negativos ($FN_{BH}\%$) calculados como en (2.15), es decir,

$$FP_{BH}\% = \frac{\sum_{i=1}^N I(\tilde{p}_i \leq \alpha) I(\mu_i = 0)}{\sum_{i=1}^N I(\mu_i = 0)} \times 100,$$

Tabla 3.16: Resultados obtenidos mediante el procedimiento bayesiano, con la regla bayes y con el criterio de decisión EHNF, para los parámetros iniciales $(\alpha, \beta) = (1, 1)$ y $(a, b) = (0, 0)$ y para distintos valores de c .

	Regla bayes		Criterio EHNF		
	$c = 0.003$	$c = 0.0142$	$c = 0.0007$	$c = 0.00192$	
$R_C\%$	7.32	7.96	$\widehat{N}_1\%$	7.26	7.96
$FP_C\%$	0.13	0.40	$FP_{EHNF}\%$	0.13	0.40
$FN_C\%$	34.43	30.78	$FN_{EHNF}\%$	34.97	30.78

$$FN_{BH}\% = \frac{\sum_{i=1}^N I(\tilde{p}_i > \alpha) I(\mu_i \neq 0)}{\sum_{i=1}^N I(\mu_i \neq 0)} \times 100,$$

siendo $\tilde{p}_i = \min_{j=i, \dots, N} \left\{ \min \left(\frac{N}{j} p_{(j)}, 1 \right) \right\}$, para $i = 1, \dots, N$, los p-valores ajustados para este procedimiento.

Se puede observar que con el procedimiento de Benjamini y Hochberg (1995) para controlar el FDR se obtienen unos porcentajes muy pequeños de hipótesis nulas rechazadas, con unos porcentajes muy bajos de falsos positivos pero con unos porcentajes muy elevados de falsos negativos. Sin embargo, con la aproximación Empírico Bayes con cualquiera de los dos criterios de decisión, principalmente con el criterio de decisión EHNF, se obtiene un porcentaje de falsos negativos significativamente menor que el que se obtiene con el procedimiento de Benjamini y Hochberg (1995), al mismo tiempo que los porcentajes de falsos positivos se mantienen en unos niveles aceptables ($< 10\%$).

Además, en la tabla 3.16 se muestran los resultados obtenidos con el procedi-

miento bayesiano aplicando la regla bayes y el criterio de decisión EHNF, para las mismas distribuciones iniciales de p y ϕ que en la tabla anterior, pero para los valores de c con los que se obtienen unos porcentajes de falsos positivos iguales a los que se obtienen con el procedimiento de Benjamini y Hochberg (1995) para controlar el FDR a los niveles $\alpha = 0.05$ y $\alpha = 0.1$.

Se puede observar, comparando estos resultados con los obtenidos en la tabla 3.15 por el método de Benjamini y Hochberg (1995), que con el procedimiento bayesiano, con cualquiera de los dos criterios de decisión, se obtiene un número más alto de hipótesis nulas rechazadas que el que se obtiene con el procedimiento frecuentista. Este número adicional de hipótesis nulas rechazadas con el procedimiento bayesiano son necesariamente falsas, ya que con ambos procedimientos se obtiene el mismo porcentaje de falsos positivos. Por tanto, con el procedimiento bayesiano, con cualquiera de los dos criterios de decisión, se obtiene una reducción en el porcentaje de falsos negativos que puede llegar a ser muy significativa, como se ve en el ejemplo con datos simulados.

Así, podemos concluir que el procedimiento bayesiano resulta más conveniente que el procedimiento frecuentista ya que, obteniendo el mismo porcentaje de falsos positivos, mediante el procedimiento bayesiano se detectan más hipótesis nulas falsas, reduciendo así el porcentaje de falsos negativos.

3.6 Aplicación a datos reales procedentes de experimentos con biochips de ADN

En esta sección se aplica el procedimiento Empírico Bayes, descrito en la sección anterior, a los datos sobre cáncer de colon de Alon et al. (1999) para identificar genes con expresión diferencial. Alon et al. (1999) utilizaron biochips de oligonucleótidos de Affymetrix para estudiar el nivel de expresión de más de 6500 genes humanos, con 22 muestras de tejido normal de colon y 40 de tejido tumoral. Las muestras fueron tomadas de 40 pacientes diferentes, proporcionando 22 de estos pacientes una muestra de tejido normal y una de tejido tumoral. Alon et al. (1999) centraron el estudio en los 2000 genes con la mayor intensidad mínima. La información detallada está disponible en <http://www.stat.ucla.edu/wxl/research/microarray/DBC/index.htm> y en <http://microarray.princeton.edu/oncology/>.

Por tanto, la matriz de datos consta de 2000 filas (genes) y 62 columnas (muestras de tejido normal y tumoral), que en Alon et al. (1999) no están expuestas consecutivamente según el tipo de tejido. Nosotros organizamos esta matriz de manera que los datos de las 22 primeras columnas corresponden a las muestras de tejido normal y las 40 últimas columnas al tejido tumoral.

Con el fin de contrastar, para cada gen, si existen diferencias significativas entre el nivel de expresión en tejido normal y el nivel de expresión en tejido tumoral, para cada gen se utilizó el estadístico T_i , diferencia de medias entre tejido normal y tumoral, considerando el modelo Gaussiano descrito en la sección 3.3, con las distribuciones iniciales dadas en (3.11) con $c_i = c$, para $i = 1, \dots, 2000$. Para los parámetros de las distribuciones iniciales de $p \sim Beta(\alpha, \beta)$ y de $\phi \sim Gamma(a/2, b/2)$ se eligieron los mismos valores que con los datos simulados.

