



UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERÍA MATEMÁTICA

TEORÍA DE PRUEBAS DE MÚLTIPLES HIPÓTESIS ESTADÍSTICAS Y APLICACIONES

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL MATEMÁTICO

ANDRÉS MARCELINO ITURRIAGA JOFRÉ

PROFESOR GUÍA
RAÚL GOUET BAÑARES

MIEMBROS DE LA COMISIÓN
NANCY LACOURLY VENTRE
JAIME SAN MARTÍN ARISTEGUI

SANTIAGO DE CHILE
JULIO 2011

RESUMEN DE LA MEMORIA
PARA OPTAR AL TÍTULO DE
INGENIERO CIVIL MATEMÁTICO
POR: ANDRÉS ITURRIAGA J.
FECHA: JULIO 2011
PROF. GUÍA: Sr. RAÚL GOUET B.

“TEORÍA DE PRUEBAS DE MÚLTIPLES HIPÓTESIS ESTADÍSTICAS Y APLICACIONES”

Esta memoria trata acerca de la teoría de pruebas de múltiples hipótesis estadísticas, tema que está a la vanguardia de la investigación estadística actual con importantes aplicaciones en detección de fuentes en astronomía, estudio de neuro-imágenes y análisis de microarrays. Esta metodología surge cuando se quiere contrastar simultáneamente varias hipótesis, dando una protección adicional a la probabilidad de rechazar hipótesis nulas verdaderas.

Este trabajo comienza con un breve repaso de la teoría inferencial frecuentista y bayesiana clásica, procurando establecer un marco de referencia común para posteriormente exponer los fundamentos de los test de hipótesis múltiples, abreviado THM. Luego se realiza una revisión general de las principales ideas detrás de los THM, con énfasis en los trabajos posteriores al de Benjamini y Hochberg en 1995, en el cual se desarrolla el importante concepto de tasa de falsos positivos, abreviado FDR en inglés. La relevancia de este último artículo se debe a que la FDR ha dado lugar a procedimientos estadísticos que se adecúan mejor a problemas en los cuales se dispone de grandes bases de datos con miles e incluso millones de variables que se desean analizar simultáneamente.

Posteriormente se discuten algunas soluciones propuestas recientemente para resolver THM, desde una perspectiva frecuentista primero, una bayesiana después y finalmente, desde una perspectiva computacional. Las metodologías escogidas fueron aquellas que se caracterizan por sus interesantes desarrollos matemáticos y el impacto provocado en la comunidad científica, medido a través de la cantidad de citas documentadas por la base de datos MathSciNet.

Finalmente, inspirados en un artículo reciente de Rudloff y Karatzas del año 2010 sobre la construcción de test de hipótesis no múltiples optimales, se usan ideas de análisis convexo para proponer un resultado de existencia de THM optimales, en el contexto de hipótesis nulas y alternativas compuestas. Se destaca la ausencia de resultados como este en la literatura estadística explorada durante el desarrollo de esta memoria.

AGRADECIMIENTOS

Quiero comenzar agradeciendo el apoyo financiero otorgado por el proyecto Fondecyt 1090216 para el desarrollo de esta memoria y a los profesores de mi comisión Nancy Lacourly y Jaime San Martín, quienes muy gentilmente accedieron a formar parte de este proyecto. También quiero expresar toda mi gratitud y afecto a mi profesor guía Raúl Gouet, quien me ayudó durante todo el desarrollo de este trabajo. Su apoyo fue fundamental para dar forma y terminar esta memoria, además de hacer nacer en mí el gusto por la inferencia estadística clásica. De todo corazón muchas gracias por todo.

Deseo agradecer también a toda la gente que pasó por mi vida durante estos años de estudio y que de una a otra forma contribuyeron a mi crecimiento profesional y personal. Palabras especiales tengo para mis amigos de generación, con quienes viví momentos inolvidables. De estos, en especial quiero agradecer a Héctor Olivero, quien sin llegar a pensarlo ni quererlo se transformó en el mejor amigo de todos, en mi mejor amigo. Muchas gracias por ser mi “brou”.

De la manera más especial de todas y con todo el amor del mundo quiero agradecer a mis padres Mónica Jofré y Andrés Iturriaga. A ustedes les debo absolutamente todo lo que soy y mis logros son los suyos. Solamente nosotros sabemos cuanto costó, pero al final lo logramos papi y mami, y ahora su hijo por fin es ingeniero. Simplemente gracias por creer siempre en mí y principalmente por ser los mejores padres que podría haber tenido. Los amo con todo mi corazón y esta memoria es para ustedes.

Finalmente quiero agradecerle al destino la oportunidad que me dio de conocer al amor de mi vida, Gabriela Sandoval, quien ha sido mi compañera durante estos años. Comenzaste siendo mi amiga, después mi polola y ahora eres mi amada esposa y la madre de nuestro hijo que viene en camino. Gracias simplemente por existir y darme la fuerza de transitar por este, a veces, pedregoso camino al que llamamos vida.

Andrés Iturriaga Jofré.
Julio, 2011.

Índice general

1. Introducción	1
1.1. Marco estructural de los test de hipótesis	2
1.2. Test de hipótesis	3
1.3. Teoría de Neyman-Pearson	4
1.4. El paradigma bayesiano	6
1.5. Test de hipótesis múltiples	9
1.6. Hipótesis simultáneas	12
1.7. Desarrollo histórico y revisión bibliográfica	14
1.8. Aplicaciones	21
2. Test de Hipótesis Múltiples, enfoque Frecuentista	22
2.1. FDR y el cambio de paradigma de los THM	22
2.2. El problema de la dependencia en los THM	24
2.3. Teoría de decisión múltiple: el principio de partición	26
2.4. La FDP vista como proceso estocástico	32
2.5. Extendiendo la FWER	36
3. Test de Hipótesis Múltiples, enfoque Bayesiano	40
3.1. La versión bayesiana de la FDR	40
3.2. Exploración de aspectos bayesianos de los THM	44
3.3. Enfoque bayesiano no paramétrico de los THM	47
3.4. Teoría de decisión múltiple: versión bayesiana	50
4. Aplicaciones Computacionales	53
4.1. El programa R	53
4.2. El paquete “multtest”	54
4.3. Aplicación: datos de microarray	56
4.4. Comentarios finales y recomendaciones	66
5. Análisis Convexo aplicado a los Test de Hipótesis Múltiples	67
5.1. Enunciado del Teorema	67
5.2. Demostración del Teorema	69
5.3. Limitaciones y alcances del Teorema	76

6. Conclusiones y Trabajo Futuro

78

Bibliografía

80

NOTACIÓN

Ω	: Espacio muestral abstracto.
\mathcal{F}	: σ -álgebra de conjuntos sobre Ω .
\mathcal{X}	: Espacio muestral asociado al conjunto de datos.
\mathcal{A}	: σ -álgebra de conjuntos sobre \mathcal{X} .
X, θ	: Elementos aleatorios.
$\mathcal{P}(\mathcal{A})$: Conjunto de las medidas de probabilidad sobre $(\mathcal{X}, \mathcal{A})$.
P_X	: Ley de X .
\mathcal{P}	: Modelo estadístico.
P_θ	: Medida de probabilidad indexada por parámetro θ .
Θ	: Espacio de parámetros.
\mathcal{T}	: σ -álgebra de conjuntos sobre Θ .
Θ_0^i	: Espacio de parámetros asociado a la hipótesis nula del test i .
Θ_1^i	: Espacio de parámetros asociado a la hipótesis alternativa del test i .
θ_X	: Verdadero parámetro asociado a X .
$\dim(\Theta)$: Dimensión del espacio Θ .
\mathbb{R}	: Conjunto de los números reales.
\mathbb{R}_+	: Conjunto de los números reales positivos.
$\mathcal{B}(A)$: σ -álgebra de los borelianos asociados al conjunto A .
$\frac{dP}{dR}$: Derivada de Radon-Nikodym de la medida P con respecto a R .
ϕ, φ	: Reglas de decisión.
H_0^i, H_1^i	: Hipótesis nula y alternativa respectivamente, del test i .
$\alpha(\theta)$: Probabilidad de rechazo en θ .
α	: Nivel de significancia de un test.
π	: Función de potencia.
UMP	: Uniformemente más potente.
$\mathcal{R}(\alpha)$: Región de rechazo de un test al nivel α .
\hat{p}	: p -valor.
Π	: Medida sobre (Θ, \mathcal{T}) .
K	: Kernel probabilista.
$\mathcal{A} \otimes \mathcal{T}$: σ -álgebra generada por $\mathcal{A} \times \mathcal{T}$.
\mathcal{Q}	: Medida de probabilidad sobre $(\mathcal{X} \times \Theta, \mathcal{A} \otimes \mathcal{T})$.
$p, p(\cdot \theta)$: Densidad de probabilidad y densidad condicional respectivamente.
$\pi, \pi(\cdot x)$: Densidad a priori y densidad a posteriori respectivamente.
B	: Factor bayesiano.
$L(\cdot, \cdot)$: Función de pérdida.
\mathcal{D}	: Espacio de decisiones.
\ln	: Logaritmo en base e .
E_θ	: Valor esperado con respecto a la medida P_θ .
$\mathcal{N}(\mu, \sigma^2)$: Ley normal de media μ y varianza σ^2 .

Var, Cov	:	Operadores varianza y covarianza respectivamente.
THM	:	Test de hipótesis múltiple.
FWER	:	Family-wise error rate.
FDR	:	False discovery rate.
pFDR	:	Positive false discovery rate.
FDP	:	False discovery proportion.
m	:	Número total de test en un THM.
R	:	Número de hipótesis rechazadas en un THM.
Q	:	Número de hipótesis aceptadas en un THM.
V	:	Número de hipótesis incorrectamente rechazadas en un THM.
T	:	Número de hipótesis incorrectamente aceptadas en un THM.
U	:	Número de hipótesis correctamente aceptadas en un THM.
S	:	Número de hipótesis correctamente rechazadas en un THM.
m_0	:	Número de hipótesis nulas verdaderas en un THM.
d	:	Proporción de hipótesis nulas verdaderas en un THM.
F	:	Función de distribución acumulada.
\mathfrak{H}_i	:	Indicador de si la i -ésima hipótesis es verdadera o no.
L^1	:	Espacio de las funciones integrables.
L^p	:	Espacio de las funciones a la p integrables.
L^∞	:	Espacio de las funciones acotadas.
\subseteq	:	Relación de ser subconjunto o igual.
\subset	:	Relación de ser subconjunto estricto.

Introducción

En esta memoria se estudia la teoría de los test de hipótesis múltiples¹ (THM), tema en torno al cual se ha desarrollado mucha investigación, sobre todo en los últimos 16 años, como consecuencia del trabajo de Benjamini y Hochberg (ver [4]), y que según la opinión de connotados estadísticos, está a la vanguardia de la exploración estadística actual. En palabras de Efron (ver [21]): “*Estamos en la era de la producción masiva de datos y preguntas complejas*”, para cuyas respuestas se necesitan nuevas metodologías y teorías. Estos nuevos desafíos han ido de la mano del avance tecnológico, especialmente en genómica, con técnicas como el microarray², que obligan al investigador a analizar miles, e incluso millones, de test de hipótesis de manera simultánea.

El principal objetivo de esta memoria es presentar un estado del arte del tema en cuestión, poniendo énfasis en los aspectos matemáticos de la teoría. Es relevante destacar que existe una vasta literatura respecto de los THM, la cual está dirigida principalmente a expertos o a personas que necesitan centrarse en las aplicaciones, por lo que el material disponible deja en segundo plano, y en algunos casos hasta olvida, los aspectos matemáticos de las distintas metodologías. Dada la gran cantidad de información existente se hizo necesario filtrar algunas de las publicaciones. El criterio utilizado fue incluir las más importantes, lo que se midió a través de la cantidad de citas documentadas por MathSciNet, y las que fueran innovadoras en cuanto a la aplicación de metodologías matemáticas. Un objetivo secundario de esta memoria es realizar alguna contribución de tipo teórica para determinar THM que tengan alguna propiedad de optimalidad, que como se verá más adelante, es uno de los puntos más criticables de la teoría a presentar.

La composición de este documento es la siguiente: en la sección 1.1 se plantea el modelo matemático de interés para este trabajo, en las secciones 1.2, 1.3 se expone brevemente la teoría de los test de hipótesis frecuentistas, siguiendo el texto de Lehmann y Romano (ver [44]), en la sección 1.4 se describe la contraparte bayesiana de las secciones 1.2, 1.3 y en las secciones 1.5 a 1.8 se explica en términos generales la teoría, desarrollo histórico

¹También denominado pruebas o contrastes de múltiples hipótesis estadísticas.

²Es una técnica utilizada especialmente en biología molecular para analizar la expresión diferencial de genes, visualizándose los niveles de miles de ellos de forma simultánea.

y principales aplicaciones de las pruebas de múltiples hipótesis estadísticas, poniendo en evidencia una dificultad inherente a este tipo de test, que se denomina “el problema de la multiplicidad”. En los capítulos 2 y 3 se presenta el detalle matemático de los contrastes de hipótesis simultáneos desde la óptica frecuentista y bayesiana respectivamente. En el capítulo 4 se muestran algunas implementaciones computacionales existentes orientadas a resolver THM. Finalmente, en el capítulo 5 se propone, usando ideas de análisis convexo, una nueva manera de abordar el problema de encontrar THM óptimas.

1.1 Marco estructural de los test de hipótesis

En esta sección se plantea el modelo matemático para la inferencia estadística que se utilizará en esta memoria, fijando de paso la notación a utilizar. Los elementos son los siguientes: sean (Ω, \mathcal{F}) , $(\mathcal{X}, \mathcal{A})$ espacios medibles, $X : \Omega \rightarrow \mathcal{X}$ un elemento aleatorio³ (i.e., una función \mathcal{F}, \mathcal{A} medible) y $\mathcal{P}(\mathcal{A})$ el espacio de las medidas de probabilidad definidas sobre $(\mathcal{X}, \mathcal{A})$. En este contexto el problema inferencial es el siguiente: se supone que X es un observable que sigue una ley P_X , que es desconocida, pero que pertenece a un conjunto $\mathcal{P} \subseteq \mathcal{P}(\mathcal{A})$, denominado **modelo estadístico**. Más aún, se considera a los elementos de \mathcal{P} indexados por un parámetro θ , de manera que

$$\mathcal{P} = \{P_\theta : \theta \in \Theta\}, \quad (1.1)$$

donde el conjunto Θ es denominado **espacio de parámetros** del modelo estadístico. En relación a lo anterior, un principio en el que se sustenta la teoría inferencial estadística es que las realizaciones de X proveen información respecto de su distribución de probabilidad. Así, bajo el supuesto de **identificabilidad** del modelo estadístico, es decir, que la función $\theta \rightarrow P_\theta$ sea inyectiva, al observar realizaciones de X se obtiene información respecto del verdadero valor de θ , denotado θ_X , el cual por definición satisface la relación $P_{\theta_X} = P_X$. El objetivo de la inferencia estadística es entonces decidir, en base a realizaciones de X , acerca del parámetro θ_X .

En términos generales, de acuerdo a la dimensión del espacio de parámetros se distinguen dos corrientes estadísticas denominadas respectivamente paramétrica y no paramétrica. En la primera, $\dim(\Theta) < \infty$ y en la segunda, $\dim(\Theta) = \infty$. Lo anterior provoca que el problema de determinar el parámetro θ_X sea de una naturaleza matemática bastante diferente. Un ejemplo de modelo paramétrico es el siguiente:

Ejemplo 1.1. Sea $\mathcal{X} = \mathbb{R}$, $\mathcal{A} = \mathcal{B}(\mathbb{R})$ (borelianos de \mathbb{R}) y \mathcal{P} la familia de leyes normales de media μ y varianza σ^2 , esto es,

$$\mathcal{P} = \left\{ P_\theta : \frac{dP_\theta}{dx}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[-\frac{(x - \mu)^2}{2\sigma^2} \right], \theta = (\mu, \sigma) \in \Theta := \mathbb{R} \times \mathbb{R}_+ \right\}, \quad (1.2)$$

³ X puede ser un vector, una matriz o una función aleatoria (proceso).

donde $\frac{dP_\theta}{dx}$ es la derivada de Radon-Nikodym (densidad) de P_θ con respecto a la medida de Lebesgue. El problema inferencial es entonces determinar el verdadero parámetro θ_X utilizando realizaciones de X . Notar que en este caso la familia \mathcal{P} está completamente determinada salvo los parámetros μ y σ . \parallel

Un ejemplo de problema no paramétrico es el siguiente:

Ejemplo 1.2. Sea $\mathcal{X} = \mathbb{R}$, $\mathcal{A} = \mathcal{B}(\mathbb{R})$ y $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ donde

$$\Theta = \left\{ \theta : \theta(x) = \frac{dP_\theta}{dx}(x) \text{ c.s.}, \int_{\mathbb{R}} (\theta''(x))^2 dx \leq c^2 \right\}, \quad (1.3)$$

con $c \in \mathbb{R}$ (constante). En palabras, Θ corresponde a la familia de todas las medidas de probabilidad sobre $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ que son absolutamente continuas con respecto a la medida de Lebesgue y cuyas derivadas de Radon-Nikodym (densidades) no son tan onduladas. El problema inferencial en este caso es determinar la función de densidad correspondiente, usando realizaciones de X . \parallel

Se hace evidente de los ejemplos 1.1 y 1.2 que en los problemas no paramétricos el espacio de parámetros es de una complejidad mayor. Lo anterior provoca el desarrollo de metodologías, por definición, muy diferentes. El lector interesado en una exposición moderna de estadística no paramétrica puede consultar [83].

En la sección siguiente se exponen brevemente los fundamentos de los test de hipótesis, el cual es uno de los temas más importantes de la inferencia estadística.

1.2 Test de hipótesis

Antes de definir lo que es un test de hipótesis se debe precisar lo que se entiende por una hipótesis estadística. Como ya se señaló en la sección anterior, el problema inferencial consiste en determinar P_X . Así, una hipótesis estadística es cualquier afirmación respecto de la ley de X , del estilo $P_X \in \mathcal{P}_0$ donde $\mathcal{P}_0 \subseteq \mathcal{P}$, que por la parametrización (1.1) se traduce en una afirmación sobre θ_X de la forma $\theta_X \in \Theta_0$ con $\Theta_0 \subseteq \Theta$.

Ejemplo 1.3. Siguiendo con el ejemplo 1.1, una hipótesis estadística puede ser que el parámetro μ sea estrictamente positivo, en cuyo caso $\Theta_0 = \{(\mu, \sigma) \in \Theta : \mu > 0\} = \mathbb{R}_+^2$. \parallel

Dado Θ_0 , las hipótesis $H_0 : \theta_X \in \Theta_0$ y $H_1 : \theta_X \in \Theta_1$ se denominan respectivamente hipótesis nula y alternativa, donde $\Theta_1 := \Theta \setminus \Theta_0$, y el problema de contrastar hipótesis estadísticas consiste en utilizar realizaciones de X para decidir entre alguna de ellas. Así, un test de hipótesis no aleatorizado ϕ , se define como una función medible

$$\phi : \mathcal{X} \rightarrow \{0, 1\} \quad (1.4)$$

tal que si $\phi(X) = 1$ entonces se rechaza H_0 y si $\phi(X) = 0$ entonces se acepta H_0 . En este último caso, la terminología estadística usual es afirmar que no hay suficiente evidencia para

rechazar la hipótesis nula, en lugar de decir que se acepta H_0 . Aunque en lo que sigue sólo se hablará de test de hipótesis no aleatorizados, para efectos de futuros resultados que se presentarán en esta memoria es importante definir lo que se entiende por un test de hipótesis aleatorizado. Este consiste en una función medible

$$\phi : \mathcal{X} \rightarrow [0, 1] \quad (1.5)$$

tal que $\phi(X)$ indica la probabilidad con la que se rechaza la hipótesis nula. El lector interesado en una exposición más detallada acerca de los test de hipótesis aleatorizados puede consultar [44].

Tras la definición de lo que se entiende por un test de hipótesis surge, de manera inmediata, el problema de tener que determinar la calidad de un test y escoger el óptimo, de acuerdo a cierto criterio, en caso de existir. En relación a esto, es importante tener presente que hay dos tipos de error inherentes a la aplicación de cualquier contraste de hipótesis. Estos se denominan **error de tipo I** y **II**, y corresponden a “rechazar H_0 cuando esta es verdadera” y “no rechazar H_0 cuando esta es falsa”, respectivamente⁴. Cabe destacar que dada la aleatoriedad asociada a las realizaciones de X , lo usual es controlar la probabilidad con la cual los errores de tipo I y II ocurren.

Dado un diseño muestral con un tamaño de muestra fijo, no es posible escoger un test de hipótesis que minimice de manera simultánea ambas probabilidades de error. En la sección siguiente se expone brevemente la teoría de Neyman-Pearson, la cual propone una solución a este problema.

1.3 Teoría de Neyman-Pearson

Esta teoría propone controlar la probabilidad de error de tipo I, manteniéndola bajo un nivel predeterminado. Dado un test ϕ , se define la probabilidad de rechazo como

$$\alpha(\theta) = P_\theta(\phi = 1) = E_\theta[\phi], \quad \theta \in \Theta. \quad (1.6)$$

Sea $\alpha \in [0, 1]$, se dice que un test ϕ tiene nivel α si

$$\sup_{\theta \in \Theta_0} \alpha(\theta) \leq \alpha. \quad (1.7)$$

Típicamente se impone que α tome un valor pequeño, de donde se aprecia el tratamiento diferente que, en esta teoría, se le da a la probabilidad de error de tipo I y II respectivamente, siendo más grave cometer un error de tipo I.