Tabla 3.17: Estimación final de c para distintos valores de los parámetros (α, β) y (a, b) . Estimación final de p y ϕ , porcentaje de genes con expresión diferencial aplicando la regla bayes ($R_C\%$) y aplicando el criterio de decisión EHNF ($\widehat{N}_1\%$) y probabilidad final estimada más alta con la que se rechaza cada hipótesis nula ($prob_{\widehat{N}_1}$) bajo el criterio EHNF, para los diferentes valores de los parámetros (α, β) y (a, b) y con las respectivas estimaciones finales de c , para los datos sobre cáncer de colon de Alon et al. (1999).

(a, b)	(α, β)	\hat{c}	\hat{p}	$\hat{\phi}$	$R_C\%$	$\widehat{N}_1\%$	$prob_{\widehat{N}_1}$
(0, 0)	(1, 25)	0.0040	0.72	0.00069	24.25	28.10	0.7594
	(1, 1)	0.0041	0.75	0.00059	21.95	24.90	0.7002
	(25, 1)	0.0041	0.76	0.00047	21.65	23.90	0.6737
(1, 10)	(1, 25)	0.0060	0.71	0.00065	24.40	28.60	0.7744
	(1, 1)	0.0056	0.74	0.00060	22.80	26.15	0.7078
	(25, 1)	0.0048	0.75	0.00058	21.90	24.60	0.6906
(10, 1)	(1, 25)	0.0040	0.71	0.00073	24.90	28.95	0.7754
	(1, 1)	0.0054	0.73	0.00063	23.55	26.50	0.7081
	(25, 1)	0.0047	0.75	0.00061	22.75	25.45	0.6882

En primer lugar, con objeto de estimar el parámetro c de la distribución inicial de las medias del modelo, se aplicó la aproximación Empírico Bayes descrita en la sección anterior con distintos valores de los parámetros (α, β) y (a, b) . Posteriormente, para los mismos valores de los parámetros (α, β) y (a, b) y para las respectivas estimaciones de c , se estimaron p , ϕ y la probabilidad final de cada hipótesis nula y, finalmente, para identificar los genes que presentan expresión diferencial, se aplicó el criterio de decisión EHNF y la regla bayes (3.8) considerando $C_{0i} = C_{1i}$ para $i = 1, \dots, 2000$.

En la tabla 3.17 se muestra, para los datos sobre cáncer de colon de Alon et al. (1999), la estimación final de c para los distintos valores considerados de los

parámetros (α, β) y (a, b) . También se muestran en esta tabla, para los mismos valores de los parámetros (α, β) y (a, b) y para las respectivas estimaciones finales de c , el porcentaje de genes con expresión diferencial obtenido aplicando la regla bayes ($R_C\%$) y aplicando el criterio de decisión EHNF ($\widehat{N}_1\%$), así como la probabilidad final estimada más alta con la que se rechaza cada hipótesis nula ($prob_{\widehat{N}_1}$) cuando se aplica el criterio de decisión EHNF. Se puede observar que, como sucedía con los datos simulados, los resultados son robustos respecto a la elección de los parámetros de las distribuciones iniciales.

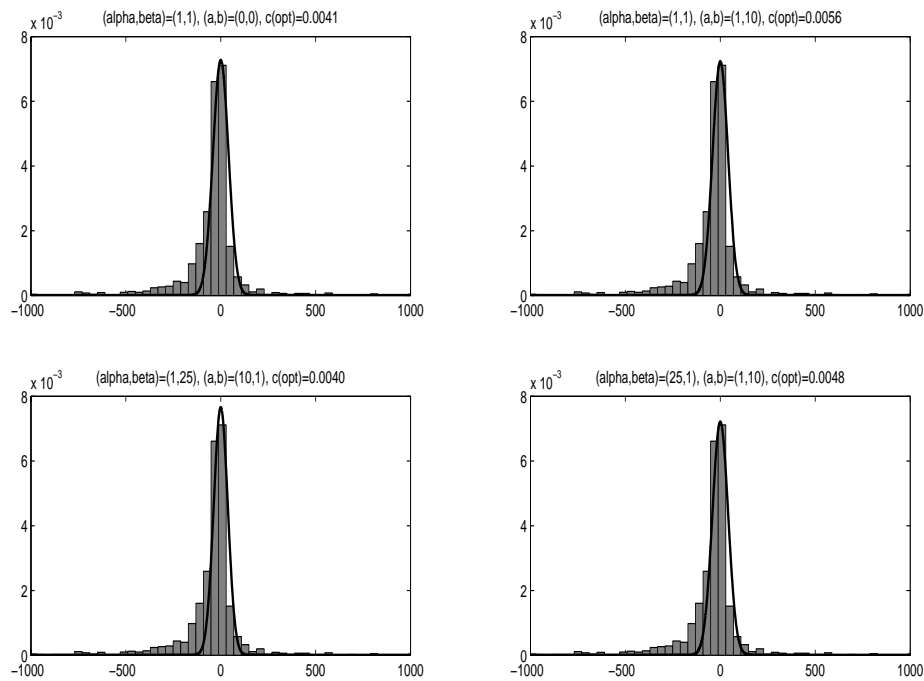
En la figura 3.1 se puede observar cómo el modelo ajusta los datos sobre cáncer de colon de Alon et al. (1999). Para distintos valores de los parámetros (α, β) y (a, b) de las respectivas distribuciones iniciales de p y ϕ y para los correspondientes valores estimados de c ; en cada gráfica se muestra el histograma de los datos junto con la densidad marginal ajustada $f(t_i|\hat{p}, \hat{\phi}, \hat{c})$, donde

$$f(t_i|\hat{p}, \hat{\phi}, \hat{c}) = \hat{p}f_0(t_i|\hat{\phi}) + (1 - \hat{p})f(t_i|\mu_i \neq 0, \hat{\phi}),$$

siendo $f_0(t_i|\hat{\phi}) \sim N(0, \hat{\phi}^{-1/2})$ y $f(t_i|\mu_i \neq 0, \hat{\phi}) = \int_{\mu_i \neq 0} f_1(t_i|\mu_i, \hat{\phi})\pi(\mu_i|\hat{\phi}, \hat{c})d\mu_i$, con $f_1(t_i|\mu_i, \hat{\phi}) \sim N(\mu_i, \hat{\phi}^{-1/2})$ y $\pi(\mu_i|\hat{\phi}, \hat{c})$ es la distribución dada en (3.11). Se puede ver fácilmente que $f(t_i|\mu_i \neq 0, \hat{\phi})$ es la distribución $N(0, (\hat{c}\hat{\phi}/(\hat{c} + 1))^{-1/2})$. Como se puede observar en la figura el modelo recoge bien las características básicas de los datos.

También se aplicó a estos datos el procedimiento de Benjamini y Hochberg (1995) para controlar el FDR, obteniendo un 6% y un 10.90% de genes con expresión diferencial para los niveles $\alpha = 0.05$ y $\alpha = 0.1$, respectivamente, mientras que con la aproximación Empírico Bayes, dependiendo de los valores de los parámetros de las distribuciones iniciales, se obtiene entre un 21.65% y un 24.90% de genes con expresión diferencial cuando se utiliza la regla bayes, y entre un 23.90% y un 28.95%

Figura 3.1: Histograma de las diferencias de medias entre tejido normal y tumoral de los datos sobre cáncer de colon de Alon et al. (1999). La curva es la densidad marginal ajustada, para diferentes valores de los parámetros.



de genes con expresión diferencial cuando se utiliza el criterio EHNf. Estos valores están muy próximos a los obtenidos por Do et al. (2005), que igualmente utilizaron esta base de datos para realizar un estudio mediante una aproximación bayesiana no paramétrica, y obtienen un 21.1% y un 28.95% de genes con expresión diferencial cuando controlan el FDR a los niveles 0.005 y 0.01, respectivamente.

Por tanto, la aproximación Empírico Bayes, principalmente con el criterio de decisión EHNf que nosotros proponemos, detecta más genes con expresión diferencial; lo que resulta deseable como se comentó en la sección anterior, mientras que el procedimiento de Benjamini y Hochberg (1995) para controlar el FDR es muy

conservador al respecto.

La aproximación bayesiana que se propone da lugar a un procedimiento, relativamente sencillo, para identificar genes con expresión diferencial. Además, al basar la inferencia en las probabilidades finales, se puede evaluar la influencia de cada gen. Por ejemplo, con los datos sobre cáncer de colon de Alon et al. (1999), hemos obtenido 223 genes con una probabilidad final de que se expresen diferencialmente igual a uno, por lo que se puede considerar que estos genes están altamente relacionados con el cáncer de colon.

3.7 Conclusiones y extensiones

Para contrastar simultáneamente N hipótesis del tipo

$$H_{0i} : \mu_i = 0 \quad \text{frente a} \quad H_{1i} : \mu_i \neq 0, \quad \text{para } i = 1, \dots, N,$$

se observan estadísticos T_i , para cada $i = 1, \dots, N$, en nuestro caso la media muestral, que se suponen independientes e idénticamente distribuidos de tal manera que

$$T_i | H_{0i} \sim N(0, 1/\sqrt{\phi}) \quad \text{y} \quad T_i | H_{1i} \sim N(\mu_i, 1/\sqrt{\phi}), \quad \mu_i \neq 0,$$

donde ϕ es la precisión desconocida y μ_i , con $i = 1, \dots, N$, los parámetros de interés.

Además, asumimos que la probabilidad inicial de H_{0i} es p para todo i . De esta manera el conjunto de parámetros del modelo es $\boldsymbol{\theta} = (p, \phi, \mu_1, \dots, \mu_N)$. La distribución inicial considerada para $\boldsymbol{\theta}$ es

$$\pi(\boldsymbol{\theta}) = \pi(p|\alpha, \beta) \pi(\phi|a, b) \prod_{i=1}^N \pi(\mu_i|\phi),$$

donde $\pi(p|\alpha, \beta)$, $\pi(\phi|a, b)$ y $\pi(\mu_i|\phi)$, para $i = 1, \dots, N$, son las distribuciones dadas en (3.11).

En este contexto el procedimiento que nosotros proponemos funciona aceptablemente bien. Además, podemos afirmar que este procedimiento es robusto respecto a la elección de los parámetros de las distribuciones iniciales propuestas, excepto respecto del parámetro c_i asociado a la varianza de las distribuciones iniciales de las medias μ_i , ver (3.11).

El problema de la sensibilidad respecto a c_i se ha resuelto estimando dicho parámetro por medio de una aproximación Empírico Bayes comprobando, mediante un ejemplo de simulación, que el valor estimado que se obtiene de este parámetro está dentro del rango de valores para los que se obtienen buenos resultados; es decir, para los que se obtienen las estimaciones de p y ϕ más próximas a los valores con los que se generaron los datos ($p = 0.9$ y $\phi = 5$).

Además, con esa estimación de c_i se obtienen los porcentajes más bajos de falsos negativos mientras que los porcentajes de falsos positivos se mantienen en unos niveles aceptables ($< 10\%$).

Debemos observar que la metodología Empírico Bayes implica el uso de una distribución inicial que depende de los datos observados. Sin embargo, como señala Casella (2001), las distribuciones iniciales dependientes de los datos son perfectamente válidas y su uso es frecuente en la literatura bayesiana.