La imposición de la condición (1.7) no es suficiente para encontrar un test adecuado, dado que el contraste $\phi(x) = 0, \forall x \in \mathcal{X}$, es de nivel α para cualquier valor de esta constante, sin

⁴En el primer caso se hablará de un **falso positivo** y en el segundo de un **falso negativo**

embargo, esta elección está lejos de ser idónea⁵. Lo anterior lleva a tener que controlar el error de tipo II de alguna manera. Para esto se define la función de potencia, π , como la restricción de $\alpha(\theta)$ a Θ_1 y se interpreta como la probabilidad de rechazar correctamente la hipótesis nula. Es decir,

$$\begin{aligned} \pi : \Theta_1 &\rightarrow [0, 1] \\ \theta &\rightarrow \pi(\theta) = \alpha(\theta). \end{aligned} \quad (1.8)$$

Es evidentemente deseable que un test tenga una potencia alta. Lo anterior justifica la siguiente definición: un test ϕ^* , de nivel α , se dice uniformemente más potente (UMP) si para todo test ϕ , de nivel α , se tiene que

$$\pi^*(\theta) \geq \pi(\theta), \quad \forall \theta \in \Theta_1, \quad (1.9)$$

donde π^* y π son las funciones de potencia de ϕ^* y ϕ respectivamente. No obstante, no siempre existen test con la propiedad anterior, lo que lleva a tener que imponer otras propiedades de interés como insesgamiento, invarianza, etc. Para más detalles ver [67] y [44].

Antes de concluir esta sección, se presenta un resultado de existencia y unicidad de test de hipótesis UMP aleatorizados, en el caso $\Theta_0 = \{\theta_0\}$ y $\Theta_1 = \{\theta_1\}$. En esta situación, las hipótesis H_0 y H_1 se suelen denominar **simples**. La referencia usada es [70].

Teorema 1.1 (Lema de Neyman-Pearson). *Sean $\Theta_0 = \{\theta_0\}$ y $\Theta_1 = \{\theta_1\}$, y $p_i = \frac{dP_{\theta_i}}{d\nu}$ para $i = 0, 1$, las derivadas de Radon-Nikodym con respecto a una medida σ -finita ν sobre $(\mathcal{X}, \mathcal{A})$. Entonces:*

- (i) (Existencia de un test UMP aleatorizado) *Para todo $\alpha \in [0, 1]$, existe un test UMP de nivel α , dado por*

$$\phi^*(X) = \begin{cases} 1 & \text{si } p_1(X) > cp_0(X), \\ \gamma & \text{si } p_1(X) = cp_0(X), \\ 0 & \text{si } p_1(X) < cp_0(X), \end{cases} \quad (1.10)$$

donde $\gamma \in [0, 1]$ y $c \geq 0$ son constantes elegidas de modo que $E_{\theta_0}(\phi^*) = \alpha$ ($c = \infty$ es permitido).

- (ii) (Unicidad) *Si ϕ^{**} es un test UMP de nivel α , entonces*

$$\phi^{**}(X) = \begin{cases} 1 & \text{si } p_1(X) > cp_0(X), \\ 0 & \text{si } p_1(X) < cp_0(X), \end{cases} \quad \nu - c.s. \quad (1.11)$$

El lector interesado en la demostración del teorema anterior, puede consultar las pags. 394-395 de [70].

El teorema 1.1 es la base de la teoría de los test de hipótesis clásicos, también llamados frecuentistas, y posee varias extensiones, siendo las más conocidas las de Karlin-Rubin y el

⁵Esto pues, independiente de x , nunca se rechaza H_0 .

test de razón de verosimilitud. La extensión de Karlin-Rubin no siempre es aplicable, ya que requiere una hipótesis conocida como **propiedad de razón de verosimilitud monótona**, sin embargo, el test resultante es UMP. El test de razón de verosimilitud siempre es aplicable, no obstante, no siempre entrega test de hipótesis UMP. Para detalles ver cap. 8 de [16]. De lo anterior se destaca que la potencia aparece como un criterio de optimalidad frecuentista a la hora de construir test.

Una forma alternativa de rechazar o no un test de hipótesis, que no cae directamente dentro de la teoría de Neyman-Pearson y que será muy utilizado al momento de explicitar metodologías asociadas a los THM, es el enfoque basado en p -valores, que se define a continuación: considere una familia de test no aleatorizados $\mathcal{T} = \{\phi_\alpha : \alpha \in [0, 1]\}$ tal que ϕ_α es un test de nivel α . Sea $R(\alpha) := \phi_\alpha^{-1}(\{1\}) \subseteq \mathcal{X}$ la región de rechazo de $\phi_\alpha \in \mathcal{T}$ y suponga que $\{R(\alpha) : \alpha \in [0, 1]\}$ es una familia creciente de subconjuntos de \mathcal{X} tal que $R(0) = \emptyset$ y $R(1) = \mathcal{X}$. Para cada $x \in \mathcal{X}$ se define

$$\hat{p}(x) := \inf\{\alpha \in [0, 1] : x \in R(\alpha)\}. \quad (1.12)$$

El p -valor asociado a la familia de tests \mathcal{T} se define como la v.a. $\hat{p} = \hat{p}(X)$. Así, la regla de decisión para un test de nivel α , en base al p -valor, consiste en

$$\phi(x) = \begin{cases} 1 & \text{si } \hat{p} \leq \alpha, \\ 0 & \text{en otro caso.} \end{cases} \quad (1.13)$$

Para más referencias ver pags. 63-65 de [44].

En la sección siguiente se exponen brevemente los fundamentos de otra importante corriente estadística, en la cual cambia la interpretación estocástica que se le da al espacio de parámetros. La referencia usada es [30].

1.4 El paradigma bayesiano

El paradigma bayesiano se diferencia del enfoque frecuentista, visto en la sección anterior, en que el parámetro θ del modelo estadístico se visualiza como si fuera un elemento aleatorio θ a valores en un espacio medible (Θ, \mathcal{T}) , con \mathcal{T} una σ -álgebra de conjuntos sobre Θ , y definido sobre algún espacio de probabilidad abstracto. Lo anterior se consigue definiendo una medida de probabilidad \mathbf{Q} sobre $(\mathcal{X} \times \Theta, \mathcal{A} \otimes \mathcal{T})$ que se pueda descomponer como

$$\mathbf{Q}(dx, d\theta) = K(dx, \theta)\Pi(d\theta), \quad (1.14)$$

donde Π es una medida sobre (Θ, \mathcal{T}) denominada **distribución a priori** y K es un kernel probabilístico, esto es, una función $K : \mathcal{A} \times \Theta \rightarrow [0, 1]$ tal que $\forall \theta \in \Theta, K(\cdot, \theta) \in \mathcal{P}(\mathcal{A})$ y $\forall A \in \mathcal{A}, K(A, \cdot)$ es una función $\mathcal{T}, \mathcal{B}([0, 1])$ medible, donde $\mathcal{B}([0, 1])$ son los borelianos de $[0, 1]$. La medida a priori Π , o probabilidad a priori si integra 1, se interpreta como un conocimiento previo sobre los parámetros del modelo y el kernel K , como (una versión de)

la ley condicional de X dado θ , donde se hace la suposición que X observado tiene esta ley, es decir, se identifica $K(\cdot, \theta)$ con $P_\theta(\cdot)$. Esta manera de ver el espacio de parámetros provoca profundas diferencias entre los enfoques frecuentista y bayesiano, y muchas controversias. Es interesante en este punto destacar que en el paradigma bayesiano el parámetro es fijo, sin embargo, se modela la incerteza acerca de su verdadero valor a través de una medida de probabilidad. El lector interesado en una defensa de la perspectiva bayesiana y de sus fundamentos, desde la teoría de decisión, puede consultar [11] y [12].

Sea ν una medida σ -finita sobre $(\mathcal{X}, \mathcal{A})$ tal que $\forall \theta \in \Theta$, $P_\theta \ll \nu$ y sea $p(\cdot|\theta)$ la derivada de Radon-Nikodym correspondiente, la cual es la densidad condicional con respecto a ν de X dado θ . Sea P la distribución marginal de X . Se puede probar usando el teorema de Tonelli (ver cap. 1 de [67]) que $P \ll \nu$, con derivada de Radon-Nikodym dada por

$$p(x) = \int_{\Theta} p(x|\theta)\Pi(d\theta). \quad (1.15)$$

El objeto fundamental en que se sustentan todas las inferencias bayesianas es la **distribución a posteriori**, $\Pi(\cdot|x)$, que corresponde a la distribución condicional de θ dado X . Utilizando una versión general del teorema de Bayes presentada en [67] se prueba que la distribución a posteriori es absolutamente continua con respecto a la distribución a priori, con derivada de Radon-Nikodym dada por

$$\tilde{\pi}(\theta|x) = \frac{p(x|\theta)}{\int_{\Theta} p(x|\theta)\Pi(d\theta)}. \quad (1.16)$$

En el caso que $\Pi \ll \rho$ donde ρ es una medida σ -finita sobre (Θ, \mathcal{T}) y π la densidad correspondiente, se obtiene la siguiente fórmula para la **densidad a posteriori**

$$\pi(\theta|x) = \pi(\theta) \frac{p(x|\theta)}{p(x)}, \quad (1.17)$$

donde $p(x) = \int_{\Theta} \pi(\theta)p(x|\theta)\rho(d\theta)$. Típicamente ρ es la medida de Lebesgue, de manera que para mantener la notación usual del cálculo se escribe $p(x) = \int_{\Theta} \pi(\theta)p(x|\theta)d\theta$. Se concluye así el marco estructural matemático de la inferencia bayesiana considerada en este trabajo. En lo que sigue se explica de manera general, en qué consiste la teoría de test de hipótesis bayesiana.

Como dice Silvey en [71], la teoría de Neymann-Pearson no trata precisamente con las hipótesis, es decir, con los conjuntos de parámetros Θ_0 y Θ_1 , sino que busca una solución al problema de decisión asociado, intentando satisfacer ciertos criterios de manera que el procedimiento resultante tenga alguna propiedad de optimalidad. En cambio el enfoque bayesiano de test de hipótesis si admite una inferencia directamente sobre el espacio de parámetros, gracias a la existencia de la distribución a posteriori.

Desde el punto de vista bayesiano, lo único importante a la hora de contrastar hipótesis es calcular la probabilidad a posteriori de los conjuntos Θ_0 y Θ_1 respectivamente, y comparar

cuál de estas probabilidades es más grande. Lo anterior tiene una justificación desde el punto de vista de la teoría de decisión, ya que en cierto sentido, un test ϕ es un procedimiento que intenta estimar la función $\mathbb{1}_{\Theta_0}(\theta)$. Así, con la incorporación de una adecuada función de pérdida $L(\theta, \phi)$, es posible derivar un estimador de Bayes para el problema en estudio. La función de pérdida propuesta por Neyman-Pearson es

$$L(\theta, \phi) = \begin{cases} 1 & \text{si } \phi \neq \mathbb{1}_{\Theta_0}(\theta), \\ 0 & \text{en otro caso.} \end{cases} \quad (1.18)$$

Así, es posible probar que el estimador de Bayes ϕ^π asociado al problema anterior es

$$\phi^\pi(x) = \begin{cases} 1 & \text{si } \Pi(\Theta_0|x) > \Pi(\Theta_1|x), \\ 0 & \text{en otro caso,} \end{cases} \quad (1.19)$$

y el problema de test de hipótesis está completamente determinado desde la perspectiva bayesiana. Para más detalles ver cap. 5 de [52].

Según Berger (ver [11]), aunque las probabilidades a posteriori de Θ_0 y Θ_1 son las medidas principales para los test de hipótesis bayesianos, también es de interés el concepto de factor bayesiano:

Definición 1.1 (Factor bayesiano). Se define el **factor bayesiano** como

$$B = \frac{\Pi(\Theta_0|x)\Pi(\Theta_1)}{\Pi(\Theta_1|x)\Pi(\Theta_0)}. \quad (1.20)$$

El concepto anterior puede interpretarse algunas veces como la razón de H_0 a H_1 dado los datos. Al respecto, hay varias reglas para aplicar este concepto, quizás la más conocida sea la de Jeffreys (ver [40]), quien estableció la siguiente convención:

1. Si $\log_{10}(B)$ varía entre 0 y 0,5 entonces la evidencia en contra de H_0 es decisiva.
2. Si $\log_{10}(B)$ varía entre 0,5 y 1 entonces la evidencia en contra de H_0 es fuerte.
3. Si $\log_{10}(B)$ varía entre 1 y 2 entonces la evidencia en contra de H_0 es substancial.
4. Si $\log_{10}(B)$ es mayor que 2 entonces la evidencia en contra de H_0 es pobre.

Como ya se explicitó, hay muchas otras convenciones, pero para efectos de este trabajo bastará con la anterior.

Con esto se da por terminado el marco estructural asociado a la teoría de los test de hipótesis, en su enfoque frecuentista y bayesiano, que servirá de base para explicar la teoría de pruebas de múltiples hipótesis estadísticas, el tema central de esta memoria. No obstante, es importante destacar que hay varios tópicos que, por motivos de simplificación de la teoría

y espacio, no se presentaron en las secciones anteriores y que son de relativa importancia para esta memoria. Un ejemplo de esto, es el tema de la inferencia bayesiana no paramétrica, que actualmente está siendo foco de mucha investigación y que tiene algunas aplicaciones a los test de hipótesis múltiples. Al respecto, el lector interesado puede consultar la excelente monografía de Ghosh y Ramamoorthi (ver [36]).

1.5 Test de hipótesis múltiples

De manera informal se puede afirmar que los procedimientos de THM intentan controlar el número de resultados incorrectamente significativos mientras se testean simultáneamente un gran número de hipótesis estadísticas. Para realizar lo anterior es fundamental definir lo que se entiende por una adecuada **tasa de error**, que es básicamente una manera de conceptualizar los errores que se cometen al decidir en paralelo un gran número de contrastes de hipótesis. El problema anterior tuvo sus orígenes a principios del siglo XX y fue bien estudiado en el contexto de pocas hipótesis (del orden de 10 hipótesis). De estos estudios surgió la FWER (familywise error rate), que corresponde a la probabilidad que haya al menos una hipótesis erróneamente rechazada, como la tasa de error a controlar, no obstante, cambios tecnológicos actuales han provocado la necesidad de analizar un número muy elevado de hipótesis al mismo tiempo (del orden de miles o incluso millones de hipótesis), lo que ha obligado a definir otras tasas de error, esto pues, en dichos casos la FWER no tiene un buen comportamiento, como puede verse en el siguiente ejemplo.

Ejemplo 1.4. Se quieren contrastar m hipótesis estadísticas. Se definen las v.a's V , que representa el número de falsos positivos (esto es, el número de hipótesis incorrectamente rechazadas) cometidos al resolver los m test, y V_i , que toma el valor 1 si la hipótesis i es incorrectamente rechazada y 0 en otro caso. Formalmente

$$V = \sum_{i=1}^m V_i \quad \text{y} \quad V_i = \mathbf{1}\{\theta \in \Theta_0^i\} \mathbf{1}\{\phi_i = 1\}, \quad (1.21)$$

donde $\Theta_0^i = \{\theta_0\}$ es el conjunto de parámetros bajo la i -ésima hipótesis nula y ϕ_i es un test de la i -ésima hipótesis, $i = 1, \dots, m$. Suponiendo que $(V_i)_{i=1}^m$ son independientes y que el nivel de significación de cada test individual ϕ_i es $\alpha \in [0, 1]$, es decir, $P_{\theta_0}(V_i = 1) = \alpha$, entonces

$$\begin{aligned} P_{\theta_0}(V \geq 1) &= 1 - P_{\theta_0}(V = 0) \\ &= 1 - P_{\theta_0}(V_1 = 0, V_2 = 0, \dots, V_m = 0) \\ &= 1 - \prod_{i=1}^m P_{\theta_0}(V_i = 0) \\ &= 1 - (P_{\theta_0}(V_1 = 0))^m \\ &= 1 - (1 - P_{\theta_0}(V_1 = 1))^m \\ &= 1 - (1 - \alpha)^m. \end{aligned}$$

Para entender las repercusiones del resultado anterior, se procede a graficar la probabilidad de que se cometa al menos un falso positivo en función del número de hipótesis individuales a contrastar.

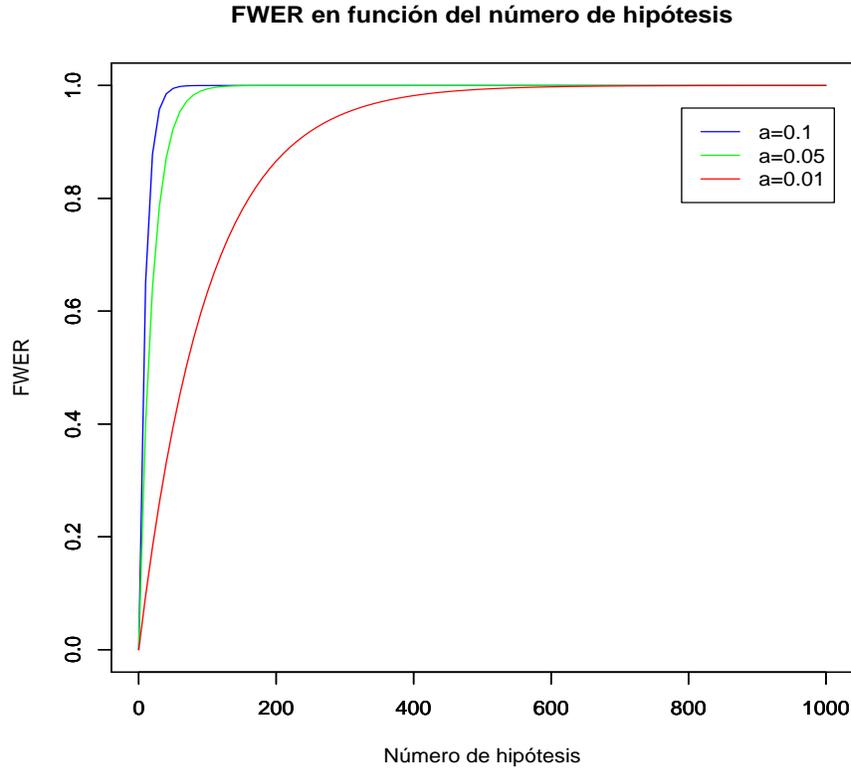


Figura 1.1: $P_{\theta_0}(V \geq 1)$ en función de m .

De la figura 1.1 se observa que para alrededor de 400 hipótesis, la probabilidad de cometer al menos un error de tipo I es prácticamente 1, lo que demuestra que en el contexto de muchas hipótesis⁶ controlar la FWER no es conveniente bajo el supuesto de independencia entre hipótesis. El fenómeno anterior se conoce con el nombre de **problema de la multiplicidad**.

||

El problema de los test de hipótesis múltiples surge cuando es necesario considerar de manera unificada varios problemas de contraste de hipótesis, como los descrito en las secciones anteriores. Los ingredientes de un problema de este tipo, en el caso que se tienen m hipótesis, son familias paramétricas $\mathcal{P}_i = \{P_{\theta^i} | \theta^i \in \Theta^i\}$, elementos aleatorios X^i a valores en $(\mathcal{X}^i, \mathcal{A}^i)$, hipótesis nulas $H_0^i : \theta^i \in \Theta_0^i$, hipótesis alternativas $H_1^i : \theta^i \in \Theta_1^i$ con $\Theta_0^i \cup \Theta_1^i = \Theta^i$ y distribuciones a priori Π^i en el caso bayesiano, para $i = 1, \dots, m$. En cada caso se dispone además de un test $\phi^i : \mathcal{X}^i \rightarrow \{0, 1\}$ a través del cual se decide rechazar o no cada hipótesis nula H_0^i . En la tabla 1.1 se definen algunos ingredientes adicionales importantes, relacionados

⁶Como sucede con los microarrays por ejemplo.

con los test de hipótesis múltiples, que serán utilizados frecuentemente en los capítulos 2 y 3 de este trabajo. La notación usada en esta tabla corresponde a la empleada en [4].

Tabla 1.1: Marco conceptual de los THM.

Hipótesis	Decisión		
	Aceptadas	Rechazadas	Total
Verdadadera	U	V	m_0
Falsa	T	S	$m - m_0$
Total	Q	R	m

Las definiciones de la tabla 1.1 se pueden formalizar como sigue: sean $I = \{1, \dots, m\}$ e $I(\theta) = \{i \in I : \theta_i \in \Theta_0^i\}$. Entonces

$$V = \sum_{i \in I(\theta)} \mathbb{1}\{\phi_i = 1\}; \quad U = \sum_{i \in I(\theta)} \mathbb{1}\{\phi_i = 0\}; \quad S = \sum_{i \in I \setminus I(\theta)} \mathbb{1}\{\phi_i = 1\},$$

$$T = \sum_{i \in I \setminus I(\theta)} \mathbb{1}\{\phi_i = 0\}; \quad R = \sum_{i \in I} \mathbb{1}\{\phi_i = 1\}; \quad Q = \sum_{i \in I} \mathbb{1}\{\phi_i = 0\}.$$

No es claro qué significa un test múltiple desde el punto de vista matemático porque, de hecho, todavía es de interés testear cada una de las hipótesis H_0^i . Por lo mismo, suponiendo que se dispone de test ϕ^i que son buenos, incluso óptimos, ¿por qué razón habría que hacer algo distinto que aplicarlos secuencialmente o simultáneamente? La respuesta a esta interrogante está en la premisa básica de los THM que, según Miller (ver [48]), es que “*hay que dar protección adicional a la hipótesis nula, correspondiente a un grupo de parámetros*”. Además este insiste, y con mucha razón, en no descuidar los errores que se cometen cuando las hipótesis alternativas son ciertas.

En el siguiente ejemplo se explica en qué consiste la protección adicional de la hipótesis nula cuando se contrastan simultáneamente varias hipótesis estadísticas.

Ejemplo 1.5. Sean $X \sim \mathcal{N}(\mu, 1)$ e $Y \sim \mathcal{N}(\nu, 1)$ variables aleatorias independientes y las hipótesis nulas $H_0^1 : \mu = 0$, $H_0^2 : \nu = 0$ versus las alternativas $H_1^1 : \mu \neq 0$, $H_1^2 : \nu \neq 0$. Dado un $\alpha \in [0, 1]$, para decidir H_0^1 se usa el test $|X| > z_{\alpha/2}$ y para H_0^2 , $|Y| > z_{\alpha/2}$, con $z_{\alpha/2} = \Phi^{-1}(1 - \alpha/2)$ y Φ la función de distribución de la $\mathcal{N}(0, 1)$. Se procede a agrupar las hipótesis nulas individuales en una sola hipótesis nula $H_0^{12} : \theta = 0$ contra la alternativa $H_1^{12} : \theta \neq 0$, donde $\theta = (\mu, \nu)$. Claramente se rechaza H_0^{12} cuando $|X| > z_{\alpha/2}$ o $|Y| > z_{\alpha/2}$, lo que equivale a rechazar H_0^{12} cuando $T = \max\{|X|, |Y|\} > z_{\alpha/2}$. Notemos ahora que bajo

la hipótesis nula H_0^{12} se verifica que,

$$\begin{aligned}
 P_\theta(T > z_{\alpha/2}) &= 1 - P_\theta(|X| \leq z_{\alpha/2}, |Y| \leq z_{\alpha/2}) \\
 &= 1 - P_\theta(|X| \leq z_{\alpha/2})P_\theta(|Y| \leq z_{\alpha/2}) \\
 &= 1 - (1 - \alpha)^2 \\
 &= 2\alpha - \alpha^2 > \alpha.
 \end{aligned}$$

Se concluye entonces que H_0^{12} no tiene el mismo nivel de protección que H_0^1 o H_0^2 , puesto que, la probabilidad de rechazar H_0^{12} es mayor. \parallel

La manera estándar de enfrentar el problema de la multiplicidad, así como la protección adicional de la hipótesis nula, es definiendo una adecuada tasa de error. Al respecto hay muchas posibles definiciones; de hecho, uno de los problemas aún abiertos en esta teoría es cómo definir una tasa que sea aplicable para cualquier estructura de dependencia entre las hipótesis a analizar. Sin embargo, una vez escogida una tasa de error, el camino a seguir es muy parecido al de los test de hipótesis no múltiples, esto es, se busca un procedimiento estadístico o, como se definirá en la sección siguiente, una regla de decisión múltiple que haga que la tasa de error escogida sea menor o igual que un valor determinado $\alpha \in [0, 1]$. Se aprecia así el por qué las tasas de error son la generalización de la probabilidad de error de tipo I, vista en las secciones anteriores.

En la sección siguiente se muestra que una familia de hipótesis estadísticas siempre pueden agruparse de manera que todas las hipótesis nulas se expresen en una única hipótesis nula, con respecto a un parámetro global.

1.6 Hipótesis simultáneas

Considere el marco estructural planteado en la sección anterior para los test individuales. Lo que sigue es ver que se pueden extender dichos conceptos al contexto de los test múltiples. En relación a lo anterior, es posible definir una nueva familia de medidas de probabilidad $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ definidas sobre un espacio medible $(\mathcal{X}, \mathcal{A})$, un elemento aleatorio X a valores en dicho espacio y un conjunto Θ_0 , para los cuales a su vez existen funciones medibles⁷ epiyectivas $\varphi_i : \Theta \rightarrow \Theta^i$ tales que $\varphi_i(\theta) = \theta^i$ y funciones medibles epiyectivas $\psi_i : \mathcal{X} \rightarrow \mathcal{X}^i$ tales que $\psi_i(X) = X^i$. El conjunto Θ_0 asociado a la hipótesis nula simultánea se define como $\Theta_0 = \bigcap_{i=1}^m \varphi_i^{-1}(\Theta_0^i)$, de manera que

$$\theta \in \Theta_0 \Leftrightarrow \theta^i = \varphi_i(\theta) \in \Theta_0^i, \quad i = 1, \dots, m. \quad (1.22)$$

Por otra parte, X es un observable sobre $(\mathcal{X}, \mathcal{A})$, con \mathcal{A} una σ -álgebra sobre \mathcal{X} , tal que si su ley es P_θ , entonces la ley de $X^i = \psi_i(X)$ es $P_{\varphi_i(\theta)}$ donde $\theta^i = \varphi_i(\theta)$. En otras palabras

$$P_\theta \circ \psi_i^{-1} = P_{\varphi_i(\theta)}, \quad \forall i = 1, \dots, m, \quad \forall \theta \in \Theta. \quad (1.23)$$

⁷Esta propiedad de medibilidad sólo es necesaria en un contexto bayesiano.

El conjunto Θ_0 se escribe como la intersección de m regiones $\varphi_i^{-1}(\Theta_0^i)$ de Θ , cada una de las cuales está asociada a las hipótesis nulas originales H_0^i . Es importante notar que θ puede tener más, o incluso menos, de m componentes puesto que en una formulación conjunta hay que definir la ley conjunta de (X^1, X^2, \dots, X^m) , que puede requerir más, o tal vez menos, de m parámetros. Esta situación se discute en el siguiente ejemplo.

Ejemplo 1.6. Se disponen de hipótesis $H_0^1 : \mu_1 = 0$ vs $H_1^1 : \mu_1 \neq 0$ y $H_0^2 : \mu_2 = 0$ vs $H_1^2 : \mu_2 \neq 0$ en el contexto de modelos gaussianos $\mathcal{N}(\mu_1, \sigma_1^2)$ y $\mathcal{N}(\mu_2, \sigma_2^2)$ respectivamente, con σ_1 y σ_2 desconocidos. Se poseen además observaciones X^1 y X^2 de dichos modelos probabilísticos. En este entorno de trabajo $\theta^1 = (\mu_1, \sigma_1)$ y $\theta^2 = (\mu_2, \sigma_2)$, y se puede construir un modelo que albergue a las hipótesis nulas H_0^1 y H_0^2 del siguiente modo: Se define el espacio $\mathcal{X} = \mathbb{R}^2$, el observable $X = (X_1, X_2)$ y la familia $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ de medidas de probabilidad sobre $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2))$, donde $\mathcal{B}(\mathbb{R}^2)$ es la σ -álgebra de los borelianos sobre \mathbb{R}^2 , como

$$P_\theta := \mathcal{N}(\mu_1, \sigma_1^2) \otimes \mathcal{N}(\mu_2, \sigma_2^2) \quad (1.24)$$

donde \otimes indica la medida producto, en este caso, de medidas gaussianas sobre \mathbb{R} . Esto último traduce el hecho que las componentes X^1, X^2 del vector X son independientes. En tal caso, claramente se define el parámetro $\theta = (\mu_1, \sigma_1, \mu_2, \sigma_2)$ y el espacio de parámetros $\Theta = \Theta^1 \times \Theta^2 = \mathbb{R} \times \mathbb{R}_+ \times \mathbb{R} \times \mathbb{R}_+$. Así, las funciones $\varphi_1, \varphi_2, \psi_1, \psi_2$ son simplemente las proyecciones y todo es bastante sencillo. Sin embargo, puede ocurrir que el modelo conjunto P_θ definido anteriormente no sea adecuado, porque por ejemplo X_1 y X_2 no son independientes. En este caso, si aún fuera razonable creer que la ley conjunta es gaussiana, se podría considerar como modelo probabilista

$$P_\theta := \mathcal{N} \left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{pmatrix} \right) \quad (1.25)$$

donde $\sigma_{12} = Cov(X_1, X_2) = \sigma_{21}$. Así, el parámetro $\theta = (\mu_1, \sigma_1, \mu_2, \sigma_2, \sigma_{12})$, es decir, ahora el modelo tiene 5 parámetros reales, y el espacio de parámetros $\Theta = (\mathbb{R} \times \mathbb{R}_+)^2 \times \mathbb{R}$, agregando la restricción que $\sigma_1^2 \sigma_2^2 > \sigma_{12}^2$ para que la matriz de varianza-covarianza de X sea definida positiva. Las funciones ψ_i, φ_i siguen siendo las proyecciones y se aprecia ahora que el modelo tiene más parámetros que los de los modelos originales. No obstante, hay muchas maneras alternativas de construir medidas de probabilidad sobre $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2))$ de manera que las marginales sean $\mathcal{N}(\mu_1, \sigma_1^2)$ y $\mathcal{N}(\mu_2, \sigma_2^2)$. En tal caso, P_θ podría ya no ser una ley gaussiana y quizás habría que incorporar muchos parámetros más para describir la ley de X , siendo necesarios incluso una infinidad, lo que obligaría a entrar en el dominio de los modelos no paramétricos. ||

El ejemplo anterior muestra que puede ser muy difícil especificar el modelo probabilista adecuado para el observable X , a partir de los elementos aleatorios X^1 y X^2 relacionados con las hipótesis nulas H_0^1 y H_0^2 respectivamente.

En lo que sigue, se supone que es posible especificar lo que se denomina el modelo conjunto, que se designa simplemente como $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$, y el observable X a valores

en el espacio medible $(\mathcal{X}, \mathcal{A})$. Se considera además que se dispone de conjuntos $\Theta_0^i \subseteq \Theta$, $i = 1, \dots, m$, cada uno de los cuales define la hipótesis H_0^i . En tal caso, la hipótesis nula conjunta o simultánea se especifica como

$$H_0 : \theta \in \Theta_0 := \bigcap_{i=1}^m \Theta_0^i \neq \emptyset, \quad (1.26)$$

quedando la hipótesis alternativa especificada por Θ_1 , el complemento de Θ_0 en Θ . Es muy importante destacar la importancia que $\Theta_0 \neq \emptyset$, para que el test múltiple asociado tenga sentido.

Dado un test ϕ , para las hipótesis H_0 vs H_1 especificadas más arriba, y $\alpha \in [0, 1]$, se puede en principio proteger la hipótesis nula, siguiendo la teoría de Neyman-Pearson, imponiendo que el test tenga nivel α , es decir,

$$\sup_{\theta \in \Theta_0} \alpha(\theta) \leq \alpha \quad (1.27)$$

quedando resuelto el problema de las hipótesis múltiples, al menos en lo que respecta a controlar el error de tipo I, sin embargo, la respuesta que se obtiene en caso de rechazar la hipótesis nula H_0 , no es satisfactoria, esto pues, es indispensable señalar cuáles son las hipótesis nulas originales H_0^i , $i = 1, \dots, m$, que son rechazadas. Esta es la razón por la cual la teoría de Neyman-Pearson no es adecuada para analizar el problema de los test de hipótesis múltiples.

Para abordar este problema se deben abandonar los test clásicos del tipo $\phi : \mathcal{X} \rightarrow \{0, 1\}$ y pasar a reglas de decisión del tipo $\phi : \mathcal{X} \rightarrow \{0, 1\}^m$ donde

$$\phi(x) = (\phi_1(x), \dots, \phi_i(x), \dots, \phi_m(x)) \quad (1.28)$$

y $\forall i = 1, \dots, m$, ϕ_i es un test de hipótesis que, por simplicidad, se considera no aleatorizado (ver (1.4)). Otra manera de enfrentar el problema de los test de hipótesis múltiples, bastante usada en la práctica, es definir una regla de decisión del tipo $\phi : \mathcal{X} \rightarrow 2^{\{1, \dots, m\}}$ donde

$$\phi(x) \subseteq \{1, \dots, m\} \quad (1.29)$$

y representa el conjunto de índices cuyas hipótesis nulas son rechazadas, es decir, si la aplicación $\phi(x) = \{2, 7, m - 1\}$, entonces las hipótesis nulas rechazadas son H_0^2 , H_0^7 y H_0^{m-1} .

En la sección siguiente se expone el desarrollo histórico de los test de hipótesis múltiples, basada en una revisión bibliográfica realizada durante el desarrollo de esta memoria.

1.7 Desarrollo histórico y revisión bibliográfica

La teoría de los test de hipótesis múltiples tardó en surgir. Al respecto, se fueron generando una gran cantidad de resultados provenientes de los análisis de regresión y varianza, que

permitieron que se formaran las bases de esta teoría. En este sentido, el período entre 1902 y 1946 podría considerarse como un inicio rudimentario del área donde connotados investigadores, como Pearson (padre e hijo), Student y Fisher, hicieron sus primeros aportes. Para detalles y una lista exhaustiva de las publicaciones de esta época, ver [48] y sus referencias. Sin perjuicio de lo anterior, es entre 1947 y 1955 cuando se sientan de manera explícita las bases de esta teoría, gracias al aporte de tres investigadores: D.R. Duncan, H. Scheffé y J.W. Tukey. Es muy importante destacar que fueron ellos quienes identificaron el problema de multiplicidad como un problema en sí mismo, más que una dificultad derivada de un problema de regresión múltiple. Posterior a esta época, se fueron generando una gran cantidad de publicaciones en el área, sin embargo, faltaba que alguien condensara todo ese material y lo presentara de manera clara y accesible en un libro. Esto no ocurrió hasta 1966, cuando R. G. Miller publicó la primera edición de su libro: “*Simultaneous Statistical Inference*”, en la que hace una revisión bibliográfica de las publicaciones más connotadas hasta 1966. En una segunda edición de este, la revisión se extiende hasta 1976. Es muy importante destacar la importancia del texto de Miller, debido a que expone el problema de los test de hipótesis múltiples a un público cuyo interés son las aplicaciones y que no tenía un acercamiento a dicho material, por encontrarse en revistas muy especializadas o de difícil acceso. En este punto hay que prevenir al lector interesado en estos temas, que la exposición de Miller se parece mucho a la de un libro de recetas de cocina, lo cual puede ser muy útil para quien quiera aplicar estas metodologías, pero no muy adecuada para quien se interese en los fundamentos matemáticos de los THM. Entrando un poco más en los detalles técnicos, durante esta época la tasa natural de error a controlar era la “*familywise error rate*”, abreviada FWER, cuya definición corresponde a

$$FWER = P_{\theta}(V \geq 1), \quad (1.30)$$

donde V es el número de falsos positivos cometidos en la aplicación de los m test de hipótesis (ver tabla 1.1). Un punto interesante de esta tasa de error es que si se piensa un problema de test de hipótesis múltiple desde el punto de vista de la teoría de Neyman-Pearson, esta es la generalización natural de la probabilidad de error de tipo I. El control más común utilizado sobre esta tasa de error es el procedimiento de Bonferroni, el cual consiste en testear cada hipótesis individual al nivel $\frac{\alpha}{m}$, para así lograr un control de nivel α para la FWER (para más detalles ver [48]). No obstante, cuando el número total de hipótesis a contrastar, m , es muy grande, se obtiene un procedimiento que es demasiado conservativo como se vio en el ejemplo 1.4, es decir, se rechazan muy pocas hipótesis nulas. Los investigadores de la época, conscientes de lo anterior, siguieron usando este criterio pues, en la práctica no se necesitaban testear tantas hipótesis al mismo tiempo. Al final de la década de los 80' e inicios de los 90' esta situación comenzó a cambiar, debido principalmente al estudio de los microarray, contexto en el cual se necesitan analizar miles e incluso millones de genes de manera simultánea. Lo anterior provocó la necesidad de extender la teoría existente, lo que sucedió de dos maneras: generalizando la FWER y sus formas de control, y definiendo nuevas tasas de error. En relación a lo primero, se han propuesto varias generalizaciones al procedimiento de Bonferroni. Por ejemplo, Simes en [72] propuso trabajar con la secuencia $\alpha_i = i \frac{\alpha}{m}$ de valores críticos para compararlos con los p -valores de cada test individual. En lo

que respecta a generalizar la FWER, el trabajo más influyente es sin duda el de Lehmann y Romano (ver [43]), quienes propusieron la k-FWER como la tasa de error natural a controlar y cuya definición corresponde a

$$\text{k-FWER} = P_{\theta}(V \geq k). \quad (1.31)$$

En relación a la definición de nuevas tasas de error se encuentra la “*per comparison error rate*”, abreviada PCER, y cuya definición es

$$PCER = E_{\theta} \left(\frac{V}{m} \right) = \int \left(\frac{V}{m} \right) dP_{\theta}. \quad (1.32)$$

Esta tasa se caracteriza porque la significancia de cada hipótesis individual es decidida sin considerar las del resto, debido a la linealidad del operador esperanza. Lo anterior hace que sea un concepto relativamente fácil de aplicar, no obstante, no toma en cuenta las posibles dependencias que puedan existir entre las distintas hipótesis a contrastar. Por lo mismo, permite un control muy liberal en el sentido que se rechazan muchas hipótesis, aumentando por ende la posibilidad que se cometan más errores de tipo I. Sin perjuicio de lo anterior, esta tasa es defendida por algunos autores como Saville (ver [66]), quien propone usarla debido a su simplicidad. Volviendo al contexto histórico, es importante destacar que las publicaciones a inicios de los 90' se caracterizan por una búsqueda de una nueva tasa de error que no fuera tan conservadora como la FWER, pero que tampoco fuera tan liberal como la PCER. Finalmente, una tasa de error distinta y mejor apareció, y fue propuesta en 1995 por Benjamini y Hochberg (ver [4]). Esta es la conocida “*false discovery rate*”, abreviada FDR, y cuya definición corresponde a

$$FDR = E_{\theta} \left(\frac{V}{R} \right) = \int \left(\frac{V}{R} \right) dP_{\theta}, \quad (1.33)$$

donde el cociente se considera cero cuando $R = 0$ (ver tabla 1.1). Esta tasa cambió el panorama de los test de hipótesis múltiples, ya que incentivó el desarrollo de numerosas nuevas investigaciones centradas tanto en la búsqueda de nuevas tasas de error derivadas de (1.33), así como de procedimientos alternativos, al propuesto por BH⁸, para controlar la FDR. Esta revolucionaria tasa se caracteriza, además, por tener muchas propiedades interesantes vistas tanto en [4] como en otros artículos de interés. Algunas de estas propiedades se comentan a continuación:

1. Si todas las hipótesis nulas son verdaderas, entonces controlar la FDR es equivalente a controlar la FWER.
2. $PCER \leq FDR \leq FWER$ (para detalles ver [32]).
3. Existen algoritmos para controlar la FDR bajo condiciones generales del comportamiento conjunto de los test estadísticos, esto es, cuando existen ciertas estructuras de dependencia entre los test estadísticos a contrastar.

⁸Benjamini y Hochberg

El punto 2 es muy interesante, pues significa que cualquier procedimiento que controle la FWER también controla la FDR, pero no necesariamente al revés. Otra característica muy importante de la FDR, es que existen procedimientos, como el de BH, que son más potentes que los procedimientos existentes para controlar la FWER, esto pues, se usan niveles de significación mayores para los test individuales, lo que repercute en una mayor probabilidad de error de tipo I, una menor probabilidad de error de tipo II y por tanto, en una mayor potencia para cada test individual. De lo anterior se desprende que una característica importante al momento de escoger test múltiples es que estos rechacen la mayor cantidad posible de hipótesis, garantizando el control de una adecuada tasa de error. En relación a esto último, hay dos tipos de controles que se pueden realizar sobre una determinada tasa de error. El primero se denomina control débil y consiste a realizar el control suponiendo que todas las hipótesis nulas son verdaderas. El segundo se denomina control fuerte y consiste en realizar el control para cualquier configuración de hipótesis nulas verdaderas.

Posterior al año 1995, las publicaciones más relevantes se centran alrededor de la FDR, de manera que en lo sigue la revisión bibliográfica se enfocará en este tema. Volviendo al procedimiento de BH para controlar la FDR, este se caracteriza por ser un procedimiento de tipo secuencial, también llamado *step-up*, donde los p -valores $\hat{p}_1, \dots, \hat{p}_m$ asociados a los test individuales, son secuencialmente comparados, comenzando con el más grande, con una serie de valores hasta encontrar el p -valor crítico, de manera que todas las hipótesis con p -valores menores a este valor crítico son rechazadas. En específico, se rechazan todas las hipótesis cuyos p -valores sean menores a $\hat{p}_{(\hat{k})}$ donde $\hat{k} = \max\{i : \hat{p}_{(i)} \leq \alpha_i\}$ con $\alpha_i = i \frac{\alpha}{m}$ y adicionalmente $\hat{p}_{(1)} \leq \dots \leq \hat{p}_{(m)}$ son los p -valores ordenados. Posteriormente, Benjamini y Yekutieli en [8], Sarkar en [58] y Storey et al. en [79], probaron que bajo el procedimiento de BH, se verifica que

$$FDR = m_0 \frac{\alpha}{m}, \quad (1.34)$$

(ver tabla 1.1) en el caso donde los test estadísticos individuales y los p -valores asociados son independientes, y $FDR < m_0 \frac{\alpha}{m}$, cuando los test estadísticos tienen una dependencia positiva, concepto que será explicado en el capítulo 2 de esta memoria. Lo anterior es muy relevante, ya que indica que el procedimiento de BH se vuelve conservativo a una tasa proporcional al número de hipótesis nulas verdaderas, esto es, m_0 . Por otro lado, hay otras formas de controlar la FDR, por ejemplo en [58] se sugiere un procedimiento denominado *step-down*, en el cual ninguna de las hipótesis asociadas con el p -valor $\hat{p}_{(\hat{l})}$ y mayores son rechazadas, donde $\hat{l} = \min\{i : \hat{p}_{(i)} > \alpha_i\}$ con $\alpha_i = i \frac{\alpha}{m}$. Con este procedimiento, aún en el caso independiente, se cumple que la FDR puede ser menor que $m_0 \frac{\alpha}{m}$.

En lo que sigue se considera el procedimiento *step-up* propuesto por BH en el caso independiente. Así, se tiene que $FDR = m_0 \frac{\alpha}{m}$ y si la proporción de hipótesis nulas verdaderas $d = \frac{m_0}{m}$ es conocida, entonces α puede ser manipulado de manera que la FDR satisfaga el control deseado. Por ejemplo, si $\alpha_i = i \frac{\alpha}{md}$ entonces se puede demostrar que $FDR = \alpha$. No obstante, en la práctica el valor de m_0 es desconocido, de manera que no es trivial lograr que la FDR tenga algún valor predeterminado. De lo anterior se deduce la importancia de

poder determinar de alguna forma el valor de \mathbf{d} para tener alguna noción de la bondad del procedimiento resultante.