En este capítulo también se comprueba, al igual que en el capítulo precedente, que el criterio de decisión EHNF, basado en la estimación del número de hipótesis nulas falsas, que nosotros proponemos es menos conservador que la regla bayes (tomando $C_{0i} = C_{1i}$ para $i = 1, \dots, N$); en el sentido de que se rechazan más hipótesis nulas.

Además, con nuestro criterio de decisión se alcanza un mayor equilibrio entre los

porcentajes de falsos positivos y falsos negativos que el que se consigue con la regla bayes, puesto que obtenemos un menor porcentaje de falsos negativos (ya que se detecta un mayor porcentaje de hipótesis nulas falsas), manteniéndose el porcentaje de falsos positivos en niveles aceptables ($< 10\%$). Ello resulta en general muy adecuado, especialmente en el contexto de los experimentos con biochips de ADN.

Por otro lado, con el criterio de decisión EHNF no tenemos que fijar los costes C_{0i} y C_{1i} , para $i = 1, \dots, N$, que son necesarios para poder aplicar la regla bayes y que no siempre son fáciles de fijar.

Con los datos simulados podemos concluir que nuestro procedimiento es significativamente menos conservador que el procedimiento de Benjamini y Hochberg (1995), con el que se obtiene un elevado porcentaje de falsos negativos cuando se controla el FDR a los niveles usualmente utilizados en la literatura.

Es más, eligiendo el parámetro c para obtener, con nuestro procedimiento, el mismo porcentaje de falsos positivos que con el procedimiento frecuentista, resulta que con nuestro procedimiento rechazamos un porcentaje mayor de hipótesis nulas, detectando así más hipótesis nulas falsas. Como consecuencia se obtiene un porcentaje de falsos negativos significativamente menor que con el procedimiento frecuentista de Benjamini y Hochberg (1995).

Por otro lado, hay que señalar que el coste computacional del procedimiento bayesiano es bajo. Por todo ello resulta más conveniente aplicar el procedimiento bayesiano en el contraste de hipótesis múltiples con cualquiera de los dos criterios de decisión, preferentemente con el nuestro, que el procedimiento frecuentista.

Finalmente, cabe destacar que nuestra aproximación da lugar a un procedimiento bayesiano relativamente sencillo de aplicar y útil para la identificación de genes con

expresión diferencial, como hemos puesto de manifiesto en la sección 3.6. Asimismo, es importante observar que permite evaluar la influencia de cada gen, debido a que la inferencia está basada en la probabilidad final de que cada gen se exprese diferencialmente.

Parte de los resultados de este capítulo han sido aceptados para su publicación, véase Ausín et al. (2010).

En esta memoria se asume un modelo normal para las observaciones, ya que los datos procedentes de experimentos con biochips de ADN son habitualmente normalizados, y distribuciones iniciales conjugadas. No obstante, la metodología propuesta puede extenderse a otros modelos diferentes, aunque probablemente con un mayor coste computacional.

Un campo a explorar es el de los modelos con distribuciones asimétricas para los datos, puesto que con frecuencia se presentan estas distribuciones de probabilidad en algunos contextos en los estudios con biochips de ADN.

Otra posibilidad es considerar para las observaciones la distribución potencial exponencial, ya que también en este tipo de datos se presentan distribuciones con colas más altas o menos altas que las de la distribución normal.

En esta memoria se ha asumido también, por simplicidad, que las varianzas son iguales para las observaciones, por lo que otra extensión natural sería considerar varianzas distintas, dado que esta característica suele también aparecer con alguna frecuencia en el contexto de los experimentos con biochips de ADN.

Apéndice 3A.1: Modelo Gaussiano. distribuciones finales condicionadas

Denotamos por $\mathbf{z}_{-i} = (z_1, \dots, z_{i-1}, z_{i+1}, \dots, z_N)$, $\mathbf{t}_{-i} = (t_1, \dots, t_{i-1}, t_{i+1}, \dots, t_N)$ y por $\boldsymbol{\theta}_{-\theta_i}$ el vector de parámetros $\boldsymbol{\theta} = (p, \phi, \mu_1, \dots, \mu_N)$ sin el parámetro indicado en el subíndice.

- *Distribución final condicionada de $z_i = 0$, dados los datos y el resto de parámetros*

$$\begin{aligned}
 Pr(z_i = 0 | \mathbf{t}, \mathbf{z}_{-i}, \boldsymbol{\theta}) &= \frac{\pi(\boldsymbol{\theta})f(\mathbf{t}, z_i = 0, \mathbf{z}_{-i} | \boldsymbol{\theta})}{\sum_{k=0}^1 \pi(\boldsymbol{\theta})f(\mathbf{t}, z_i = k, \mathbf{z}_{-i} | \boldsymbol{\theta})} \\
 &= \frac{\pi(\boldsymbol{\theta})Pr(z_i = 0 | p)f(t_i | z_i = 0, \phi, \mu_i)f(\mathbf{t}_{-i}, \mathbf{z}_{-i} | \boldsymbol{\theta}_{-\mu_i})}{\sum_{k=0}^1 \pi(\boldsymbol{\theta})Pr(z_i = k | p)f(t_i | z_i = k, \phi, \mu_i)f(\mathbf{t}_{-i}, \mathbf{z}_{-i} | \boldsymbol{\theta}_{-\mu_i})} \\
 &= \frac{Pr(z_i = 0 | p)f(t_i | z_i = 0, \phi, \mu_i)}{Pr(z_i = 0 | p)f(t_i | z_i = 0, \phi, \mu_i) + Pr(z_i = 1 | p)f(t_i | z_i = 1, \phi, \mu_i)} \\
 &= \frac{p\phi^{\frac{1}{2}}(2\pi)^{-\frac{1}{2}}e^{-\frac{\phi}{2}t_i^2}}{p\phi^{\frac{1}{2}}(2\pi)^{-\frac{1}{2}}e^{-\frac{\phi}{2}t_i^2} + (1-p)\phi^{\frac{1}{2}}(2\pi)^{-\frac{1}{2}}e^{-\frac{\phi}{2}(t_i - \mu_i)^2}} \\
 &= \left(1 + \frac{1-p}{p}e^{-\frac{\phi}{2}(\mu_i^2 - 2\mu_i t_i)}\right)^{-1}.
 \end{aligned}$$