A consecuencia de lo anterior, durante la última década se han introducido otros procedimientos secuenciales para controlar la FDR, en donde se propone un **estimador adaptativo**⁹ $\widehat{\mathbf{d}}$ para \mathbf{d} y los valores críticos son modificados a $\alpha_i = i\alpha/m\widehat{\mathbf{d}}$. De manera heurística parece razonable pensar que este procedimiento debería llevar a que la FDR tome el valor $\mathbf{d}\alpha E(\widehat{\mathbf{d}}^{-1})$, de manera que si $\widehat{\mathbf{d}}$ es un estimador eficiente (es decir, de varianza pequeña) de \mathbf{d} , entonces la FDR debería tomar un valor cercano a α . En este sentido, Storey en [76] sugirió usar $\widehat{\mathbf{d}} = \frac{m-R(\lambda)+1}{m(1-\lambda)}$ donde el término $R(\lambda) = \sum_{i=1}^m \mathbb{1}\{\hat{p}_i \leq \lambda\}$ es el número de p -valores más pequeños que λ , donde $0 < \lambda < 1$ es una constante por determinar. Se destaca que estimadores similares fueron sugeridos mucho antes por Schweder y Spjøtvoll en [68]. Así, suponiendo que λ ha sido determinado, entonces se escoge la siguiente secuencia de valores críticos:

$$\alpha_i = \min \left\{ \lambda, \frac{i\alpha(1-\lambda)}{m-R(\lambda)+1} \right\}. \quad (1.35)$$

Es muy importante señalar que en términos generales los procesos adaptativos permiten un control adecuado de la FDR, en el sentido que permiten mejorar significativamente las propiedades de potencia. Sin embargo, el rendimiento de los procedimientos resultantes dependen fuertemente de la elección de λ . Para más detalles ver [5], [13] y sus referencias. En cuanto a la elección de la constante λ ha habido varias propuestas. Por ejemplo, Storey en [77] sugirió varios procedimientos para determinar su valor, mientras que Benjamini et al. en [7] sugirieron elegir $\lambda = \frac{\alpha}{1+\alpha}$. En este último artículo se analizan además las propiedades de potencia del procedimiento resultante. El lector interesado en las propiedades teóricas de estos procedimientos puede consultar [62]. Por último, Gavrilov et al. en [31] sugieren procedimientos adaptativos *step-down* análogos.

En términos generales, los procedimientos estándar para controlar la FDR pueden ser descritos como procedimientos en los cuales la tasa de error es fija y donde las hipótesis individuales son testeadas a diferentes niveles de significancia para mantener constante la tasa de error global. Sin embargo, ha habido extensiones de esta idea, por ejemplo, Storey en [76] y [77] introduce un enfoque basado en una región de rechazo fija, donde $\alpha_i = \alpha$ para todo $i = 1, \dots, m$. Así, dada esta región de rechazo, la FDR puede ser estimada de los datos y entonces α es escogido de manera que la FDR tenga el nivel deseado.

Volviendo al tema de las generalizaciones basadas en la FDR, Storey en [76] propone la “*positive false discovery rate*”, abreviada *pFDR*, y cuya definición es

$$pFDR = E_{\theta} \left(\frac{V}{R} \mid R > 0 \right). \quad (1.36)$$

⁹Por definición un estimador adaptativo es un estimador eficiente para un parámetro de interés, para el cual el modelo estadístico asociado sólo es conocido parcialmente.

Según Storey, esta tasa de error es una versión más adecuada que la FDR. Más aún, mostró que si se supone un modelo mixto sobre las hipótesis, es decir, si se supone que las verdaderas hipótesis nulas se generan de manera aleatoria de acuerdo a una variable aleatoria de Bernoulli de parámetro \mathbf{d} , entonces se cumple que

$$pFDR(\alpha) = \frac{\mathbf{d}\alpha}{F(\alpha)}, \quad (1.37)$$

donde F es la función de distribución marginal de los p -valores. Es importante destacar que la expresión anterior no es cierta si los p -valores son dependientes, lo cual es una gran complicación para problemas provenientes de los microarrays, los cuales se caracterizan por presentar ciertas estructuras de dependencia. No obstante, debido a la simplicidad de su expresión, es ideal para un contexto de estimación. Así, una vez que una estimación \widehat{pFDR} de $pFDR$ ha sido obtenida, el procedimiento se reduce a rechazar todas las hipótesis cuyos p -valores que son menores o iguales a $\hat{\gamma}$ donde

$$\hat{\gamma} = \max\{\gamma : \widehat{pFDR}(\gamma) \leq \alpha\}. \quad (1.38)$$

Siguiendo con lo anterior, Storey en [76] mostró que el procedimiento de BH puede ser visto como un caso particular del procedimiento anterior, donde \mathbf{d} es estimado por la constante 1. Más aún, mostró que este procedimiento puede mejorar significativamente la potencia en determinados casos.

En relación a la $pFDR$, Storey en [77] mostró que esta tiene la siguiente interpretación bayesiana: “la probabilidad a posteriori que una hipótesis nula sea verdadera dado que esta ha sido rechazada”. Esta interpretación es muy interesante, ya que conecta al enfoque frecuentista y bayesiano, en el contexto de los test de hipótesis múltiples. Se destaca que han surgido varios procedimientos alternativos de estimación de (1.37), no obstante, dado que $\mathbf{d}\alpha \leq F(\alpha)$, existe la posibilidad de violar esta desigualdad si es que $F(\alpha)$ y \mathbf{d} se estiman de manera independiente. En un intento por solucionar lo anterior, Tang et al. (ver [81]) proponen un modelo bayesiano no paramétrico para estimar simultáneamente \mathbf{d} y $F(\alpha)$, en el contexto de un modelo mixto o jerárquico en el que naturalmente se restringen los estimadores a cumplir con la desigualdad deseada. Como resultado, se obtiene una estimación de la $pFDR$ que es bastante eficiente.

Como ya se ha mencionado anteriormente, la existencia de estructuras de dependencia entre los p -valores transforma al problema de los THM en uno muy complejo y además, de mucho interés práctico. En este sentido, ha habido mucha investigación, sin embargo, la mayoría de los procedimientos propuestos se centran en introducir algún tipo muy especial de dependencia entre los test estadísticos o los p -valores. Por ejemplo, Benjamini y Yekutieli en [8] proponen un procedimiento que controla la FDR bajo una dependencia que ellos llaman de “regresión positiva”, donde los valores críticos utilizados son

$$\alpha_i = \frac{i\alpha}{m \sum_{j=1}^i j^{-1}}. \quad (1.39)$$

Lamentablemente, este procedimiento resulta ser muy conservativo. Por otro lado, Sarkar en [62] propone un conjunto alternativo de valores críticos y también investiga el rendimiento del procedimiento resultante bajo ciertas estructuras especiales de dependencia. Por su lado, Finner y Roters en [28] y Efron en [20] sugieren modelar la transformación probit de los p -valores como una distribución conjunta normal, para así capturar la dependencia entre ellos. Un procedimiento similar fue sugerido por Roy y Ghosal en [53], quienes usaron una mezcla de densidades normales inclinadas para incorporar la dependencia entre p -valores, donde esta distribución mixta es estimada usando técnicas bayesianas no paramétricas. De lo anterior se observa que ha habido mucha investigación en torno a este problema, pero que sin embargo, aún no se obtiene una respuesta adecuada para una estructura de dependencia más general.

Las publicaciones en torno a nuevas tasas de error no se han detenido. Por ejemplo, Efron en [20] propone la denominada “*local FDR*” y abreviada l-FDR. Esta se diferencia de la FDR, en que no se trabaja directamente con la función de distribución marginal de los p -valores, $F(\alpha)$, sino más bien con la densidad marginal de estos. Otras generalizaciones de la FDR pueden ser encontradas en [60], [61], [63] y sus referencias. En relación a este último artículo, los autores definen la k-FDR, en un intento por homologar a lo acontecido con la k-FWER. El lector interesado en ver los últimos resultados respecto de esta tasa debe consultar Sarkar y Guo (ver [64]).

Respecto del enfoque bayesiano del problema de los test de hipótesis múltiples, algunas propuestas han sido discutidas en [17], [81], [35], [53] y [33]. En esta última publicación los autores proponen el uso de una secuencia de p -valores ponderados, que incorporan la información a priori acerca de las hipótesis individuales, para controlar la FDR. Por otro lado está la estimación empírica bayesiana de la FDR, la cual ha sido discutida en [23]. Es importante recalcar que una característica interesante del enfoque bayesiano, es que resulta fácil predecir algunas cantidades como, por ejemplo, la “*false discovery proportion*”, abreviada FDP y cuya definición es $FDP = \frac{V}{R}$. En relación de este último concepto, un trabajo notable es [34], en el que se propone un enfoque basado en procesos estocásticos para controlar la FDR. Más aún, se demuestra que

$$FDP(\alpha) = \frac{\sum_{i=1}^m I_i(\alpha)(1 - \mathfrak{H}_i)}{\sum_{i=1}^m I_i(\alpha) + \prod_{i=1}^m (1 - I_i(\alpha))}, \quad (1.40)$$

donde $I_i(\alpha) = \mathbf{1}\{\hat{p}_i < \alpha\}$ y \mathfrak{H}_i es el indicador de veracidad de la i -ésima hipótesis alternativa, tomando el valor 1 en caso de ser verdadera y 0 en caso contrario.

Suponiendo que $(\mathfrak{H}_i, I_i(\alpha))$ para $i = 1, \dots, m$, son v.a. intercambiables, Roy y Ghosal (ver [53]) probaron que $FDR(\alpha) = \mathbf{d}b(\alpha)P_\theta(R > 0)$, donde $b(\alpha)$ es el valor esperado de una función que depende de las funciones indicadoras $\mathfrak{H}_i, I_i(\alpha)$, lo cual implica que $pFDR(\alpha) = \mathbf{d}b(\alpha)$. Una expresión similar fue derivada en [8] y [59]. En particular, Sarkar en [59] mostró que la cantidad $\frac{b(\alpha)}{\alpha}$ es el valor esperado del estimador jackknife de $E[(1 + R)^{-1}]$.

Dentro de la gran cantidad de temas que rodean a los test de hipótesis múltiples, durante la última década han aparecido algunas importantes contribuciones dentro de lo que se puede llamar la teoría de decisión múltiple. Dentro de los trabajos a destacar en esta área está el de Finner y Strassburger (ver [29]) en donde exponen, a través del principio de la partición, una forma de contruir test más potentes. La mayoría de los trabajos en esta área se dirigen a determinar tests que tengan ciertas propiedades de optimalidad, que justifiquen la elección de realizada, utilizando para ello la teoría de la decisión estadística. En este camino se encuentran, por ejemplo, los trabajos de Storey (ver [78]) y Sun y Cai (ver [80]).

Para concluir este breve resumen acerca del desarrollo metodológico de los test múltiples, se quiere mencionar el artículo de Hall y Wang (ver [37]). El problema de la dependencia estadística no está del todo resuelto, ya que los procedimientos para enfrentar esta dificultad suelen ser muy conservativos, lo que de cierta manera va en contra del espíritu de los THM. Por lo mismo, muchos investigadores suelen suponer independencia de los test estadísticos, lo cual puede ser en algunos casos una imposición muy fuerte. Bueno, esto último no es del todo cierto como lo demuestran en su artículo Hall y Wang, ya que cuando el tamaño muestral es más grande que el logaritmo del número de test, la hipótesis de independencia no tiene consecuencias en los resultados inferenciales, en el contexto del test t, aún cuando no sea una suposición válida.

1.8 Aplicaciones

Los problemas de test de hipótesis múltiples aparecen en el contexto de análisis de datos en los cuales el interés es extraer las características significativas del fenómeno en estudio, de entre un gran conjunto de candidatos. Así, esta metodología es aplicable en cualquier circunstancia en que el enfoque deseado sea el anterior. Los ejemplos clásicos son:

- (1) Análisis de microarrays.
- (2) Detección de fuentes en astronomía.
- (3) Análisis de neuro-imágenes.

De estos, sin duda, el análisis de microarrays es el que más investigación y publicaciones ha generado. Aquí el interés es básicamente determinar qué genes están diferencialmente expresados, en el contexto de tumores y cánceres principalmente. Al respecto, algunas referencias importantes son de Wiel y van Wieringen (ver [82]) y Dudoit y van der Laan (ver [19]). En relación al ejemplo (2), el objetivo es decidir si las observaciones son consistentes con las predicciones del modelo teórico. El lector interesado en estos temas puede consultar a Miller et al. (ver [47]). El punto (3), en términos bastante simples, busca analizar qué zonas del cerebro se relacionan con determinados estímulos. Algunas referencias interesantes son Pantazis et al. (ver [50]) y Benjamini y Heller (ver [3]).

Test de Hipótesis Múltiples, enfoque Frecuentista

En este capítulo se enfatizan los detalles técnicos de las ideas más importantes de los test de hipótesis múltiples, desde una perspectiva frecuentista. Es importante enfatizar que en este capítulo, y el siguiente, se comentan y relacionan estas ideas, que son aportes originales de este trabajo. Se destaca también, que el criterio usado para su inclusión fue su importancia, medida a través de su impacto en MathSciNet, y la innovación en cuanto a ideas matemáticas.

2.1 FDR y el cambio de paradigma de los THM

Como ya se hizo referencia en el capítulo 1, la introducción de la FDR (tasa de falsos positivos) por parte de BH (ver [4]) cambió el paradigma de los test de hipótesis múltiples, en el sentido que provocó una avalancha de nuevas investigaciones y artículos orientados a definir tasas de error que fuesen apropiadas para los problemas de contrastes múltiples que estaban intentando abordar los científicos de la época, en especial, los genetistas.

En lo que sigue se supone el marco estructural discutido en la sección 1.6 del capítulo 1 y el marco conceptual dado en la tabla 1.1. Lo primero que se destaca de la información contenida en esta, es que R , U , V , T y S son variables aleatorias, no obstante, solamente R es observable. Adicionalmente, notar que si se testea cada hipótesis individual a nivel $\alpha \in [0, 1]$, entonces $R = R(\alpha)$ es variable aleatoria creciente en función de α , es decir, $P_\theta(R(\alpha) \geq R(\alpha')) = 1$ si $\alpha > \alpha'$.

Sea $FDP = \frac{V}{\max\{R, 1\}}$ la proporción de falsos positivos. Este concepto puede ser visto como una variable aleatoria que no es observable, debido a V , y es muy importante en la teoría de los test de hipótesis múltiples, ya que hay varias tasas de error basadas en ella. Como ya fue mencionado en la sección 1.7 del capítulo 1, BH definieron la tasa de falsos positivos, abreviada FDR, como

$$FDR = E_\theta(FDP) = \int FDP dP_\theta. \quad (2.1)$$

Algunas propiedades importantes de la FDR se resumen en la siguiente proposición:

Proposición 2.1. *Para la FDR se verifica lo siguiente:*

- (a) *Si todas las hipótesis nulas son verdaderas (i.e. $m_0 = m$), entonces $FDR = FWER$.*
- (b) *Cuando $m_0 < m$, entonces $FDR \leq FWER$.*

La demostraciones anteriores son sencillas de manera que se omiten. En todo caso, el lector interesado en verlas puede consultar la pag. 4 de [4]. En el contexto de la propiedad (a) se deduce que controlar la FDR equivale a controlar la FWER en un sentido denominado débil, esto es, cuando $m_0 = m$, y en relación a la propiedad (b), se concluye que cualquier procedimiento que controle la FWER también controlará la FDR. Además, Benjamini y Hochberg señalan, con mucha razón, que mientras más hipótesis nulas existan, eventualmente mayor es la potencia obtenida controlando la FDR en vez de la FWER.

El procedimiento de BH para controlar la FDR, es un método tipo Bonferroni que se describe a continuación: sean $\hat{p}_{(1)} \leq \dots \leq \hat{p}_{(m)}$ los p -valores ordenados asociados a las hipótesis nulas $H_0^{(i)}$. Entonces el test de hipótesis múltiple ϕ^{BH} definido por BH se caracteriza por rechazar las hipótesis nulas dadas por

$$\phi^{BH}(x) = \{\pi^*(1), \dots, \pi^*(\hat{k})\}, \quad (2.2)$$

donde $\hat{k} = \max\{i \in \{1, \dots, m\} : \hat{p}_{(i)} \leq \alpha_i\}$ con $\alpha_i = i \frac{\alpha}{m}$ y π^* es el inverso del operador que ordena los índices de los p -valores. Una manera alternativa de describir este test es a través de

$$\phi_i^{BH}(x) = \begin{cases} 1 & \text{si } (\pi^*)^{-1}(i) \leq \hat{k}, \\ 0 & \text{en otro caso,} \end{cases} \quad (2.3)$$

$\forall i = 1, \dots, m$, donde $(\pi^*)^{-1}$ es la función inversa de π^* y $\phi^{BH}(x) = (\phi_1^{BH}(x), \dots, \phi_m^{BH}(x))$. En ambos casos, $x \in \mathcal{X}$ con \mathcal{X} el espacio asociado a los datos muestrales, es decir, x es una realización del elemento aleatorio X . Es muy importante destacar en este punto que en la práctica, es muy común identificar $\mathcal{X} \equiv [0, 1]$ y considerar como datos muestrales el vector de p -valores, esto es, se identifica $X \equiv \hat{p}(X) = (\hat{p}_1, \dots, \hat{p}_m)$. Por otro lado, se destaca que los procedimientos de este tipo se conocen con el nombre procedimientos *step-down*.

Teorema 2.1. *Para test estadísticos independientes y cualquier configuración de hipótesis nulas falsas, el procedimiento anterior controla la FDR al nivel α .*

La demostración del teorema 2.1 se basa en el siguiente lema, cuya demostración se encuentra en el apéndice A del artículo de BH.

Lema 2.1. *Para cualesquiera m_0 p -valores independientes correspondientes a las hipótesis nulas verdaderas, con $0 \leq m_0 \leq m$, y $m_1 = m - m_0$ p -valores correspondientes a las hipótesis nulas falsas, el procedimiento de BH satisface la siguiente desigualdad*

$$E(FDP | \hat{p}_{m_0+1}, \dots, \hat{p}_m) \leq \alpha \frac{m_0}{m} \quad c.s. \quad (2.4)$$

La demostración del teorema 2.1 se deduce directamente tomando esperanza en ambos términos de la desigualdad (2.4), resultando que $FDR \leq \alpha \frac{m_0}{m} \leq \alpha$.

Una observación importante de la demostración del teorema 2.1 es que sólo se necesita independencia de los p -valores asociados a las hipótesis nulas verdaderas.

Un punto débil de la formulación presentada hasta el momento para controlar la FDR, es que el procedimiento resultante para controlar la tasa de falsos positivos no surge a priori de ningún problema de maximización o minimización, de manera que no pareciera poseer alguna propiedad de optimalidad. Lo anterior no es cierto, como puede verse en la siguiente proposición:

Proposición 2.2. *Sea $R(q)$ el número de hipótesis rechazadas al nivel q , entonces el procedimiento de BH es óptimo en el sentido que*

$$\hat{p}_{(\hat{k})} = \arg \max_{q \in [0,1]} \left\{ R(q) : \frac{qm}{R(q)} \leq \alpha \right\}. \quad (2.5)$$

Demostración. Notar que $\forall q \in [0, 1]$ tal que $\hat{p}_{(i)} \leq q < \hat{p}_{(i+1)}$, entonces $R(q) = i$. Así, la expresión $\frac{qm}{R(q)}$ crece en el rango de valores para el cual $R(\cdot)$ es constante. Además, dado que $R(q)$ es creciente en q , para resolver (2.5) basta restringirse a los valores de q para los cuales este es igual a algún p -valor. Por demostrar que $q = \hat{p}_{(\hat{k})}$ resuelve el problema. En efecto,

$$\frac{q}{R(q)} = \frac{\hat{p}_{(\hat{k})}}{\hat{k}} \leq \frac{\alpha}{m}. \quad (2.6)$$

Finalmente recordar que por definición $\hat{p}_{(\hat{k})}$ es el más grande de los p -valores que satisfacen la desigualdad en (2.6). ■

Una conclusión de la proposición anterior es que el procedimiento de BH controla la tasa de falsos positivos, maximizando simultáneamente el número de hipótesis nulas rechazadas.

2.2 El problema de la dependencia en los THM

El artículo de BH es notable en muchos sentidos, sin embargo, la hipótesis de independencia de los test estadísticos es muy fuerte en la gran mayoría de los casos. Por otro lado, como quedó de manifiesto en el ejemplo 1.6 del capítulo 1, el tema de la dependencia es un tópico de gran complejidad en el contexto de las pruebas de múltiples hipótesis estadísticas. En este sentido, la publicación de Benjamini y Yekutieli (ver [8]) es una de las primeras en abordar esta dificultad de manera más o menos general cuando se usa la FDR como la tasa de error a controlar.

Uno de los principales aportes del trabajo de BY¹ es demostrar que el procedimiento propuesto por BH también controla la tasa de falsos positivos en caso que los test estadísticos cumplan una determinada propiedad llamada dependencia de regresión positiva (PRDS) y que será definida más adelante. Se destaca el trabajo previo de Benjamini et al. (ver [6]) en el que los autores muestran a través de un estudio de simulación que lo anterior es cierto para test estadísticos distribuidos normalmente y con una estructura de correlación igual y positiva. En este sentido, los resultados de [8] vienen a generalizar los del artículo anterior.

En todo lo que sigue se supone que en el contexto de los test de hipótesis múltiples siempre hay al menos una hipótesis nula verdadera, ya que en caso contrario la FDR es idénticamente cero.

Dos resultados importantes se presentan a continuación:

Teorema 2.2. *Si la distribución conjunta de los test estadísticos satisface la propiedad PRDS sobre el conjunto de índices asociado a las hipótesis nulas verdaderas, entonces el procedimiento de BH controla la FDR a un nivel menor o igual a $\frac{m_0}{m}\alpha$.*

Teorema 2.3. *Si se aplica el procedimiento de BH tomando $\alpha/(\sum_{i=1}^m \frac{1}{i})$ en vez de α en la definición de \hat{k} , entonces sin importar la estructura de dependencia de los test estadísticos, este procedimiento controla la FDR a un nivel menor o igual a $\frac{m_0}{m}\alpha$.*

Es importante destacar la relevancia de los teoremas 2.2 y 2.3, ya que se manifiesta la robustez del procedimiento de BH.

En el contexto de esta sección $X = (X_1, \dots, X_m)$ corresponde a un vector de estadísticos tal que $\forall i = 1, \dots, m$, el p -valor $\hat{p}_i = 1 - F_{X_i}(X_i)$ donde F_{X_i} es la función de distribución acumulada asociada a X_i .

A continuación se detalla en qué consiste la propiedad denominada PRDS y se discute la aplicabilidad de esta propiedad en situaciones prácticas.

Definición 2.1. Un conjunto D se denomina creciente (decreciente) si dado $x \in D$ e $y \geq x$ ($y \leq x$), entonces $y \in D$, donde \geq es una relación de orden en un conjunto que contiene a D .

En la definición anterior típicamente $D \subseteq \mathbb{R}^m$ y la desigualdad $y \geq x$ debe entenderse componente a componente.

Definición 2.2 (Propiedad PRDS sobre I_0). El elemento aleatorio X satisface la propiedad de dependencia de regresión positiva sobre un conjunto de índices I_0 , abreviada PRDS sobre I_0 , si para cualquier conjunto creciente (decreciente) $D \in \mathcal{A} = \sigma(X)$ y para cada $i \in I_0$, $P_\theta(X \in D | X_i = x)$ es no decreciente (no creciente) en x para todo $P_\theta \in \mathcal{P}(\mathcal{A})$.

¹Benjamini y Yekutieli

En estricto rigor la definición anterior es una versión relajada de la propiedad de dependencia de regresión positiva definida por Sarkar [65] del siguiente modo: X satisface la propiedad de dependencia de regresión positiva si para cualquier conjunto creciente $D \in \mathcal{A}$, $P_\theta(X \in D | X_1 = x_1, \dots, X_i = x_i)$ es no decreciente en (x_1, \dots, x_i) para todo $i = 1, \dots, m$.

Se enfatiza que en el contexto de las pruebas de múltiples hipótesis estadísticas el conjunto I_0 , donde es de interés probar la propiedad PRDS, es el conjunto de índices asociado a las hipótesis nulas verdaderas.

Un concepto de dependencia más fuerte que la de regresión positiva es la positividad total multivariada de orden 2, abreviada MTP_2 , y cuya definición se da a continuación.

Definición 2.3. X satisface la MTP_2 si para todo $x, y \in \mathcal{X}$,

$$f(x) \cdot f(y) \leq f(\min\{x, y\}) \cdot f(\max\{x, y\}), \quad (2.7)$$

donde es la función de densidad conjunta (o función de probabilidad conjunta en el caso discreto) y el mínimo y máximo son evaluados componente a componente.

Es importante destacar que la MTP_2 implica la PRDS y que en general, es más fácil de verificar. En la literatura de los THM se pueden encontrar más definiciones de dependencia, que bajos ciertos supuestos implican la PRDS. Una buena referencia al respecto es [8].

Antes de concluir con esta sección, se enuncian algunos resultados que complementan el trabajo de BH.

Teorema 2.4. *Para test estadísticos independientes, el procedimiento propuesto por BH controla la FDR a un nivel menor o igual a $\frac{m_0}{m}\alpha$. Si los estadísticos con los cuales se calculan los p-valores son v.a's continuas, entonces la FDR es exactamente igual a $\frac{m_0}{m}\alpha$.*

Teorema 2.5. *Para test estadísticos independientes unilaterales, si la distribución en cada una de las hipótesis nulas compuestas es estocásticamente más pequeña que la distribución nula bajo la cual los p-valores fueron calculados, entonces el procedimiento de Benjamini y Hochberg controla la FDR a un nivel menor o igual a $\frac{m_0}{m}\alpha$.*

Las demostraciones de todos los teoremas enunciados en esta sección se encuentran en [8].

2.3 Teoría de decisión múltiple: el principio de partición

En esta sección² se presenta un procedimiento desarrollado por Finner y Strassburger en [29], que es otra interesante forma de abordar el problema de los THM. Se destaca que, a diferencia de otros enfoques existentes en la literatura, este método es uno de los pocos en los cuales se desarrollan conceptos de optimalidad. Curiosamente, a diferencia de otras

²La notación utilizada en esta sección es un poco diferente de la usada en el resto de la memoria.

metodologías mencionadas en esta memoria, este método no ha sido tan difundido. La explicación de lo anterior se puede deber a que la tasa de error subyacente es la FWER. En cualquier caso, este procedimiento es muy interesante, sobre todo en lo que respecta a una formalización de la teoría de pruebas de múltiples hipótesis estadísticas.

Se comienza esta sección introduciendo un poco de notación. Sea $\mathcal{H} = \{H_i : i \in I\}$ una familia de hipótesis nulas con $I = \{1, \dots, m\}$ y tales que $\emptyset \neq H_i \subset \Theta$ para todo $i \in I$. Además, $H_i \neq H_j$ para $i \neq j, i, j \in I$. Las hipótesis alternativas se denotan por $K_i = \Theta \setminus H_i$. Destacar el cambio notacional empleado en esta sección, ya que las hipótesis nula y alternativa ya no se denotan por H_0 y H_1 . Además, se identifican las hipótesis con el espacio de parámetros subyacente (restringido). Si todas las hipótesis en \mathcal{H} son disjuntas de a pares, entonces \mathcal{H} se denomina disjunto. Si $H_i \subset H_j$ para algunos $i, j \in I$, entonces \mathcal{H} se llama jerárquico. Además, \mathcal{H} se dice cerrado para la intersección si para todo $\emptyset \neq J \subseteq I$, entonces $\bigcap_{j \in J} H_j \in \mathcal{H}$ o $\bigcap_{j \in J} H_j = \emptyset$. Adicionalmente, si $H_I = \bigcap_{i \in I} H_i \in \mathcal{H}$, entonces H_I se denomina hipótesis global. Por otro lado, una hipótesis $H_i \in \mathcal{H}$ se dice minimal (maximal) si no existe $H_j \in \mathcal{H}$ tal que $H_j \subset H_i$ ($H_j \supset H_i$). Si $\theta \in \Theta$ es el verdadero parámetro, entonces H_i se denomina verdadera si $\theta \in H_i$. En este sentido, $I(\theta) = \{i \in I : \theta \in H_i\}$ es el conjunto de índices asociado a las hipótesis nulas verdaderas, cuando θ es el verdadero parámetro. En este contexto, un test de hipótesis múltiple no aleatorizado se designa por $\varphi = (\varphi_i : i \in I)$ donde $\varphi_i : \mathcal{X} \rightarrow \{0, 1\}$ es un test de hipótesis no aleatorizado, para todo $i \in I$. También se pueden definir los test de hipótesis múltiples aleatorizados vía la extensión obvia, sin embargo, en este caso no es tan claro como interpretar la probabilidad de error de tipo I.

Un test de hipótesis múltiple $\varphi = (\varphi_i : i \in I)$ se dice de nivel múltiple $\alpha \in [0, 1]$ si

$$\forall \theta \in \Theta, \quad P_\theta \left(\bigcup_{i \in I(\theta)} \{\varphi_i = 1\} \right) \leq \alpha, \quad (2.8)$$

donde la convención $\bigcup_{i \in \emptyset} A_i = \emptyset$ es usada. El conjunto de todos los test múltiples φ para \mathcal{H} que satisfacen la condición (2.8) es denotado por $\Phi_\alpha(\mathcal{H})$. Se puede demostrar que la condición (2.8) es equivalente a

$$\forall \emptyset \neq J \subseteq I, \forall \theta \in H_J = \bigcap_{j \in J} H_j : P_\theta \left(\bigcup_{j \in J} \{\varphi_j = 1\} \right) \leq \alpha. \quad (2.9)$$

Nótese que la expresión anterior equivale a lo que se conoce como controlar la FWER en sentido fuerte.

Otra definición interesante es la siguiente. Un test múltiple $\varphi = (\varphi_i : i \in I)$ se dice de nivel local α si $\varphi_i \in \Phi_\alpha(\{H_i\})$, para todo $i \in I$. Se denota por $\Phi_\alpha^{loc}(\mathcal{H})$ el conjunto de los test múltiples de nivel local α . Se puede demostrar que si \mathcal{H} es disjunto, entonces $\Phi_\alpha(\mathcal{H}) = \Phi_\alpha^{loc}(\mathcal{H})$.

Los siguientes son conceptos que se inventaron para evitar contradicciones en los resultados de los test de hipótesis múltiples.

Definición 2.4. Un test de hipótesis múltiple $\varphi = (\varphi_i : i \in I) \in \Phi_\alpha(\mathcal{H})$ se dice:

- (a) Coherente, si $\forall i, j \in I: (H_i \subset H_j \Rightarrow \varphi_i \geq \varphi_j)$.
- (b) Consonante, si $\forall i \in I, \forall x \in \mathcal{X}: (\varphi_i(x) = 1 \text{ y } \exists H_j \supset H_i \Rightarrow \max_{r: H_r \supset H_i} \varphi_r(x) = 1)$.

Conceptualmente, la coherencia es la noción de mayor interés para la teoría de los test de hipótesis múltiples, ya que si $H_i \subset H_j$ y H_j es rechazada, entonces lo más razonable es que la hipótesis H_i sea rechazada.

Una forma de comparar test múltiples es la siguiente:

Definición 2.5. Sean $\varphi^1, \varphi^2 \in \Phi_\alpha(\mathcal{H})$. Entonces:

- (a) $\varphi^1 \geq \varphi^2$, si $\forall i \in I, \forall x \in \mathcal{X}: \varphi_i^1(x) \geq \varphi_i^2(x)$.
- (b) $\varphi^1 > \varphi^2$, si $\varphi^1 \geq \varphi^2$ y $\exists i \in I, \exists x \in \mathcal{X}$ tq $\varphi_i^1(x) > \varphi_i^2(x)$.
- (c) Si $\varphi^1 > \varphi^2$, entonces se dice que φ^2 es un test más conservativo que φ^1 .

Un resultado para construir test múltiples que sean coherentes es:

Teorema 2.6 (Principio de clausura). *Sea \mathcal{H} cerrado bajo intersecciones y $\varphi \in \Phi_\alpha^{loc}(\mathcal{H})$. Se define*

$$\forall i \in I: \bar{\varphi}_i = \min_{j: H_j \subseteq H_i} \varphi_j. \quad (2.10)$$

Entonces $\bar{\varphi} = (\bar{\varphi}_i : i \in I) \in \Phi_\alpha(\mathcal{H})$ y $\forall \emptyset \neq I' \subset I, \bar{\varphi}' = (\bar{\varphi}_i : i \in I') \in \Phi_\alpha(\{H_i : i \in I'\})$. Además, $\bar{\varphi}$ y $\bar{\varphi}'$ son coherentes.

Una propiedad interesante de señalar es la siguiente: si \mathcal{H} es cerrado para la intersección y si $\varphi \in \Phi_\alpha^{loc}(\mathcal{H})$ es coherente, entonces $\varphi = \bar{\varphi} \in \Phi_\alpha(\mathcal{H})$, donde $\bar{\varphi}$ está dado por (2.10) (ver [75]). En relación al teorema (2.6) se desprende que no hay restricciones para considerar una familia de hipótesis $\bar{\mathcal{H}}$, cerrada para la intersección, generada por una familia $\mathcal{H} = \{H_i : i \in I\}$ tal que $\bar{\mathcal{H}} = \{H \subset \Theta : \exists J \subseteq I, H_J = H \neq \emptyset\}$, donde H_J está dado por (2.9). Por lo tanto, $\bar{\mathcal{H}}$ será denominada la cerradura de \mathcal{H} .

Los niveles de las componentes de un test múltiple φ de nivel local α son a menudo denominados niveles nominales, mientras que el nivel de una componente de $\bar{\varphi}$, esto es,

$$\alpha_i = \sup_{\theta \in H_i} P_\theta(\bar{\varphi}_i = 1), \quad (2.11)$$

es a veces llamado el verdadero nivel de $\bar{\varphi}_i$. Notar que la región de aceptación de $\bar{\varphi}_i$ está dada por

$$\{\bar{\varphi}_i = 0\} = \bigcup_{j: H_j \subseteq H_i} \{\varphi_j = 0\}, \quad (2.12)$$

y que esta región puede ser mucho más grande que las regiones de aceptación de $\{\varphi_i = 0\}$ correspondientes a los test de nivel local α , φ_i . De lo anterior se deduce que una aplicación esquemática del principio de clausura puede llevar a procedimientos innecesariamente conservativos.

Teorema 2.7. *Sea $\varphi \in \Phi_\alpha(\mathcal{H})$ y $\underline{\varphi} = (\underline{\varphi}_i : i \in I)$ con*

$$\underline{\varphi}_i = \max_{j: H_j \supseteq H_i} \varphi_j, \quad (2.13)$$

$i \in I$. Entonces $\underline{\varphi}$ cumple las siguientes propiedades:

- (a) $\underline{\varphi} \in \Phi_\alpha(\mathcal{H})$.
- (b) $\underline{\varphi}$ es coherente.
- (c) $\underline{\varphi} \geq \varphi$.

Uno de las principales conclusiones de este teorema es que no hay razón para usar test múltiples de nivel α que no sean coherentes, ya que pueden ser reemplazados por test que si lo sean y sin exceder el nivel α . A modo de comentario es importante mencionar que la propiedad de coherencia no es sólo un requerimiento natural de los test de hipótesis múltiples sino más bien uno minimal.

Un tema bastante delicado en relación a los test múltiples es qué hipótesis aceptar dado $\varphi \in \Phi_\alpha(\mathcal{H})$. Desde el punto de vista de la teoría de decisión una hipótesis nula H_i rechazada se interpreta como una decisión a favor de la hipótesis alternativa $K_i = \Theta \setminus H_i$. Sin embargo, como fue mencionado por Lehmann en [42], si H_i es aceptado puede ser mejor no concluir nada. Así, en un problema de contraste de múltiples hipótesis hay dos formas de interpretar los resultados. La primera es aplicar el razonamiento anterior a cada hipótesis individual φ_i . No obstante, este procedimiento puede resultar en numerosas inconsistencias y contradicciones. La segunda es intentar de resumir todas las decisiones parciales en una decisión conjunta. Para esto se definen los conjuntos

$$H_{\varphi(x)} = \bigcup_{j: \varphi_j(x)=1} H_j \quad \text{y} \quad K_{\varphi(x)} = \Theta \setminus H_{\varphi(x)}, \quad x \in \mathcal{X}. \quad (2.14)$$

El conjunto $H_{\varphi(x)}$ puede ser interpretado como el de todas las hipótesis rechazadas por una de las componentes de φ dado $x \in \mathcal{X}$. Así, la decisión inducida por φ , usando leyes de Morgan, es

$$K_{\varphi(x)} = \bigcap_{j: \varphi_j(x)=1} K_j, \quad (2.15)$$

lo que va en la línea de Lehmann (ver [42]). Esto sugiere que una hipótesis nula H_i tal que $H_i \subseteq H_{\varphi(x)}$ debiera ser rechazada. Lo anterior llevó a Finner (ver [27]) a dar la siguiente definición:

Definición 2.6. Un test múltiple $\varphi \in \Phi_\alpha(\mathcal{H})$ se dice fuertemente coherente si

$$\forall x \in \mathcal{X}, \forall i \in I : (H_i \subseteq H_{\varphi(x)} \Rightarrow \varphi_i(x) = 1). \quad (2.16)$$

Teorema 2.8 (Principio fuerte de clausura de Finner). *Sea $\varphi \in \Phi_\alpha(\mathcal{H})$ y $\bar{\varphi} = (\bar{\varphi}_i : i \in I)$ con*

$$\bar{\varphi}_i = \min_{\theta \in H_i} \max_{j \in I(\theta)} \varphi_j, \quad \forall i \in I. \quad (2.17)$$

Entonces $\bar{\varphi}$ cumple las siguientes propiedades:

- (a) $\bar{\varphi} \in \Phi_\alpha(\mathcal{H})$.
- (b) $\bar{\varphi}$ es fuertemente coherente.
- (c) $\bar{\varphi} \geq \varphi$.
- (d) $\bigcup_{i \in I(\theta)} \{\varphi_i = 1\} = \bigcup_{i \in I(\theta)} \{\bar{\varphi}_i = 1\}, \quad \forall \theta \in \Theta$.
- (e) $H_{\varphi(x)} = H_{\bar{\varphi}(x)}, \quad \forall x \in \mathcal{X}$.

A continuación se introduce el principio de partición, el cual es uno de los principales aportes de la metodología propuesta por Finner y Strassburger en [29].

Teorema 2.9 (Principio de partición general (GPP)). *Sea $\mathcal{H} = \{H_i : i \in I\}$ una familia de hipótesis, $\Theta_J = \{\Theta_j : j \in J\}$ una partición de Θ' con un conjunto adecuado de índices J y $\bigcup_{i \in I} H_i \subseteq \Theta' \subseteq \Theta$ tal que*

$$\forall j \in J : \Theta_j \neq \emptyset \quad y \quad \forall i \in I : \exists J(i) \subseteq J, \quad H_i = \bigcup_{j \in J(i)} \Theta_j. \quad (2.18)$$

Sea además $\phi = (\phi_j : j \in J) \in \Phi_\alpha^{loc}(\Theta_J)$ y $\varphi = (\varphi_i : i \in I)$ dado por

$$\varphi_i = \min_{j \in J(i)} \phi_j, \quad \forall i \in I. \quad (2.19)$$

Entonces $\varphi \in \Phi_\alpha(\mathcal{H})$ y es fuertemente coherente. Finalmente, para las decisiones conjuntas inducidas por φ y ϕ se tiene que

$$K_{\varphi(x)} = \left(\bigcup_{i \in I: \varphi_i(x)=1} H_i \right)^c \supseteq \left(\bigcup_{j \in J: \phi_j(x)=1} \Theta_j \right)^c = K_{\phi(x)}. \quad (2.20)$$

El hecho de base en el que se sustenta el principio de partición es que si \mathcal{H} es una familia disjunta, entonces $\Phi_\alpha(\mathcal{H}) = \Phi_\alpha^{loc}(\mathcal{H})$. La utilidad de este resultado es que ϕ puede llevar a una decisión más precisa $K_{\phi(x)}$ que φ . Una forma de aplicar lo anterior es definiendo la partición más fina $\Theta_\Theta = \{\{\theta\} : \theta \in \Theta\}$, sin embargo, este enfoque tiene algunas dificultades. Para detalles ver [49].

El siguiente lema muestra que cualquier familia de hipótesis $\mathcal{H} = \{H_i : i \in I\}$ puede ser usada para generar una partición natural de $\bigcup_{i \in I} H_i$.

Lema 2.2. Sea $\Theta_J^* = \{\theta \in \Theta : I(\theta) = J\}$ con $\emptyset \neq J \subseteq I$ y $\mathcal{J} = \{\emptyset \neq J \subseteq I : \Theta_J^* \neq \emptyset\}$. Entonces

$$\Theta_{\mathcal{J}} = \{\Theta_J^* : J \in \mathcal{J}\} \quad (2.21)$$

es una partición de $\bigcup_{i \in I} H_i$ y $H_i = \bigcup_{J \in \mathcal{J}: J \ni i} \Theta_J^*$ para todo $i \in I$. Además, la familia de hipótesis nulas, cerrada para intersecciones, inducida por \mathcal{H} genera la misma partición que \mathcal{H} y para $H_J = \bigcap_{j \in J} H_j \neq \emptyset$, esta es $H_J = \bigcup_{R \in \mathcal{J}: R \supseteq J} \Theta_R^*$.

Observación 2.1. Algunas observaciones importantes:

(a) La partición natural $\Theta_{\mathcal{J}}$ es la más gruesa entre todas las particiones de $\bigcup_{i \in I} H_i$ con la propiedad que cada H_i pueda ser representado como una unión disjunta de conjuntos de la partición subyacente.

(b) Sea $\mathcal{H} = \{H_i : i \in I\}$ cerrado para la intersección, $\Theta_i^{**} = H_i \cap \left(\bigcup_{j: H_j \subset H_i} H_j \right)^c$ para $i \in I$ y $J_p = \{i \in I : \Theta_i^{**} \neq \emptyset\}$. Entonces la partición natural generada por \mathcal{H} (ver Lema 2.2) está dada por

$$\Theta(J_p) = \{\Theta_i^{**} : i \in J_p\}, \quad (2.22)$$

esto es, $\Theta(J_p) = \Theta_{\mathcal{J}}$. Si $J_p = I$, entonces cada hipótesis H_i puede ser identificada con Θ_i^{**} y viceversa. En todo caso, en general $I \neq J_p$.

En lo que sigue se muestran otras versiones del principio de partición.

Teorema 2.10 (Principio de partición débil). Sea $\mathcal{H} = \{H_i : i \in I\}$ una familia de hipótesis cerrada para la intersección y $\Theta(J_p) = \{\Theta_i^{**} : i \in J_p\}$ la partición natural generada por \mathcal{H} (descrita como en la observación 2.1 (b)). Para cada $i \in J_p$ se elige un test ϕ_i para contrastar H_i versus K_i tal que

$$\sup_{\theta \in \Theta_i^{**}} P_{\theta}(\phi_i = 1) \leq \alpha, \quad (2.23)$$

esto es, $\phi = (\phi_i : i \in J_p) \in \Phi_{\alpha}^{loc}(\Theta(J_p))$. Se define $\varphi = (\varphi_i : i \in I)$ por

$$\varphi_i = \min_{j \in J_p: \Theta_j \subseteq H_i} \phi_j, \quad \forall i \in I. \quad (2.24)$$

Entonces $\varphi \in \Phi_{\alpha}(\mathcal{H})$ y es fuertemente coherente.

Teorema 2.11 (Principio de partición fuerte). Bajo las suposiciones del teorema 2.10, se elige un test φ_i para contrastar H_i versus K_i tal que

$$\forall i \in J_p : \varphi_i = \min_{j \in J_p: \Theta_j \subseteq H_i} \phi_j \quad (2.25)$$

y

$$\forall i \in J_p : \sup_{\theta \in \Theta_i} P_\theta(\varphi_i = 1) \leq \alpha. \quad (2.26)$$

Además, sea $\varphi_i = \min_{j \in J_p: \Theta_j \subseteq H_i} \phi_j$ para $i \in I \setminus J_p$. Entonces $\varphi = (\varphi_i : i \in I) \in \Phi_\alpha(\mathcal{H})$ y es fuertemente coherente.

Finner y Strassburger en [29] afirman, y lo muestran a través de algunos ejemplos, que la aplicación del principio débil o fuerte de partición puede llevar a procedimientos más potentes que aquellos obtenidos a través del principio de clausura.

2.4 La FDP vista como proceso estocástico

Genovese y Wasserman (ver [34]) extienden algunos resultados asociados al FDR y a su control, y desarrollan un enfoque basado en procesos estocásticos para resolver el problema de los test de hipótesis múltiples, modelando la FDP como una función aleatoria en términos de los p -valores individuales de cada test. Se destaca lo innovadora de esta propuesta, ya que antes de esta no existían en la literatura trabajos en tal dirección.