- *Distribución final condicionada de p , dados los datos y el resto de parámetros*

$$f(p | \mathbf{t}, \mathbf{z}, \boldsymbol{\theta}_{-p}) = \frac{\pi(\boldsymbol{\theta})f(\mathbf{t}, \mathbf{z} | \boldsymbol{\theta})}{\int_0^1 \pi(\boldsymbol{\theta})f(\mathbf{t}, \mathbf{z} | \boldsymbol{\theta}) dp}$$

$$\begin{aligned}
 &\propto \pi(\boldsymbol{\theta}) p^{n_0} (1-p)^{n_1} \prod_{i:z_i=0} f(t_i|z_i=0, \phi, \mu_i) \prod_{i:z_i=1} f(t_i|z_i=1, \phi, \mu_i) \\
 &\propto \pi(p|\alpha, \beta) p^{n_0} (1-p)^{n_1} \\
 &\propto p^{\alpha-1} (1-p)^{\beta-1} p^{n_0} (1-p)^{n_1} \sim \text{Beta}(n_0 + \alpha, n_1 + \beta).
 \end{aligned}$$

- Distribución final condicionada de ϕ , dados los datos y el resto de parámetros

$$\begin{aligned}
 f(\phi|\mathbf{t}, \mathbf{z}, \boldsymbol{\theta}_{-\phi}) &= \frac{\pi(\boldsymbol{\theta}) f(\mathbf{t}, \mathbf{z}|\boldsymbol{\theta})}{\int_0^\infty \pi(\boldsymbol{\theta}) f(\mathbf{t}, \mathbf{z}|\boldsymbol{\theta}) d\phi} \\
 &\propto \pi(\boldsymbol{\theta}) p^{n_0} (1-p)^{n_1} \prod_{i:z_i=0} f(t_i|z_i=0, \phi, \mu_i) \prod_{i:z_i=1} f(t_i|z_i=1, \phi, \mu_i) \\
 &\propto \pi(\phi|a, b) \prod_{i=1}^N \pi(\mu_i|\phi) \prod_{i:z_i=0} f(t_i|z_i=0, \phi, \mu_i) \prod_{i:z_i=1} f(t_i|z_i=1, \phi, \mu_i) \\
 &\propto \phi^{\frac{a}{2}-1} e^{-\frac{b}{2}\phi} \prod_{i=1}^N \left(\phi^{\frac{1}{2}} e^{-\frac{\phi}{2} c_i \mu_i^2} \right) \prod_{i:z_i=0} \left(\phi^{\frac{1}{2}} e^{-\frac{\phi}{2} t_i^2} \right) \prod_{i:z_i=1} \left(\phi^{\frac{1}{2}} e^{-\frac{\phi}{2} (t_i - \mu_i)^2} \right) \\
 &= \phi^{\frac{a+2N}{2}-1} e^{-\frac{\phi}{2} \left(b + \sum_{i=1}^N c_i \mu_i^2 + \sum_{i:z_i=0} t_i^2 + \sum_{i:z_i=1} (t_i - \mu_i)^2 \right)} \\
 &\sim \text{Gamma} \left(\frac{a+2N}{2}, \frac{K}{2} \right),
 \end{aligned}$$

donde $K = b + \sum_{i=1}^N c_i \mu_i^2 + \sum_{i:z_i=0} t_i^2 + \sum_{i:z_i=1} (t_i - \mu_i)^2$.

- Distribución final condicionada de μ_i , dados los datos, $z_i = 0$ y el resto de parámetros

$$\begin{aligned}
 f(\mu_i|\mathbf{t}, z_i=0, \mathbf{z}_{-i}, \boldsymbol{\theta}_{-\mu_i}) &= \frac{\pi(\boldsymbol{\theta}) f(\mathbf{t}, z_i=0, \mathbf{z}_{-i}|\boldsymbol{\theta})}{\int_{\mu_i \neq 0} \pi(\boldsymbol{\theta}) f(\mathbf{t}, z_i=0, \mathbf{z}_{-i}|\boldsymbol{\theta}) d\mu_i} \\
 &\propto \pi(\boldsymbol{\theta}) Pr(z_i=0|p) f(t_i|z_i=0, \phi, \mu_i) f(\mathbf{t}_{-i}, \mathbf{z}_{-i}|\boldsymbol{\theta}_{-\mu_i}) \\
 &\propto \pi(\mu_i|\phi) f(t_i|z_i=0, \phi, \mu_i)
 \end{aligned}$$

$$\propto e^{-\frac{\phi}{2}c_i\mu_i^2} \sim N\left(0, \frac{1}{\sqrt{c_i\phi}}\right).$$

- Distribución final condicionada de μ_i , dados los datos, $z_i = 1$ y el resto de parámetros

$$\begin{aligned} f(\mu_i|\mathbf{t}, z_i = 1, \mathbf{z}_{-i}, \boldsymbol{\theta}_{-\mu_i}) &= \frac{\pi(\boldsymbol{\theta})f(\mathbf{t}, z_i = 1, \mathbf{z}_{-i}|\boldsymbol{\theta})}{\int_{\mu_i \neq 0} \pi(\boldsymbol{\theta})f(\mathbf{t}, z_i = 1, \mathbf{z}_{-i}|\boldsymbol{\theta}) d\mu_i} \\ &\propto \pi(\boldsymbol{\theta})Pr(z_i = 1|p)f(t_i|z_i = 1, \phi, \mu_i)f(\mathbf{t}_{-i}, \mathbf{z}_{-i}|\boldsymbol{\theta}_{-\mu_i}) \\ &\propto \pi(\mu_i|\phi)f(t_i|z_i = 1, \phi, \mu_i) \\ &\propto e^{-\frac{\phi}{2}c_i\mu_i^2}e^{-\frac{\phi}{2}(t_i - \mu_i)^2} \\ &\propto e^{-\frac{\phi}{2}(1+c_i)(\mu_i - \frac{t_i}{1+c_i})^2} \sim N\left(\frac{t_i}{1+c_i}, \frac{1}{\sqrt{(1+c_i)\phi}}\right). \end{aligned}$$

Apéndice 3A.2: Modelo Gaussiano jerárquico. distribución final condicionada

Distribución final condicionada de c_i , dados los datos y el resto de parámetros

$$\begin{aligned}
 f(c_i | \mathbf{t}, \mathbf{z}, \boldsymbol{\theta}_{-c_i}) &= \frac{\pi(\boldsymbol{\theta}) f(\mathbf{t}, \mathbf{z} | \boldsymbol{\theta})}{\int_0^\infty \pi(\boldsymbol{\theta}) f(\mathbf{t}, \mathbf{z} | \boldsymbol{\theta}) dc_i} \\
 &\propto \pi(\boldsymbol{\theta}) p^{n_0} (1-p)^{n_1} \prod_{i: z_i=0} f(t_i | z_i=0, \phi, \mu_i) \prod_{i: z_i=1} f(t_i | z_i=1, \phi, \mu_i) \\
 &\propto \pi(\mu_i | \phi, c_i) \pi(c_i | e, f) \\
 &\propto \left(c_i^{\frac{1}{2}} e^{-\frac{\phi}{2} c_i \mu_i^2} \right) c_i^{e-1} e^{-f c_i} \\
 &= c_i^{e+\frac{1}{2}-1} e^{-c_i(f+\frac{\phi}{2}\mu_i^2)} \sim \text{Gamma} \left(e + \frac{1}{2}, f + \frac{\phi}{2}\mu_i^2 \right).
 \end{aligned}$$

donde el conjunto de parámetros es ahora $\boldsymbol{\theta} = (p, \phi, \mu_1, \dots, \mu_N, \mathbf{c})$, siendo $\mathbf{c} = (c_1, \dots, c_N)$.

Bibliografía

- [1] Aitkin, M. (1969). Multiple comparisons in psychological experiments. *British J. Math.Statist. Psychology*, **22**, 193–198.
- [2] Alon, U. Barkai, N., Notterman, D. A., Gish, K., Ybarra, S., Mack, D. y Levine, A. J. (1999). Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. *Proc. Natn. Acad. Sci. USA* **96**, 6745–6750.
- [3] Ausín, M. C., Gómez-Villegas, M. A., González-Pérez, B, Rodríguez-Bernal, M. T., Salazar, I. y Sanz, L. (2010). Bayesian analysis of multiple hypothesis testing with applications to microarray experiments. *Aceptado en Commun. in Statistics-Theory and Methods*.
- [4] Baldi, P. y Long, A. D. (2001). A Bayesian framework for the analysis of microarray expression data: regularized t-test and statistical inferences of gene changes. *Bioinformatics*, **17**, 509–519.
- [5] Bauer, P., Hommel, G. y Sonnemann, E. (1988). *Multiple Hypothesenprüfung. (Multiple Hypotheses Testing)*. Berlin: Springer-Verlag (In German and English).
- [6] Benjamini, Y. y Hochberg, Y. (1995). Controlling the false discovery rate: a

- practical and powerful approach to multiple testing. *J. Roy. Statist. Soc. Ser. B* **57**, 289–300.
- [7] Benjamini, Y. y Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Ann. Statist.* **29**, 1165–1188.
- [8] Berger, J. (1985). *Statistical Decision Theory and Bayesian Analysis*. Second Ed., New York: Springer-Verlag.
- [9] Berger, J. (2006). The case for objective Bayesian analysis (with discussion). *Bayesian Analysis* **1**, 385–402.
- [10] Berry, D. A. y Hochberg, Y. (1999). Bayesian perspectives on multiple comparisons. *J. Statist. Plann. Inference*, **82**, 215–227.
- [11] Braun, H. I. (1994). *The Collected Works of John W. Tukey. Vol. VIII: Multiple Comparisons: 1948–1983*. New York: Chapman and Hall.
- [12] Cabras, S. (2010). A note on multiple testing for composite null hypotheses. *J. Statist. Plann. Inference*, **140**, 659–666.
- [13] Casella, G. (2001). Empirical Bayes Gibbs sampling. *Biostatistics*, **2**(4), 485–500.
- [14] Chew, V. (1976a). Comparing treatment means: A compendium. *Hort. Science*, **11**, 348–357.
- [15] De la Horra, J. (2007). Bayesian Robustness of the Positive False Discovery Rate. *Commun. in Statistics-Theory and Methods*, **36**, 1905–1914.
- [16] Diebolt, J. y Robert, C.P. (1994). Estimation of finite mixture distributions through Bayesian sampling. *Journal of the Royal Statistical Society*, **56**, 363–375.