Para comenzar algunas definiciones:

Definición 2.7. Se definen la FDP (false discovery proportion) y FNP (false non-discovery proportion) por

$$FDP = \begin{cases} \frac{V}{R} & \text{si } R > 0, \\ 0 & \text{si } R = 0, \end{cases} \quad (2.27)$$

y

$$FNP = \begin{cases} \frac{T}{Q} & \text{si } R < m, \\ 0 & \text{si } R = m. \end{cases} \quad (2.28)$$

Además, se definen la $FDR = E_\theta(FDP)$ (false discovery rate) y $FNR = E_\theta(FNP)$ (false non-discovery rate).

Para todo $i = 1, \dots, m$, sea el indicador

$$\mathfrak{H}_i = \begin{cases} 1 & \text{si la } i\text{-ésima hipótesis nula es falsa,} \\ 0 & \text{si la } i\text{-ésima hipótesis nula es verdadera.} \end{cases} \quad (2.29)$$

Se definen los vectores $\mathfrak{H} = (\mathfrak{H}_1, \dots, \mathfrak{H}_m)$ y $\hat{p} = (\hat{p}_1, \dots, \hat{p}_m)$, donde para todo $i = 1, \dots, m$ \hat{p}_i , es el p -valor asociado a la i -ésima hipótesis.

En lo que sigue se usa el siguiente modelo mixto de efectos aleatorios (también llamado jerárquico), utilizado antes por Efron et. al en [23]:

$$\mathfrak{H}_1, \dots, \mathfrak{H}_m \stackrel{iid}{\sim} \text{Bernoulli}(a), \quad (2.30)$$

$$\Xi_1, \dots, \Xi_m \stackrel{iid}{\sim} \mathcal{L}_{\mathcal{F}}, \quad (2.31)$$

$$\hat{p}_i | \mathfrak{H}_i = 0, \Xi_i = \xi_i \sim \text{Uniforme}(0,1), \quad (2.32)$$

$$\hat{p}_i | \mathfrak{H}_i = 1, \Xi_i = \xi_i \sim \xi_i, \quad (2.33)$$

donde $a \in [0, 1]$, Ξ_1, \dots, Ξ_m denotan funciones de distribución y $\mathcal{L}_{\mathcal{F}}$ es una medida de probabilidad arbitraria sobre la clase \mathcal{F} de las funciones de distribución que dominan estocásticamente a la distribución uniforme en $(0, 1)$, a denotar por la letra U , es decir, $\xi(t) \leq U(t) \forall t \in \mathbb{R}$ y $\forall \xi \in \mathcal{F}$.

De lo anterior se concluye que la distribución marginal de los p -valores es

$$G = (1 - a)U + aF, \quad (2.34)$$

donde $F(t) = \int \xi(t) d\mathcal{L}_{\mathcal{F}}(\xi)$ y así, $G \leq U$. En el contexto de test de hipótesis múltiples es más común considerar un modelo en el cual $\mathfrak{H}_1, \dots, \mathfrak{H}_m$ son valores binarios fijos no aleatorizados y desconocidos, no obstante, este enfoque de efectos aleatorios presentado aquí puede ser más intuitivo. Finalmente, $m_0 = \sum_{i=1}^m (1 - \mathfrak{H}_i)$, de donde resulta que $m_0 \sim \text{Binomial}(m, 1 - a)$.

Uno de los principales aportes del trabajo de Genovese y Wasserman (ver [34]) es hacer un análisis basado en procesos estocásticos de los test múltiples, que es lo que se presenta a continuación. Se definen los procesos FDP como

$$\Gamma(t) \equiv \Gamma(t, \hat{p}, \mathfrak{H}) = \frac{\sum_{i=1}^m \mathbb{1}\{\hat{p}_i \leq t\}(1 - \mathfrak{H}_i)}{\sum_{i=1}^m \mathbb{1}\{\hat{p}_i \leq t\} + \mathbb{1}\left\{\min_{i=1, \dots, m} \hat{p}_i > t\right\}}, \quad (2.35)$$

y FNP como

$$\Xi(t) \equiv \Xi(t, \hat{p}, \mathfrak{H}) = \frac{\sum_{i=1}^m \mathbb{1}\{\hat{p}_i > t\}\mathfrak{H}_i}{\sum_{i=1}^m \mathbb{1}\{\hat{p}_i > t\} + \mathbb{1}\left\{\max_{i=1, \dots, m} \hat{p}_i \leq t\right\}}, \quad (2.36)$$

para todo $t \in [0, 1]$.

En el contexto de esta sección un procedimiento de test múltiple T se entiende como una función medible $T : [0, 1]^m \rightarrow [0, 1]$ donde los p -valores menores a $T(\hat{p})$ son rechazados. En tal caso, T recibe el nombre de umbral.

Sean

$$Q(t) = (1 - a) \frac{t}{G(t)}, \quad (2.37)$$

$$\tilde{Q}(t) = a \left(\frac{1 - F(t)}{1 - G(t)} \right). \quad (2.38)$$

El siguiente lema es un corolario del teorema 1 de Storey en [76].

Lema 2.3. *Bajo el contexto del modelo mixto presentado en esta sección, para $t > 0$,*

$$E_{\theta}(\Gamma(t)) = Q(t)(1 - (1 - G(t))^m), \quad (2.39)$$

$$E_{\theta}(\Xi(t)) = \tilde{Q}(t)(1 - G(t))^m. \quad (2.40)$$

Del lema anterior se aprecia la importancia de determinar $Q(t)$ si se quiere estimar la FDR. En relación a esto, si se dispone de estimaciones \hat{a} y \hat{G} para a y G respectivamente, entonces Q se puede estimar por

$$\hat{Q}(t) = \frac{(1 - \hat{a})t}{\hat{G}(t)}. \quad (2.41)$$

Por otro lado, notar que una de las dificultades esenciales de un procedimiento de test de hipótesis múltiple T , es que $\Gamma(T)$ es la evaluación del proceso estocástico $\Gamma(\cdot)$ en la variable aleatoria T , dependiendo ambos de los datos observados y estando en general correlacionados.

En lo que sigue se muestran algunos resultados asociados a distribuciones límites. Para esto se necesitan algunas definiciones previas. Sean

$$\Lambda_0(t) = \frac{1}{m} \sum_{i=1}^m (1 - \mathfrak{H}_i) \mathbb{1}\{\hat{p}_i \leq t\} \quad (2.42)$$

y

$$\Lambda_1(t) = \frac{1}{m} \sum_{i=1}^m \mathfrak{H}_i \mathbb{1}\{\hat{p}_i \leq t\}. \quad (2.43)$$

Para cada $c \in (0, 1)$ se define

$$\Omega_c(t) = (1 - c)\Lambda_0(t) - c\Lambda_1(t) = \frac{1}{m} \sum_{i=1}^m D_i(t), \quad (2.44)$$

donde $D_i(t) = \mathbb{1}\{\hat{p}_i \leq t\}(1 - \mathfrak{H}_i - c)$ y

$$\mu_c(t) \equiv E_\theta(\Omega_c(t)) = E_\theta(D_1(t)). \quad (2.45)$$

Se puede probar que $\mu_c(t) = (1 - a)t - cG(t)$. En efecto,

$$\begin{aligned} \mu_c(t) &= E_\theta(\mathbb{1}\{\hat{p}_1 \leq t\}(1 - \mathfrak{H}_1 - c)) = E_\theta((1 - \mathfrak{H}_1 - c)E_\theta(\mathbb{1}\{\hat{p}_1 \leq t\}|\mathfrak{H}_1)) \\ &= (1 - c)E_\theta(\mathbb{1}\{\hat{p}_1 \leq t\}|\mathfrak{H}_1 = 0)P_\theta(\mathfrak{H}_1 = 0) - cE_\theta(\mathbb{1}\{\hat{p}_1 \leq t\}|\mathfrak{H}_1 = 1)P_\theta(\mathfrak{H}_1 = 1) \\ &= (1 - c)t(1 - a) - cF(t)a = (1 - a)t - c((1 - a)t + aF(t)) \\ &= (1 - a)t - cG(t). \end{aligned}$$

Teorema 2.12. *Sea W un proceso Gaussiano de media cero y covarianza*

$$\begin{aligned} K_\Omega(s, t) &= (1 - a)(1 - c)[(1 - c)(\text{mín}\{s, t\} - (1 - a)st) + ac(tF(s) + sF(t))] \\ &\quad + ac[cF(\text{mín}\{s, t\}) - acF(s)F(t)]. \end{aligned} \quad (2.46)$$

Entonces

$$\sqrt{m}(\Omega_c - \mu_c) \xrightarrow{d} W \text{ en } \mathcal{D}[0, 1], \quad (2.47)$$

cuando $m \rightarrow \infty$.

El lector interesado en la demostración del teorema anterior, la cual está basada en la teoría de procesos empíricos, puede consultar [34]. La relevancia de este teorema, para la teoría de los test de hipótesis múltiples, es que forma parte del argumento base para la demostración del siguiente teorema.

Teorema 2.13 (Distribución límite del proceso FDP). *Para $t \in [\delta, 1]$ con $\delta > 0$, sea*

$$Z_m(t) = \sqrt{m}(\Gamma_m(t) - Q(t)). \quad (2.48)$$

Sea Z un proceso Gaussiano sobre $(0, 1]$ con media cero y kernel de covarianza

$$K_\Gamma(s, t) = a(1-a) \left(\frac{(1-a)stF(\min\{s, t\}) + aF(s)F(t)(\min\{s, t\})}{G^2(s)G^2(t)} \right). \quad (2.49)$$

Entonces $Z_m \xrightarrow{d} Z$ en $\mathcal{D}[\delta, 1]$ cuando $m \rightarrow \infty$.

Observación 2.2. La restricción sobre $[\delta, 1]$ se debe a que el proceso tiene varianza infinita en cero. Además se puede probar un resultado análogo para el proceso FNP.

En lo que viene, sea \hat{G}_m un estimador de G (función de distribución empírica o estimador de los mínimos mayorantes cóncavos de G , donde este último se define como el ínfimo del conjunto de todas las funciones de distribución acumuladas cóncavas, que son mayores que la función de distribución empírica).

Teorema 2.14. *Sea $\hat{Q}_m(t) = (1-a)t/\hat{G}_m(t)$. Para cualquier $\delta > 0$,*

$$\sqrt{m}(\hat{Q}_m(t) - Q(t)) \xrightarrow{d} W \quad (2.50)$$

en $\mathcal{D}[\delta, 1]$, donde W es un proceso Gaussiano sobre $(0, 1]$ de media cero y kernel de covarianza

$$K_Q(s, t) = Q(s)Q(t) \left(\frac{G(\min\{s, t\}) - G(s)G(t)}{G(s)G(t)} \right). \quad (2.51)$$

Volviendo a la ecuación (2.41), Storey [76] propuso el siguiente estimador para el parámetro a :

$$\hat{a}_0 = \left(\frac{\mathbb{G}_m(t_0) - t_0}{1 - t_0} \right)^+, \quad (2.52)$$

que se conoce como estimador de Storey de a , donde $t_0 \in (0, 1)$ es un parámetro fijo de ajuste y \mathbb{G}_m es la función de distribución empírica de $\hat{p} = (\hat{p}_1, \dots, \hat{p}_m)$.

Teorema 2.15. *Sea $\hat{Q}_m(t) = (1 - \hat{a}_0)t/\hat{G}_m(t)$ con \hat{a}_0 el estimador de Storey de a . Entonces*

$$\sqrt{m}(\hat{Q}_m(t) - Q(t)) \xrightarrow{d} W \quad (2.53)$$

en $\mathcal{D}[0, 1]$ cuando $m \rightarrow \infty$, donde W es un proceso Gaussiano de media cero y kernel de covarianza

$$K(s, t) = \frac{t^2}{(1 - t_0)^2 G^2(s) G^2(t)} [G(s)G(t)t_0(1 - t_0 + G(t)(1 - G(t_0)))R(s, t_0) + G(s)(1 - G(t_0))R(t, t_0) + (1 - G(t_0))^2 R(s, t)], \quad (2.54)$$

donde $R(s, t) = \min\{s, t\} - st$.

En algunos métodos, una forma de decidir un test múltiple es encontrando una cota $t(a, G)$ tal que si un p -valor es menor que $t(a, G)$ entonces se rechaza la correspondiente hipótesis nula, donde t es un funcional conocido, pero a y G no. En tales casos, un procedimiento de test de múltiple posible es considerar el estimador *plug-in* definido como $T_m = t(\hat{a}, \hat{G}_m)$, donde \hat{a} y \hat{G}_m son estimadores de a y G respectivamente. El siguiente teorema valida asintóticamente este tipo de procedimientos.

Teorema 2.16 (Validez asintótica de los métodos *plug-in*). *Se supone que G es cóncava. Sea $T_m = t(\hat{a}, \hat{G}_m)$ un estimador *plug-in* donde \hat{G}_m es la función de distribución empírica o el estimador de los mínimos mayorantes cóncavos de G y $\hat{a} \xrightarrow{P_a} a_0$ para algún $a_0 < a$. Entonces*

$$E_\theta(\Gamma(T_m)) \leq \alpha + o_m(1). \quad (2.55)$$

2.5 Extendiendo la FWER

Cuando el número de hipótesis m a contrastar es muy grande, los procedimientos resultantes de controlar la tasa de error global o *family-wise error rate* (FWER) son muy conservativos. Lo anterior es perjudicial en el sentido que se hace más difícil detectar aquellas características individuales en estudio de los test de hipótesis. Todo esto lleva a tener que definir nuevas tasas de error y procedimientos para controlarlas, que es lo que se expone en esta sección. En particular, se trabaja con la k -FWER, cuya definición es

$$k\text{-FWER} = P_\theta(V \geq k), \quad (2.56)$$

con $k \in I = \{1, \dots, m\}$. La idea detrás de esta tasa es permitir un número mayor de rechazos, controlando la cantidad de falsos rechazos, lo que se traduce en una mayor potencia del procedimiento resultante.

Teorema 2.17. *Sean $H_0^i : \theta_i \in \Theta_0^i$ las hipótesis nulas a contrastar para $i = 1, \dots, m$. Se considera el procedimiento que rechaza cualquier H_0^i para el cual $\hat{p}_i \leq \frac{k\alpha}{m}$. Entonces*

- (i) *Este procedimiento controla la k -FWER al nivel α , es decir, $k\text{-FWER} \leq \alpha$.*
- (ii) *Para este procedimiento puede existir una distribución conjunta de $(\hat{p}_1, \dots, \hat{p}_m)$ tal que $k\text{-FWER} = \alpha$.*

Para la demostración ver pag. 1143 de [43].

Al igual como ocurre con el control de Bonferroni para la FWER, el procedimiento planteado en el teorema 2.17 se puede mejorar, en cuanto a potencia, a través de un procedimiento tipo Holm, como se muestra en el siguiente teorema.

Teorema 2.18. *Bajo el contexto del teorema 2.17, considere el siguiente procedimiento: sean $\alpha_1 \leq \dots \leq \alpha_m$ constantes por definir. Si $\hat{p}_{(1)} > \alpha_1$ no se rechaza ninguna hipótesis. De otro modo, si*

$$\hat{p}_{(1)} \leq \alpha_1, \dots, \hat{p}_{(r)} \leq \alpha_r, \quad (2.57)$$

se rechazan las hipótesis $H_0^{(1)}, \dots, H_0^{(r)}$ donde r es el entero más grande que realiza (2.57). Además, si

$$\alpha_i = \begin{cases} \frac{k\alpha}{m} & \text{si } i \leq k, \\ \frac{k\alpha}{s+k-i} & \text{si } i > k, \end{cases} \quad (2.58)$$

entonces el procedimiento resultante controla la k -FWER, es decir, k -FWER $\leq \alpha$.

Para la demostración ver [43].

Procedimientos como (2.57) se denominan *stepdown*. Un punto interesante de mencionar es por qué no simplemente rechazar las hipótesis asociadas a los menores $k - 1$ p -valores y así controlar la k -FWER. Aunque esta observación es cierta, sigue siendo preferible alguno de los procedimientos planteados en los teoremas 2.17 y 2.18, ya que en estos se pueden rechazar incluso más hipótesis, manteniendo el control de la k -FWER.

Teorema 2.19. *Se verifican las siguientes situaciones:*

- (i) *Sean α_i dados en (2.58). Para cualquier $i \geq k$ existe una distribución conjunta para $\hat{p}_1, \dots, \hat{p}_m$ tal que $m + k - i$ de los \hat{p}_i están uniformemente distribuidos sobre $(0, 1)$ y*

$$P_\theta(\hat{p}_{(1)} \leq \alpha_1, \hat{p}_{(2)} \leq \alpha_2, \dots, \hat{p}_{(i-1)} \leq \alpha_{i-1}, \hat{p}_{(i)} \leq \alpha_i) = \alpha. \quad (2.59)$$

- (ii) *Para el procedimiento descrito en (2.57) con α_i dados por (2.58), no se puede incrementar ninguna de las constantes α_i , para $i \geq k$, sin violar el control de la k -FWER.*

En lo que sigue de esta sección, se enfoca el problema de los test de hipótesis múltiples desde la perspectiva de controlar la FDP (false discovery proportion), que se define como

$$FDP = \frac{\sum_{i \in I(\theta)} \mathbf{1}\{\phi_i = 1\}}{\sum_{i \in I} \mathbf{1}\{\phi_i = 1\}}, \quad (2.60)$$

donde $I(\theta) = \{i \in I : \theta_i \in \Theta_0^i\}$ y ϕ_i es el test de hipótesis individual asociado a la hipótesis nula i , con $i \in I$. En relación a este último, típicamente se tiene que $\{\phi_i = 1\} = \{\hat{p}_i \leq \alpha_i\}$.

BH (ver [4]) propusieron un método para controlar el valor esperado de (2.60), sin embargo, esto no evita que la FDP varíe. Por lo mismo, una opción es controlar del siguiente modo:

$$P_\theta(FDP > \gamma) \leq \alpha, \quad \forall \theta. \quad (2.61)$$

Un poco de notación: dados $\hat{p}_1, \dots, \hat{p}_m$ los p -valores de los test individuales, $\hat{q}_1, \dots, \hat{q}_{|I(\theta)|}$ denotan a los p -valores correspondientes a las hipótesis nulas verdaderas mientras que $\hat{r}_1, \dots, \hat{r}_{m-|I(\theta)|}$ denotan a los p -valores asociados a las hipótesis nulas falsas.

Teorema 2.20. Si $\forall i = 1, \dots, |I(\theta)|$,

$$P_\theta(\hat{q}_i \leq u | \hat{r}_1, \dots, \hat{r}_{m-|I(\theta)|}) \leq u, \quad (2.62)$$

entonces el procedimiento stepdown con α_i definido por

$$\alpha_i = \frac{(\lfloor i\gamma \rfloor + 1)\alpha}{m + \lfloor i\gamma \rfloor + 1 - i} \quad (2.63)$$

controla la FDP en el sentido de (2.61).

En el espíritu del teorema anterior:

Teorema 2.21. Sean $\hat{q}_{(1)} \leq \dots \leq \hat{q}_{(|I(\theta)|)}$ los p -valores ordenados asociados a las hipótesis nulas verdaderas y $\mathfrak{J} = \min\{\lfloor m\gamma \rfloor + 1, |I(\theta)|\}$. Entonces

(i) Para el procedimiento stepdown con α_i dado por (2.63), se tiene que

$$P_\theta(FDP > \gamma) \leq P_\theta \left(\bigcup_{i=1}^{\mathfrak{J}} \left\{ \hat{q}_{(i)} \leq \frac{i\alpha}{|I(\theta)|} \right\} \right). \quad (2.64)$$

(ii) Si la distribución conjunta de los p -valores asociados a las hipótesis nulas verdaderas satisfacen la desigualdad de Simes, esto es,

$$P_\theta \left(\left\{ \hat{q}_{(1)} \leq \frac{\alpha}{|I(\theta)|} \right\} \cup \left\{ \hat{q}_{(2)} \leq \frac{2\alpha}{|I(\theta)|} \right\} \cup \dots \cup \left\{ \hat{q}_{(|I(\theta)|)} \leq \alpha \right\} \right) \leq \alpha, \quad (2.65)$$

entonces $P_\theta(FDP > \gamma) \leq \alpha$.

Como comentario de este último teorema, destacar que hay muchas distribuciones conjuntas de p -valores asociados a variables positivamente dependientes que satisfacen la desigualdad de Simes. En este sentido, este teorema es más general que el teorema 2.20. Para más referencias ver [57] y [56].

Lo que sigue es un lema técnico.

Lema 2.4. Sean $\hat{p}_{(1)}, \dots, \hat{p}_{(t)}$ los p -valores ordenados y $0 = \beta_0 \leq \beta_1 \leq \beta_2 \leq \dots \leq \beta_s \leq 1$ constantes, para algún $s \leq t$.

(i) Entonces

$$P_{\theta}(\{\hat{p}_{(1)} \leq \beta_1\} \cup \{\hat{p}_{(2)} \leq \beta_2\} \cup \dots \cup \{\hat{p}_{(s)} \leq \beta_s\}) \leq t \sum_{i=1}^s \frac{\beta_i - \beta_{i-1}}{i}. \quad (2.66)$$

(ii) Además, el lado derecho de (2.66) es menor que 1 y existe una distribución conjunta de los p -valores que transforma la desigualdad en igualdad.

El teorema 2.21 y el lema 2.4 permiten obtener el siguiente resultado:

Teorema 2.22. Considere el procedimiento stepdown con constantes

$$\alpha'_i = \frac{\alpha_i}{C_{(\lfloor m\gamma \rfloor + 1)}}, \quad (2.67)$$

donde α_i está dado por (2.63) y C_j está definido como

$$C_j = \sum_{i=1}^j \frac{1}{i}. \quad (2.68)$$

Entonces $P_{\theta}(FDP > \gamma) \leq \alpha$.

Test de Hipótesis Múltiples, enfoque Bayesiano

En este capítulo, al igual que en el anterior, se exponen algunas ideas de los test de hipótesis múltiples desde una perspectiva bayesiana. El criterio usado para escoger los temas sigue siendo el de mayor impacto y además, desde la óptica del autor de este documento, la novedad de las técnicas matemáticas. Es importante destacar que, a diferencia del enfoque frecuentista de los test múltiples, el enfoque bayesiano no ha sido muy explotado.