-
- [17] Do, K-A, Müller, P. y Tang, F. (2005). A Bayesian mixture model for differential gene expression. *J. R. Stat. Soc. Ser. C*, **54**, 627–644.
- [18] Dudoit, S., Shaffer, J. P., y Boldrick, J. C. (2003). Multiple hypothesis testing in microarray experiments. *Stat. Sci.* **18** (1), 71–103.
- [19] Dudoit, S. y van der Laan, M. J. (2008). *Multiple Testing Procedures with Applications to Genomics*. New York: Springer.
- [20] Duncan, D. B. (1961). Bayes rules for a common multiple comparisons problem and related Student-t problems. *Ann. Math. Stat.*, **32**, 1013–1033.
- [21] Duncan, D. B. (1965). A Bayesian approach to multiple comparisons. *Technometrics*, **7**, 171–222.
- [22] Dunnett, C. W. (1970). Multiple comparisons. *Statistics in Endocrinology* (J. W. McArthur and T. Colton, eds.), Cambridge: MIT Press, 79–103.
- [23] Dunnett, C. W. y Goldsmith, C. H. (1981). When and how to do multiple comparisons. In *Statistics in Pharmaceutical Industry*, eds. C. R. Buncher and J. Y. Tsay, chapter 16, 397–434. New York: Marcel Dekker.
- [24] Efron, B. y Morris, C. (1972). Limiting the Risk of Bayes and Empirical Bayes Estimators. Part II: The Empirical Bayes Case. *Journal of the American Statistical Association*, **67**, 130–139.
- [25] Efron, B. y Morris, C. (1973). Stein's Estimation Rule and Its Competitors. An Empirical Bayes Approach. *Journal of the American Statistical Association*, **68**, 117–130.
- [26] Efron, B. y Morris, C. (1975). Data Analysis Using Stein's Estimator and Its Generalizations. *Journal of the American Statistical Association*, **70**, 311–319.

-
- [27] Efron, B., Tibshirani, R., Storey, J. D. y Tusher, V. (2001). Empirical Bayes Analysis of a Microarray Experiment. *Journal of the American Statistical Association*, **96**(456), 1151–1160.
- [28] Efron, B. (2004). Large-Scale Simultaneous Hypothesis Testing: The Choice of the Null Hypothesis. *J. Amer. Statist. Assoc.*, **99**(465), 96–103.
- [29] Efron, B. (2008). Microarrays, Empirical Bayes and the Two-Groups Model. *Statistical Science*, **23**, 1–22.
- [30] Games, P. A. (1971). Multiple comparisons of means. *Amer. Educ. Research J.*, **8**, 531–565.
- [31] Gavrilov, Y., Benjamini, Y. y Sarkar, S. K. (2009). An adaptive step-down procedure with proven FDR control under independence. *Ann. Statist.*, **37**, 619–629.
- [32] Gelfand, A. E. y Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *J. Amer. Statist. Assoc.*, **85**, 398–409.
- [33] Gelman, A., Carlin, J. B., Stern, H. S. y Rubin, D. B. (2004). *Bayesian Data Analysis*. London: Chapman and Hall/CRC.
- [34] GENOMA ESPAÑA/CIBT-FGUAM. Informe de Vigilancia Tecnológica. Microarrays y Biochips de ADN, (2002).
- [35] GENOMA ESPAÑA/CIBT-FGUAM. Informe de Vigilancia Tecnológica. Aplicaciones de los Microarrays y Biochips en salud humana, (2005).
- [36] Gill, J. L. (1973). Current status of multiple comparison of means in designed experiments. *J. Dairy Sci.*, **56**, 973–977.

-
- [37] Gottardo, R., Raftery, A. E., Yeung, K. Y. y Bumgarner, R. E. (2006). Bayesian Robust Inference for Differential Gene Expression in Microarrays with Multiple Samples. *Biometrics*, **62**, 10–18.
- [38] Harter, H. L. (1980). Early history of multiple comparison tests. In *Handbook of Statistics*, ed. PR Krishnaiah, **1**, 617–622. Amsterdam: North-Holland.
- [39] Hoppe, F. M. (1993b). *Multiple Comparisons, Selection, and Applications in Biometry*. New York: Dekker.
- [40] Hsu, J. C. (1996). *Multiple Comparisons: Theory and Methods*. New York: Chapman and Hall.
- [41] Hochberg, Y. y Tamhane, A. C. (1987). *Multiple Comparison Procedures*. New York: John Wiley.
- [42] Ibrahim, J. G., Ming-Hui Chen, M-H y Gray, R. J. (2002). Bayesian models for gene expression with DNA microarray data. *J. Am. Stat. Assoc.*, **97**, 88–99.
- [43] Jin, J. y Cai, T. (2007). Estimating the null and the proportion of nonnull effects in large-scale multiple comparisons. *J. American Statistical Association*, **102**, 495–506.
- [44] Kendzioriski, C., Newton, M., Lan, H., y Gould, M. N. (2003). On parametric empirical Bayes methods for comparing multiple groups using replicated gene expression profiles. *Statistics in Medicine*, **22**, 3899–3914.
- [45] Klockars, A. J. y Sax, G. (1986). *Multiple Comparisons*. Sage University Paper series on Quantitative Applications in the Social Sciences, 07–061. Newbury Park, CA: Sage.