3.1 La versión bayesiana de la FDR

Storey en [77] propone una nueva tasa de error a controlar, denominada *positive false discovery rate*, abreviada pFDR, y cuya definición es

$$pFDR = E_{\theta} \left(\frac{V}{R} \mid R > 0 \right). \quad (3.1)$$

En lo que sigue se presenta una interpretación bayesiana de la pFDR. Para esto es importante destacar que en esta sección la tasa anterior es calculada sobre una región de significancia fija, que no depende de los datos.

Sean T_1, \dots, T_m los estadísticos para contrastar m hipótesis estadísticas y Γ una región de significancia dada. Entonces

$$pFDR(\Gamma) = E_{\theta} \left(\frac{V(\Gamma)}{R(\Gamma)} \mid R(\Gamma) > 0 \right), \quad (3.2)$$

donde $V(\Gamma) = |\{T_i : T_i \in \Gamma, \theta_i \in \Theta_0^i\}|$ y $R(\Gamma) = |\{T_i : T_i \in \Gamma\}|$, con la hipótesis nula $H_0^i : \theta_i \in \Theta_0^i$. En términos de reglas de decisión, $\{\phi_i = 1\}$ ssi $\{T_i \in \Gamma\}$. Al igual que antes, $\mathfrak{H}_i = 0$ cuando la i -ésima hipótesis nula es verdadera y 1 en el otro caso. Se supone además que $\mathfrak{H}_i \stackrel{iid}{\sim} \text{Bernoulli}(a)$.

Teorema 3.1. Sean $H_0^i : \theta_i \in \{\theta_0^i\}$ y $H_1^i : \theta_i \in \{\theta_1^i\}$ con $i = 1, \dots, m$, las hipótesis nulas y alternativas de interés, respectivamente. Sean T_1, \dots, T_m los estadísticos asociados

para contrastar estas m hipótesis y Γ una región de significancia. Los (T_i, \mathfrak{H}_i) son vectores aleatorios i.i.d. tales que $T_i | \mathfrak{H}_i \sim (1 - \mathfrak{H}_i)F_0 + \mathfrak{H}_i F_1$ para alguna distribución nula F_0 y una distribución alternativa F_1 para $i = 1, \dots, m$. Entonces

$$pFDR(\Gamma) = P(\mathfrak{H} = 0 | T \in \Gamma), \quad (3.3)$$

donde $1 - a$ es la probabilidad a priori implícita usada en (3.3).

Observe que los índices en (3.3) se omiten dado que $P(\mathfrak{H}_i = 0 | T_i \in \Gamma)$ es la misma para todo $i = 1, \dots, m$. Así, \mathfrak{H} denota la v.a. indicadora de si la hipótesis nula es verdadera o no y T el estadístico para contrastar dicha hipótesis. Más aún,

$$pFDR(\Gamma) = \frac{(1 - a)P_\theta(T \in \Gamma | \mathfrak{H} = 0)}{(1 - a)P_\theta(T \in \Gamma | \mathfrak{H} = 0) + aP_\theta(T \in \Gamma | \mathfrak{H} = 1)}. \quad (3.4)$$

A pesar de que el modelo anterior supone que la hipótesis nula y alternativa son simples, este último caso puede extenderse al compuesto suponiendo que el parámetro de la hipótesis alternativa es aleatorio, en cuyo caso F_1 es una mezcla de distribuciones alternativas.

Corolario 3.1. *Bajo las condiciones del teorema 3.1,*

$$E_\theta \left(\frac{V(\Gamma)}{R(\Gamma)} | R(\Gamma) > 0 \right) = \frac{E_\theta(V(\Gamma))}{E_\theta(R(\Gamma))}. \quad (3.5)$$

De lo anterior se deduce que no siempre la pFDR captura el comportamiento conjunto de V y R .

A continuación se introduce un análogo del p -valor, en términos de la pFDR, que se denomina q -valor.

Hasta ahora se ha considerado sólo una región de significación, no obstante, en el contexto de test de hipótesis múltiples lo usual es considerar regiones de rechazo anidadas en términos de los niveles de significación. Si F_0 y F_1 tienen soporte común, es posible denotar sin pérdida de generalidad este conjunto de regiones de significación por $\{\Gamma_\alpha : \alpha \in [0, 1]\}$, donde α es tal que $P(T \in \Gamma_\alpha | \mathfrak{H} = 0) = \alpha$. En tal caso, $\alpha' \leq \alpha$ implica que $\Gamma_{\alpha'} \subseteq \Gamma_\alpha$.

Definición 3.1 (q -valor). Para un estadístico observado $T = t$, se define el q -valor de t como

$$q - \text{valor}(t) = \inf_{\{\Gamma_\alpha : t \in \Gamma_\alpha\}} pFDR(\Gamma_\alpha). \quad (3.6)$$

Corolario 3.2. *Bajo las condiciones del teorema 3.1,*

$$q - \text{valor}(t) = \inf_{\{\Gamma_\alpha : t \in \Gamma_\alpha\}} P(\mathfrak{H} = 0 | T \in \Gamma_\alpha). \quad (3.7)$$

Siguiendo la notación de Lehmann y Romano (ver [44]), el p -valor de un estadístico observado $T = t$ se define como

$$p\text{-valor}(t) = \inf_{\{\Gamma_\alpha: t \in \Gamma_\alpha\}} P_\theta(T \in \Gamma_\alpha | \mathfrak{H} = 0), \quad (3.8)$$

de donde se aprecia la interpretación del q -valor como la versión bayesiana del p -valor.

Proposición 3.1. *Sea $G_1(\alpha) = P(T \in \Gamma_\alpha | \mathfrak{H} = 1)$. Si $G_1(\alpha)/\alpha$ es decreciente en α , entonces*

$$\arg \min_{\{\Gamma_\alpha: t \in \Gamma_\alpha\}} P(\mathfrak{H} = 0 | T \in \Gamma_\alpha) = \arg \min_{\{\Gamma_\alpha: t \in \Gamma_\alpha\}} P_\theta(T \in \Gamma_\alpha | \mathfrak{H} = 0). \quad (3.9)$$

La proposición anterior enfatiza el hecho que el q -valor de un estadístico se calcula sobre la misma región de significancia que el p -valor.

Para lo que sigue un poco de notación: $pFDR^T(\Gamma_\alpha)$ denota la pFDR basada sobre los estadísticos originales y $pFDR^P(\{p \leq \alpha\})$ la pFDR basada sobre los p -valores.

Teorema 3.2. *Para m test de hipótesis idénticos se tiene que*

$$pFDR^T(\Gamma_\alpha) = pFDR^P(\{p : p \leq \alpha\}).$$

Esto implica que el q -valor puede ser calculado tanto de los estadísticos originales como de sus p -valores. Además, cuando los estadísticos T_i son independientes y siguen una distribución de mezcla, se cumple que

$$q\text{-valor}(t) = pFDR^P(\{p : p \leq p\text{-valor}(t)\}) \quad (3.10)$$

ssi $G_1(\alpha)/\alpha$ es decreciente en α .

La importancia del resultado anterior es que hace comparable los resultados de test de hipótesis definidos sobre espacios posiblemente diferentes, a través de la pFDR.

En lo que sigue se analiza la pFDR en el contexto de dependencia y propiedades asintóticas, suponiendo que para todos los test se usa la misma región de rechazo.

Teorema 3.3. *Bajo las suposiciones distribucionales de T_1, \dots, T_m y $\mathfrak{H}_1, \dots, \mathfrak{H}_m$ dadas en esta sección, se tiene que*

$$pFDR(\Gamma) = \sum_{k=1}^m \sum_{i_1, \dots, i_k} \frac{1}{k} \sum_{j=1}^k P \left(\mathfrak{H}_{i_j} = 0, \begin{array}{l} T_{i_1}, \dots, T_{i_k} \in \Gamma \\ T_{i_{k+1}}, \dots, T_{i_m} \notin \Gamma \end{array} \mid R(\Gamma) > 0 \right), \quad (3.11)$$

donde la segunda suma se hace sobre todos los subconjuntos distintos de tamaño k de $\{1, \dots, m\}$.

La expresión anterior es intratable, salvo bajo un modelo completamente paramétrico, en cuyo caso se puede calcular esta cantidad o una aproximación numérica de ella. Un resultado relacionado es el siguiente:

Corolario 3.3. Sean $(\mathfrak{H}_1, T_1), \dots, (\mathfrak{H}_m, T_m)$ elementos aleatorios intercambiables. Entonces

$$pFDR(\Gamma_\alpha) = \sum_{k=1}^m P\left(\mathfrak{H}_1 = 0 \mid \begin{array}{l} T_1, \dots, T_k \in \Gamma_\alpha \\ T_{k+1}, \dots, T_m \notin \Gamma_\alpha \end{array}\right) P_\theta(R = k \mid R > 0). \quad (3.12)$$

De estos resultados se desprende que el Teorema 3.1 no se mantiene bajo una estructura general de dependencia de los estadísticos. En lo que sigue se muestra una versión asintótica de este teorema. Antes, algo de notación:

$$V_m(\Gamma_\alpha) := \sum_{i=1}^m (1 - \mathfrak{H}_i) \mathbf{1}\{T_i \in \Gamma_\alpha\}, \quad (3.13)$$

$$S_m(\Gamma_\alpha) := \sum_{i=1}^m \mathfrak{H}_i \mathbf{1}\{T_i \in \Gamma_\alpha\}. \quad (3.14)$$

Teorema 3.4. Sea $G_0(\alpha) = P(T \in \Gamma_\alpha \mid \mathfrak{H} = 0)$. Si con probabilidad 1 se tiene que

$$\sum_{i=1}^m \frac{1 - \mathfrak{H}_i}{m} \rightarrow 1 - a \quad (3.15)$$

y

$$\frac{V_m(\Gamma_\alpha)}{\sum_{i=1}^m (1 - \mathfrak{H}_i)} \rightarrow G_0(\alpha), \quad \frac{S_m(\alpha)}{\sum_{i=1}^m \mathfrak{H}_i} \rightarrow G_1(\alpha), \quad (3.16)$$

para cada $\alpha > 0$, con G_0 y G_1 funciones continuas, cuando $m \rightarrow \infty$. Entonces, para cualquier $\delta > 0$ tal que $(1 - a)G_0(\delta) + aG_1(\delta) > 0$,

$$(i) \quad \lim_{m \rightarrow \infty} \sup_{\alpha \geq \delta} \left| \frac{V_m(\Gamma_\alpha)}{\max\{R_m(\Gamma_\alpha), 1\}} - P_\infty(\mathfrak{H} = 0 \mid T \in \Gamma_\alpha) \right| = 0, \quad c.s.$$

$$(ii) \quad \lim_{m \rightarrow \infty} \sup_{\alpha \geq \delta} \left| FDR_m(\Gamma_\alpha) - P_\infty(\mathfrak{H} = 0 \mid T \in \Gamma_\alpha) \right| = 0,$$

$$(iii) \quad \lim_{m \rightarrow \infty} \sup_{\alpha \geq \delta} \left| pFDR_m(\Gamma_\alpha) - P_\infty(\mathfrak{H} = 0 \mid T \in \Gamma_\alpha) \right| = 0,$$

donde

$$P_\infty(\mathfrak{H} = 0 \mid T \in \Gamma_\alpha) = \frac{(1 - a)G_0(\alpha)}{(1 - a)G_0(\alpha) + aG_1(\alpha)}. \quad (3.17)$$

Las funciones G_0 y G_1 son el error de tipo I y la potencia asintótica respectivamente, de los p -valores como función de α . En general, este último teorema dice que si los estadísticos T_i son débilmente dependientes, entonces la proporción realizada de falsos positivos, la FDR y la p FDR convergen simultáneamente sobre todas las regiones de rechazo de la probabilidad

a posteriori definida en (3.17). Esto permite que muchas de las propiedades, bajo independencia de los q -valores, se mantengan aproximadamente bajo una dependencia débil. Más aún, si uno es capaz de estimar G_0 , G_1 y $1 - a$, entonces para m grande, estas estimaciones proveen buenas aproximaciones para la proporción realizada de falsos positivos, la FDR y la pFDR, para todas las regiones de significación, simultáneamente.

3.2 Exploración de aspectos bayesianos de los THM

En esta sección se explora una visión bayesiana de los test de hipótesis múltiples enfocada a especificar distribuciones a priori para tales situaciones e intentando calcular ciertas cantidades a posteriori que son de interés.

El modelo que se presenta a continuación fue sugerido inicialmente por Scott y Berger en [69]. Sean $X = (X_1, \dots, X_m)$ las variables observadas del problema inferencial donde se hace el supuesto que los $X_i \sim \mathcal{N}(\mu_i, \sigma^2)$ con varianza σ^2 desconocida y son independientes entre si. El objetivo es determinar cuales medias μ_i son no nulas. Sea $\mu = (\mu_1, \dots, \mu_m)$. Para que este modelo esté completo se define, al igual que en secciones anteriores, $\mathfrak{H}_i = \mathbf{1}\{\mu_i \neq 0\}$, para todo $i = 1, \dots, m$. En tal caso, la función de verosimilitud se escribe

$$f(x|\mu, \sigma^2, \mathfrak{H}) = \prod_{i=1}^m \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(x_i - \mathfrak{H}_i\mu_i)^2}{2\sigma^2}\right), \quad (3.18)$$

donde $\mathfrak{H} = (\mathfrak{H}_1, \dots, \mathfrak{H}_m)$.

Un enfoque bayesiano natural de este problema consiste en suponer una probabilidad a priori común $p = 1 - a$ que cada $\mu_i = 0$ y que las μ_i son cero o no de manera independiente. Se destaca que en el contexto de microarrays típicamente se espera que p sea cercano a 1, esto es, que la mayoría de las μ_i sean nulas.

Para las medias μ_i no nulas se supone a priori que $\mu_i \sim \mathcal{N}(0, \mathcal{V})$ con varianza desconocida \mathcal{V} . La media nula podría ser reemplazada por una cantidad desconocida no nula, sin embargo, en el contexto de microarrays y otros, es razonable pensar que $\mu_i > 0$ y $\mu_i < 0$ son igualmente probables a priori.

La especificación bayesiana quedará completa en la medida que se elija una distribución a priori $\pi(p)$ y una distribución a priori conjunta $\pi(\mathcal{V}, \sigma^2)$. En lo que sigue, el interés inferencial estará dado por:

1. $p_i = P(\mathfrak{H}_i = 0|x)$, la probabilidad a posteriori que $\mu_i = 0$, dado los datos. A menudo $1 - p_i$ se conoce con el nombre de probabilidad a posteriori de inclusión.
2. $\pi(\mu_i|\mathfrak{H}_i = 1, x)$ la densidad marginal a posteriori de μ_i , dado que es no nula.
3. $\pi(\mathcal{V}, \sigma^2, p|x)$ la distribución a posteriori conjunta para los hiperparámetros claves.

Enseguida se analiza el problema de la elección de las distribuciones a priori.

En relación con las distribuciones a priori sobre \mathcal{V} y σ^2 , en ausencia de una fuerte información a priori, Scott y Berger recomiendan usar

$$\pi(\mathcal{V}, \sigma^2) = (\mathcal{V} + \sigma^2)^{-2}. \quad (3.19)$$

La motivación de esta elección viene de escribir

$$\pi(\mathcal{V}, \sigma^2) = \pi(\mathcal{V}|\sigma^2)\pi(\sigma^2) \propto \frac{1}{\sigma^2} \left(1 + \frac{\mathcal{V}}{\sigma^2}\right)^{-2} \frac{1}{\sigma^2}, \quad (3.20)$$

donde $\pi(\mathcal{V}|\sigma^2)$ es una distribución a priori propia para \mathcal{V} dado σ^2 , lo cual es necesario dado que \mathcal{V} es un hiperparámetro que no ocurre en todos los modelos, ya que a veces $\mu_i = 0$. Para referencias ver [9]. Notar que $\pi(\mathcal{V}|\sigma^2)$ es escalada por σ^2 , una situación común para hipervarianzas. Además, la forma de $\pi(\mathcal{V}|\sigma^2)$ debiera asegurar que la influencia de esta a priori no fuese demasiada en el análisis.

El uso de la a priori $\pi(\sigma^2) = 1/\sigma^2$ podría parecer problemático debido a que esta a priori es impropia, sin embargo, este parámetro es común a todos los modelos bajo consideración. Además el uso de esta a priori está justificado como una a priori predictiva de matching y por motivos de invarianza de pivotes. Para referencias ver [10].

En relación a una a priori sobre p , una elección natural es considerar $\pi(p) = 1$. No obstante, en algunos casos se dispone de una fuerte información a priori que a menudo indica que p es grande. Una forma conveniente de incorporar esta información es a través de

$$\pi(p) = (\beta + 1)p^\beta, \quad (3.21)$$

donde β es un parámetro ajustable que permite que p se concentre cerca de 1.

Un modo simple de especificar β es a través de la mejor adivinanza \hat{p} de p e interpretarla como una mediana a priori. Así,

$$\beta = \frac{\log(0,5)}{\log(\hat{p})} - 1. \quad (3.22)$$

En lo que sigue se explora el tema de la distribución a posteriori.

Bajo las suposiciones del modelo propuesto en esta sección, la densidad a posteriori de $\Theta = (p, \mathcal{V}, \sigma^2, \mathfrak{H}, \mu)$ es

$$\pi(\Theta|x) = C_1^{-1} f(x|\sigma^2, \mathfrak{H}, \mu) \left[\prod_{i:\mathfrak{H}_i=1}^m \mathcal{N}(\mu_i|0, \mathcal{V}) \right] \pi(\mathfrak{H}|p)\pi(\mathcal{V}, \sigma^2)\pi(p), \quad (3.23)$$

donde $\pi(\mathfrak{H}|p) = \prod_{i=1}^m p^{1-\mathfrak{H}_i}(1-p)^{\mathfrak{H}_i}$, $\mathcal{N}(\mu_i|0, \mathcal{V})$ denota la función de densidad de una variable aleatoria $\mathcal{N}(0, \mathcal{V})$ evaluada en el punto μ_i y C_1 es una constante de normalización.

Lema 3.1. *La distribución a posteriori en (3.23) es propia, es decir, C_1 es finita.*

Lema 3.2. *La distribución a posteriori marginal de $(\mathcal{V}, \sigma^2, p)$ está dada por*

$$\begin{aligned} \pi(\mathcal{V}, \sigma^2, p|x) &= C_1^{-1} \prod_{j=1}^m \left[\frac{p}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-x_j^2}{2\sigma^2}\right) \right. \\ &\quad \left. + \frac{(1-p)}{\sqrt{2\pi(\sigma^2 + \mathcal{V})}} \exp\left(\frac{-x_j^2}{2(\sigma^2 + \mathcal{V})}\right) \right] \pi(\mathcal{V}, \sigma^2) \pi(p). \end{aligned} \quad (3.24)$$

Lema 3.3. *Se tiene que*

$$\begin{aligned} p_i &\equiv P(\mathfrak{H}_i = 0|x) = P(\mu_i = 0|x) \\ &= \int_0^1 \int_0^\infty \int_0^\infty \left[1 + \frac{(1-p)}{p} \sqrt{\frac{\sigma^2}{\sigma^2 + \mathcal{V}}} \exp\left(\frac{x_i^2 \mathcal{V}}{2\sigma^2(\sigma^2 + \mathcal{V})}\right) \right]^{-1} \\ &\quad \times \pi(\mathcal{V}, \sigma^2, p|x) d\mathcal{V} d\sigma^2 dp. \end{aligned} \quad (3.25)$$

Lema 3.4. *Se tiene que*

$$\begin{aligned} \pi(\mu_i | \mathfrak{H}_i = 1, x) &= \int_0^1 \int_0^\infty \int_0^\infty (2\pi\tau^2)^{-1/2} \exp\left(-\frac{1}{2\tau^2}(\mu_i - \rho)^2\right) \\ &\quad \times \pi(\mathcal{V}, \sigma^2, p|x) d\mathcal{V} d\sigma^2 dp, \end{aligned} \quad (3.26)$$

donde $\rho = \mathcal{V}x_i/(\sigma^2 + \mathcal{V})$ y $\tau^2 = \mathcal{V}\sigma^2/(\sigma^2 + \mathcal{V})$.

En lo que sigue se supone que se han calculado los p_i 's y el interés ahora es decidir si las observaciones son ruido ($\mu_i \neq 0$) o no. En este último caso se hablará de señales. Una forma de realizar esto es elegir un punto de corte \mathfrak{C} y afirmar que todas las observaciones x_i con $p_i \leq \mathfrak{C}$ son señales, sin embargo, no es claro cómo escoger \mathfrak{C} .

Desde una perspectiva de teoría de decisión estadística se puede hacer lo siguiente. Sea S_i la acción “declarar que μ_i es una señal” y sea N_i la acción “declarar que μ_i es un ruido”. Entonces una función de pérdida podría ser

$$L(S_i, \mu_i) = \begin{cases} 1 & \text{si } \mu_i = 0, \\ 0 & \text{si } \mu_i \neq 0, \end{cases} \quad (3.27)$$

$$L(N_i, \mu_i) = \begin{cases} 0 & \text{si } \mu_i = 0, \\ c|\mu_i| & \text{si } \mu_i \neq 0, \end{cases} \quad (3.28)$$

donde c es un parámetro ajustable reflejando el costo relativo de cada tipo de error. La pérdida esperada a posteriori de cada acción puede ser calculada como

$$E[L(S_i, \mu_i)|x] = \int_{-\infty}^{\infty} L(S_i, \mu_i) \pi(\mu_i|x) d\mu_i = p_i \quad (3.29)$$

y

$$E[L(N_i, \mu_i)|x] = c(1 - p_i) \int_{-\infty}^{\infty} |\mu_i| \pi(\mu_i | \mathfrak{H}_i = 1, x) d\mu_i. \quad (3.30)$$

Entonces la pérdida esperada a posteriori es minimizada si se toma la acción S_i (llamar a μ_i una señal) siempre que $E[L(S_i, \mu_i)|x] < E[L(N_i, \mu_i)|x]$, esto es, siempre que

$$p_i < \frac{c \int_{-\infty}^{\infty} |\mu_i| \pi(\mu_i | \mathfrak{H}_i = 1, x) d\mu_i}{1 + c \int_{-\infty}^{\infty} |\mu_i| \pi(\mu_i | \mathfrak{H}_i = 1, x) d\mu_i}. \quad (3.31)$$

3.3 Enfoque bayesiano no paramétrico de los THM

Posterior a los años 90' ha habido literalmente una explosión de investigación en estadística bayesiana, en especial, en la denominada no paramétrica. Lo anterior se debe principalmente al perfeccionamiento de los métodos numéricos más el avance tecnológico de los computadores, lo que ha permitido popularizar métodos tales como las cadenas de Markov de Monte Carlo, abreviado MCMC.