-
- [46] Lehmann, E. L. (1957a,b). A theory of some multiple decision problems, (Parts I and II). *Annals of Mathematical Statistics*, **28**, 1–25 and 547–572.
- [47] Lehmann, E.L. y Romano, J. P. (2005). *Testing Statistical Hypotheses*. (Third Ed.), London: Springer.
- [48] Lewis, C. y Thayer, D. T. (2004). A loss function related to the FDR for random effects multiple comparisons. *J. Statist. Plann. Inference*, **125**, 49–58.
- [49] Lönnstedt, I. y Speed, T. P. (2002). Replicated microarray data. *Statistica Sinica*, **12**, 31–46.
- [50] Lönnstedt, I. y Britton, T. (2005). Hierarchical Bayes models for cDNA microarray gene expression. *Biostatistics*, **6**, 279–291.
- [51] Meinshausen, N. y Rice, J. (2006). Estimating the proportion of false null hypotheses among a large number of independently tested hypotheses. *Annals of Statistics*, **34**(1), 373–393.
- [52] Miller, R. G. (1966). *Simultaneous Statistical Inference*. New York: Wiley.
- [53] Miller, R. G. (1977). Developments in multiple comparisons 1966–1976. *J. Amer. Statist. Assoc.*, **72**, 779–788.
- [54] Miller, R. G. (1981). *Simultaneous Statistical Inference*. (Second Ed.), New York: Wiley.
- [55] Miller, R. G. (1985). Multiple comparisons. *Encyclopedia of Statistical Sciences* (S. Kotz and N. L. Johnson, eds.), New York: Wiley, 679–689.
- [56] Morris, C. (1983). Parametric Empirical Bayes Inference: Theory and Applications (with discussion). *Journal of the American Statistical Association*, **78**, 47–65.

-
- [57] Newton, M. C., Kendzierski, C. M., Richmond, C. S., Blattner, F. R., y Tsui, K. W. (2001). On differential variability of expression ratios: Improving statistical inference about gene expression changes from microarray data. *Journal of Computational Biology*, **8**, 37–52.
- [58] Newton, M., Noueir, A., Sarkar, D., y Ahlquist, P. (2004). Detecting differential gene expression with a semiparametric hierarchical mixture method. *Biostatistics*, **5**, 155–176.
- [59] O’Neill, R. T. y Wetherill, B. G. (1971). The present state of multiple comparisons methods (with discussion). *J. Roy. Statist. Soc. B*, **33**, 218–241.
- [60] Robbins, H. (1955). An Empirical Bayes Approach to Statistics. *Proceedings of the Third Berkeley Symposium Mathematical Statistics and Probability* 1, Berkeley: University of California Press, 157–164.
- [61] Robert, C. P. y Casella, G. (2004). *Monte Carlo statistical methods*. (Second Ed.), New York: Springer.
- [62] Ryan, T. A. (1959). Multiple comparisons in psychological research. *Psychol. Bull.*, **56**, 26–47.
- [63] Schena, M., Shalon, D., Davis, R. W. y Brown, P. O. (1995). Quantitative Monitoring of Gene Expression Patterns with a Complementary DNA Microarray. *Science*, **270**, 467–470.
- [64] Scott, J. G., y Berger, J. O. (2006). An exploration of aspects of Bayesian multiple testing. *Journal of Statistical Planning and Inference*, **136**, 2144–2162.
- [65] Shaffer, J. P. (1986b). Simultaneous testing. In *Encyclopedia of Statistical Sciences*, eds. S. Kotz and N. L. Johnson. New York: Wiley.

-
- [66] Shaffer, J. P. (1988). Simultaneous testing. In *Encyclopedia of Statistical Sciences*, eds. S. Kotz and N. L. Johnson, **8**, 484–490. New York: Wiley.
- [67] Shaffer, J. P. (1995). Multiple hypothesis testing: A review. *Annual Review of Psychology*, **46**, 561–584.
- [68] Shaffer, J. P. (1999). A semi-Bayesian study of Duncan’s Bayesian multiple comparison procedure. *J. Statist. Plann. Inference*, **82**, 197–213.
- [69] Smith, A. F. M. y Roberts, G. O. (1993). Bayesian computation via the Gibbs sampler and related Markov Chain Monte Carlo Methods. *Journal of the Royal Statistical Society, Ser. B*, **55**, 3–23.
- [70] Spjøtvoll, E. (1972a). On the optimality of some multiple comparison procedures. *Ann. Math. Statist.*, **43**, 398–411.
- [71] Spjøtvoll, E. (1974). Multiple testing in the analysis of variance. *Scand. J. Statist.*, **1**, 97–114.
- [72] Storey, J. D. y Tibshirani, R. (2001). Estimating false discovery rates under dependence, with applications to DNA microarrays. *Technical Report*, 2001–28. Department of Statistics, Stanford University.
- [73] Storey, J. D. (2003). The positive false discovery rate: a Bayesian interpretation and the q-value. *Ann. Statist.*, **31**, 2013–2035.
- [74] Storey, J. D. (2007). The optimal discovery procedure: a new approach to simultaneous significance testing. *Journal of the Royal Statistical Society, Ser. B*, **69**, 347–368.

-
- [75] Storey, J. D., Dai, J. Y. y Leek, J. T. (2007). The optimal discovery procedure for large-scale significance testing, with applications to comparative microarray experiments. *Biostatistics*, **8**, 414–432.
- [76] Thomas, D. A. H. (1973). Multiple comparisons among means, A review. *The Statistician*, **22**, 16–42.
- [77] Toothaker, L.E. (1991). *Multiple Comparisons for Researchers*. Newbury Park, CA: Sage.
- [78] Toothaker, L.E. (1993). *Multiple Comparison Procedures*. Newbury Park, CA: Sage.
- [79] Tukey, J. W. (1953). *The Problem of Multiple Comparisons*. Mimeographed monograph.
- [80] Westfall, P. H. y Young, S. S. (1993). *Resampling-Based Multiple Testing: Examples and Methods for p-Value Adjustment*. New York: Wiley.
- [81] Yekutieli, D. y Benjamini, Y. (1999). Resampling-based false discovery rate controlling multiple test procedures for correlated test statistics. *J. Statist. Plann. Inference*, **82**, 171–196.