En esta sección se presenta un modelo propuesto por Tang et al. (ver [81]), que se basa en una mezcla de procesos de Dirichlet para configurar la distribución de los p -valores. Para esto, se formula un modelo mixto como en [77], bajo una situación de test de hipótesis simple versus compuesta. Basado en lo anterior, es posible obtener ciertas cantidades a posteriori que son de interés en el contexto de los test de hipótesis múltiples. En específico, se estiman $\pi_1 = 1 - \pi_0$ y F_1 usando técnicas bayesianas no paramétricas, tras lo cual se propone un estimador para la pFDR.

En esta sección se usa la siguiente notación: sea $X = (X_1, \dots, X_m)$ el vector de p -valores de los m test de hipótesis y $\mathfrak{H}_i = 1$ si la i -ésima hipótesis nula es falsa. Se supone que la densidad de X_i bajo la hipótesis nula es f_0 y f_1 bajo la alternativa. Además, \mathfrak{H}_i , $i = 1, \dots, m$ son ensayos independientes de variables aleatorias de Bernoulli con probabilidad de éxito $\pi_1 = 1 - \pi_0$ y f_0 es la densidad uniforme sobre $[0, 1]$. Así, los p -valores pueden ser vistos como muestras independientes e idénticamente distribuidas de un modelo mixto con dos componentes:

$$f(x) = (1 - \pi_1)f_0(x) + \pi_1 f_1(x). \quad (3.32)$$

En 1988, Parker y Rothenberg (ver [51]) señalaron que muchas distribuciones sobre el intervalo $[0, 1]$ pueden ser modelados como una mezcla de distribuciones beta. Luego, un modelo mixto no paramétrico es extremadamente útil para modelar los p -valores en este sentido. Así,

$$f_1(x) = \int g(x|a, b) dG(a, b), \quad (3.33)$$

donde G es la función de distribución mezcladora, $g(x|a, b) = x^{a-1}(1-x)^{b-1}/B(a, b)$ es la función de densidad de una beta de parámetros a y b , y $B(a, b) = \int_0^1 x^{a-1}(1-x)^{b-1} dx$.

La densidad de los p -valores sobre la hipótesis alternativa tiende a concentrarse cerca de cero. De esta manera parece razonable modelar esta densidad para que decrezca monótonamente sobre el intervalo $[0, 1]$. Para construir esta característica adicional se debe imponer que G ponga toda la masa de probabilidad sobre el conjunto $\{(a, b) : a < 1, b \geq 1\}$. Se puede probar que el caso $b = 1$ implica que $f_1(1) > 0$ lo que provoca una no identificabilidad en términos del par (π_0, f_1) . En lo que sigue de este artículo se supone que $f_1(1) = 0$.

En el modelo propuesto por Tang et al. G es aleatoria y tomada desde un proceso de Dirichlet (ver [26]). Los parámetros del proceso de Dirichlet $DP(G_0, \tau)$ son una medida base G_0 y un parámetro de precisión escalar $\tau > 0$. La medida G_0 es la mejor adivinanza de lo que la verdadera G se cree que es y está restringida a tener soporte sobre $(0, 1) \times (1, \infty)$. Además, la realización del proceso se concentra alrededor de G_0 en la medida que τ sea más grande.

Todo lo anterior se puede resumir en el siguiente modelo jerárquico, donde a cada X_i le corresponde una variable latente $\theta_i = (a_i, b_i)$,

$$X_i | \theta_i \sim 1 - \pi_1 + \pi_1 g(x_i | \theta_i), \quad (3.34)$$

$$\theta_1, \dots, \theta_n | G \stackrel{i.i.d.}{\sim} G, \quad (3.35)$$

$$G \sim DP(G_0, \tau). \quad (3.36)$$

El modelo anterior es muy útil para desarrollar un algoritmo de MCMC para los cálculos a posteriori, como se mostrará más adelante, donde la medida aleatoria G podría ser integrada más allá de la distribución a priori y la distribución conjunta de los θ_i 's está dada por el esquema de las urnas generalizadas de Polya (ver [14]):

$$\theta_1 \sim G_0(\theta_1), \quad (3.37)$$

$$\theta_{i+1} | \theta_1, \dots, \theta_i \sim \frac{\tau}{\tau + i} G_0(\cdot) + \sum_{j=1}^i \frac{1}{\tau + i} \delta_{\theta_j}(\cdot), \quad \forall i \geq 1, \quad (3.38)$$

donde δ_{θ_j} denota la masa puntual unitaria en θ_j . Lo anterior implica que θ_i 's tienden a compartir valores en común.

En el modelo propuesto por Tang et al. (ver [81]) se reparametriza a como L_a y b como L_b , donde $a = \exp(-|L_a|)$ y $b = \exp(|L_b|)$. Además, se especifica la medida de base G_0 como $G_0(a, b) = \mathcal{N}(L_a | 0, \sigma_a^2) \mathcal{N}(L_b | 0, \sigma_b^2)$ donde, por simplificación de la notación, $\mathcal{N}(x | u, \sigma^2)$ denota tanto la distribución normal de media u y varianza σ^2 como la correspondiente función de densidad evaluada en el punto x . La distribución a priori para π_1 parametrizada como $\pi_1 = \exp(-|L_\pi|)$ es obtenida desde $L_\pi \sim \mathcal{N}(0, \sigma_\pi^2)$ y es considerada independiente de f_1 .

A continuación se describe brevemente el algoritmo de MCMC que puede ser usado para muestrear desde la distribución a posteriori de (π, f_1) . Sea $\phi = (\phi_1, \dots, \phi_k)$ el vector de elementos distintos de $\theta_1, \dots, \theta_m$ y $s = (s_1, \dots, s_m)$ el vector de configuraciones,

esto es, $s_i = j$ si y sólo si $\theta_i = \phi_j$. Así, $\theta = \{\theta_i : i = 1, \dots, m\}$ es reparametrizado como $\{\phi_1, \dots, \phi_k, s_1, \dots, s_m\}$. Sea m_i , para $i = 1, \dots, k$, el número de elementos s_j que son iguales a i . Sea $-i$ el índice que indica a todas las variables menos la i -ésima. Así, $m_{-i,j}$ será el número de elementos $s_r = j$, donde s_r son calculados para todos los θ 's excepto el i -ésimo y sea k_{-i} el conjunto $\{1, \dots, k\}$ excepto el s_i .

En el esquema estándar de MCMC basado sobre el esquema de urna de Polya generalizado (ver [25]), un paso de Gibbs es usado para muestrear θ_i (equivalentemente s_i) desde su distribución a posteriori

$$p(\theta_i | X, s_{-i}, \phi_{-i}) \propto \sum_{j \in k_{-i}} q_{i,j} \delta_{\phi_j} + q_{i,0} G_i(\theta_i), \quad (3.39)$$

$$dG_i(\theta) \propto f_\theta(X_i) dG_0(\theta), \quad (3.40)$$

$$q_{i,j} \propto m_{-i,j} f_{\phi_j}(X_i), \quad (3.41)$$

$$q_{i,0} \propto \tau \int f_\theta(X_i) dG_0(\theta). \quad (3.42)$$

Cuando G_0 no es conjugada con f , la integral $q_{i,0}$ puede ser difícil de evaluar y tomar una muestra desde G_i será extremadamente complejo. Para resolver el problema anterior, MacEachern y Muller (ver [46]) desarrollaron un método llamado algoritmo sin orificios. En este, el vector de distintos ϕ 's es aumentado con un conjunto adicional de variables como sigue:

$$\underbrace{\{\phi_1, \dots, \phi_k\}}_{\phi_F} \underbrace{\{\phi_{k+1}, \dots, \phi_m\}}_{\phi_E}, \quad (3.43)$$

con la misma a priori independiente $\phi_j \sim G_0$ y la misma definición de s . Todo lo anterior descansa en que $m_j > 0$ para $j = 1, \dots, k$ y $m_j = 0$ para $j = k+1, \dots, m$. Para más detalles acerca del algoritmo ver [81].

Suponiendo que M observaciones fueron recogidas de la simulación de MCMC, donde el superíndice (j) indica la observación que se está usando. Basado sobre la muestra a posteriori de π_1 y $\theta_1, \dots, \theta_m$, se define la j -ésima muestra a posteriori de la densidad de los p -valores por

$$f^{(j)} = 1 - \pi_1^{(j)} + \pi_1^{(j)} f_1^{(j)} \quad (3.44)$$

donde

$$f_1^{(j)}(x) = \frac{\tau \int g(x|\theta) dG_0(\theta) + \sum_{i=1}^m g(x|\theta_i^{(j)})}{\tau + m}, \quad (3.45)$$

es la j -ésima muestra a posteriori de la densidad bajo la hipótesis alternativa. Sea $F^{(j)}$ y $F_1^{(j)}$ la j -ésima muestra a posteriori correspondiente a las funciones de distribución. Así, la

proporción de hipótesis nulas verdaderas π_0 puede ser estimada como

$$\hat{\pi}_{0,Bayes} = E(\pi_0|X) \cong 1 - \frac{1}{M} \sum_{j=1}^M \pi_1^{(j)}. \quad (3.46)$$

Por otro lado, el estimador de Bayes de la pFDR es

$$\widehat{pFDR}_{Bayes}(\alpha) = E\left(\frac{\pi_0\alpha}{F(\alpha)}|X\right) \cong \frac{1}{M} \sum_{j=1}^M \frac{\pi_0^{(j)}\alpha}{F^{(j)}(\alpha)}. \quad (3.47)$$

La consistencia de un estimador es una de las principales formas de validación frecuentistas de los procedimientos bayesianos. Sin bien es cierto, este resultado no es demostrado en [81], posteriormente, Ghosal et al. (ver [35]) probaron la consistencia de (3.46) y (3.47).

3.4 Teoría de decisión múltiple: versión bayesiana

En esta sección se presenta un enfoque basado en la teoría de decisión estadística, muy similar al descrito por Lehmann en [41] y [42]. En específico se introduce una función de pérdida basada en el concepto de tasa de falsos positivos (FDR) y la regla de Bayes asociada.

Sea $\psi = f(\theta)$ un vector de m “contrastes” entre los elementos de θ . De manera informal, se quiere determinar el signo de cada elemento de ψ , dado los datos. En lenguaje de teoría de decisión, para cada ψ_i con $i = 1, \dots, m$ se tomará una acción a_i , con

$$a_i = \begin{cases} 1 & \text{si } \psi_i \text{ es positivo,} \\ -1 & \text{si } \psi_i \text{ es negativo,} \\ 0 & \text{si no es posible determinar el signo de } \psi_i. \end{cases} \quad (3.48)$$

Considerar las siguientes funciones de pérdida:

$$L_1(\psi_i, a_i) = \begin{cases} 1 & \text{si los signos de } \psi_i \text{ y } a_i \text{ difieren,} \\ 0 & \text{en otro caso,} \end{cases} \quad (3.49)$$

$$L_2(\psi_i, a_i) = \begin{cases} 1 & \text{si } a_i = 0, \\ 0 & \text{en otro caso.} \end{cases} \quad (3.50)$$

La idea de ambas funciones de pérdida es contar el número de incorrectas declaraciones de signo para L_1 y el número de signos no declarados para L_2 .

Lewis y Thayer (ver [45]) proponen una función de pérdida que combina L_1 y L_2 como sigue:

$$L_{DFDR}(\psi, a) = \frac{\sum_{i=1}^m L_1(\psi_i, a_i)}{\max\{1, m - \sum_{i=1}^m L_2(\psi_i, a_i)\}} + \left(\frac{\alpha}{2}\right) \frac{\sum_{i=1}^m L_2(\psi_i, a_i)}{m}, \quad (3.51)$$

para un $\alpha \in (0, 1)$ fijo y $a = (a_1, \dots, a_m)$. Las siglas DFDR significan directional false discovery rate.

Para identificar las acciones que minimicen el valor esperado a posteriori de la función de pérdida (3.51) es necesario introducir un poco de notación. Sea $x \in \mathcal{X}$. Si $P(\psi_i > 0|x) > 0,5$ se define $a_i^* = 1$ y $p_i = P(\psi_i < 0|x)$; por otro lado, si $P(\psi_i > 0|x) \leq 0,5$ se define $a_i^* = -1$ y $p_i = P(\psi_i > 0|x)$. Notar que a_i^* y p_i están relacionados de la siguiente manera:

$$E_{\theta|x}(L_1(\psi_i, a_i^*)|x) = p_i. \quad (3.52)$$

Sean $p_{(1)} \leq \dots \leq p_{(m)}$ los p_i 's ordenados. Se define, además, $a^{(k)}$ para $k = 1, \dots, m$ como $a_{(i)}^{(k)} = a_{(i)}^*$ para $i = 1, \dots, k$ y $a_{(i)}^{(k)} = 0$ para $i = k + 1, \dots, m$. Para $k = 0$, $a_{(i)}^{(0)} = 0$ para $i = 1, \dots, m$.

Luego, la pérdida esperada a posteriori para $a^{(k)}$ está dada por

$$E_{\theta|x}(L_{DFDR}(\psi, a^{(k)})|x) = \frac{\sum_{i=1}^k p_{(i)}}{\max\{1, k\}} + \left(\frac{\alpha}{2}\right) \left(1 - \frac{k}{m}\right). \quad (3.53)$$

Se prueba que $a^{(k)}$ minimiza la pérdida esperada a posteriori entre todos los vectores de acciones a que declaran exactamente k signos. Sea k_{DFDR} el valor de k para el cual la pérdida esperada a posteriori dada en (3.53) es minimizada. Luego, la regla de decisión de Bayes para este problema está dado por $\delta_{DFDR}(x) = a^{(k_{DFDR})}$ y el correspondiente riesgo de Bayes es

$$r(\delta_{DFDR}) = E_{\theta,x}(L_{DFDR}(\psi, \delta_{DFDR}(x))) \quad (3.54)$$

$$= E_x \left(\frac{\sum_{i=1}^{k_{DFDR}} p_{(i)}}{\max\{1, k_{DFDR}\}} + \left(\frac{\alpha}{2}\right) \left(1 - \frac{k_{DFDR}}{m}\right) \right). \quad (3.55)$$

Notar que la esperanza en (3.55) es tomada con respecto a la distribución predictiva de x y además, que $p_{(i)}$ y k_{DFDR} dependen de x .

Dado que $E_{\theta|x}(L_{DFDR}(\psi, a^{(0)})|x) = \alpha/2$ para todo $x \in \mathcal{X}$, sigue que $r(\delta_{DFDR}) \leq \alpha/2$. Por lo tanto,

$$E_{\theta|x} \left(\frac{\sum_{i=1}^m L_1(\psi_i, \delta_{DFDR,i}(x))}{\max\{1, m - \sum_{i=1}^m L_2(\psi_i, \delta_{DFDR,i}(x))\}} \right) \leq \frac{\alpha}{2}. \quad (3.56)$$

Esto último dice que la versión bayesiana de la DFDR es acotada por $\alpha/2$ cuando la regla de Bayes δ_{DFDR} es usada.

En lo que sigue se muestra una versión bayesiana del procedimiento propuesto por BH [4]. Sea $k_{DB\&H}$ el valor más grande de $k = 1, \dots, m$ tal que

$$p^{(k)} \leq \left(\frac{\alpha}{2}\right) \left(\frac{k}{m}\right), \quad (3.57)$$

con $k_{DB\&H} = 0$ si ningún tal valor de k existe. Se define $\delta_{DB\&H}(x) = a^{(k_{DB\&H})}$.

Si $k_{DB\&H} = 0$, entonces la pérdida esperada a posteriori para $a^{(k_{DB\&H})}$ es igual a $\alpha/2$. Si $k_{DB\&H} > 0$ la pérdida esperada a posteriori para $a^{(k_{DB\&H})}$ está dada por la ecuación (3.53) como

$$E_{\theta|x}(L_{DFDR}(\psi, a^{k_{DB\&H}})|x) = \frac{\sum_{i=1}^{k_{DB\&H}} p^{(i)}}{k_{DB\&H}} + \left(\frac{\alpha}{2}\right) \left(1 - \frac{k_{DB\&H}}{m}\right). \quad (3.58)$$

De la definición de $k_{DB\&H}$ se tiene que

$$p^{(i)} \leq \left(\frac{\alpha}{2}\right) \left(\frac{k_{DB\&H}}{m}\right), \quad \text{para } i = 1, \dots, k_{DB\&H}. \quad (3.59)$$

Entonces, aplicando la desigualdad (3.59) a (3.58) se obtiene:

$$E_{\theta|x}(L_{DFDR}(\psi, a^{(k_{DB\&H})})|x) \leq \left(\frac{\alpha}{2}\right) \left(\frac{k_{DB\&H}}{m}\right) + \left(\frac{\alpha}{2}\right) \left(1 - \frac{k_{DB\&H}}{m}\right) = \frac{\alpha}{2}. \quad (3.60)$$

Dado que esta desigualdad vale para todo $x \in \mathcal{X}$, resulta que $r(\delta_{DB\&H}) \leq \alpha/2$, de modo que, con δ_{DFDR} ,

$$E_{\theta,x} \left(\frac{\sum_{i=1}^m L_1(\psi_i, \delta_{DB\&H,i}(x))}{\max\{1, m - \sum_{i=1}^m L_2(\psi_i, \delta_{DB\&H,i}(x))\}} \right) \leq \frac{\alpha}{2}. \quad (3.61)$$

Esta última desigualdad dice que $\delta_{DB\&H}$ también controla la DFDR bayesiana propuesta por Lewis y Thayer en [45].

Dado que δ_{DFDR} es la regla de decisión de Bayes, entonces $r(\delta_{DFDR}) \leq r(\delta_{DB\&H})$. Finalmente, se puede demostrar que $k_{DFDR} \geq k_{DB\&H}$. El lector interesado en la demostración puede consultar [45]. En otras palabras, la regla de Bayes δ_{DFDR} declara siempre al menos tantos signos como $\delta_{DB\&H}$.

Aplicaciones Computacionales

En este capítulo se explorarán algunas implementaciones computacionales que permiten resolver el problema de los test de hipótesis múltiples. Al respecto, el software más influyente hoy por hoy es R, pues la mayoría de los grupos de investigación, dedicados a los temas de la inferencia estadística múltiple, han estado implementando rutinas para este programa. En lo que sigue se explicará qué es R y como instalarlo, para posteriormente discutir el paquete `multtest`, el cual implementa una gran cantidad de procedimientos estadísticos para resolver test de hipótesis múltiples, utilizando para esto técnicas basadas en remuestreo.

4.1 El programa R

R es un software libre y gratuito para la computación estadística y la confección de gráficos de gran calidad. Este compila y corre sobre una amplia variedad de plataformas entre las que se pueden nombrar UNIX, Windows y MacOS. Fue creado por Ross Ihaka y Robert Gentleman, y se distribuye de manera gratuita bajo los términos de la *GNU General Public License*.¹ Su desarrollo y distribución son llevados a cabo por varios grupos de investigadores, siendo uno de los más conocidos el *Grupo Nuclear de Desarrollo de R*.

Una de las características más sobresalientes de R es su enorme flexibilidad, dado que la mayoría de las rutinas implementadas se pueden adecuar a los problemas que tenga el usuario. Sin embargo, su curva de aprendizaje no es tan rápida como ocurre con otros programas estadísticos, como por ejemplo SPSS. Desde el punto de vista computacional, R puede describirse como un lenguaje orientado a objetos, lo que puede desanimar a las personas que no tengan alma de programadores. No obstante, esto es lo que le da flexibilidad al programa, ya que se puede adaptar cada implementación computacional a la problemática en estudio. Ahora, volviendo al tema de la dificultad computacional que posee R, este tiene la ventaja de poseer un lenguaje interpretado al igual que Java, y no compilado como C y Fortran entre otros, lo cual significa que los comandos escritos en el teclado son ejecutados directamente sin la necesidad de tener que construir ejecutables.

¹Para mayor información: <http://www.gnu.org/>.

La forma de instalar R varía un poco dependiendo del sistema operativo que se use. En cualquier caso, los archivos necesarios para dejar el programa funcionando se pueden descargar de la página oficial <http://www.r-project.org/> donde también se encuentran manuales y un repositorio de paquetes adicionales, los cuales no vienen incluidos en el programa base. La última versión de este programa es la 2.12.2, la cual está disponible de Febrero de este año y se puede descargar en todas las plataformas antes citadas.

Antes de instalar R es necesario escoger un servidor para descargarlo, siendo lo más recomendable, pero no obligatorio, elegir el que se sitúe más cercano al país en el que uno se encuentre². Para Chile la página de internet es <http://dirichlet.mat.puc.cl/> y está a cargo de la Pontificia Universidad Católica de Chile. Una vez instalado el programa, este se puede ejecutar de distintas formas, dependiendo del sistema operativo en uso. Por ejemplo, desde UNIX, basta abrir una terminal, escribir R y luego apretar *enter*, en cambio en Windows basta con hacerle doble click al icono que se crea de manera automática en el escritorio durante la instalación.

4.2 El paquete “multtest”

Asociados a R hay varios grupos dedicados a áreas especiales de aplicación de modelos estadísticos. Uno de ellos es **Bioconductor**, bioinformáticas con R, el cual tiene entre sus objetivos dar acceso a la comunidad investigadora a una amplia variedad de poderosos procedimientos estadísticos y gráficos, para el análisis de datos de genómica. El paquete `multtest` de R pertenece a Bioconductor, fue desarrollado por Katherine Pollard³, Sandrine Dudoit⁴ y Mark van der Laan⁵, y su última versión es la 2.7.1, disponible de la página <http://cran.r-project.org/web/packages/multtest/index.html> a partir del 3 de Febrero del presente año. Lo que hace este programa, en términos bastante simplificados, es implementar técnicas basadas en remuestreo para una variedad de procedimientos de test de hipótesis múltiples, controlando una amplia cantidad de tasas de error de tipo I, en problemas donde la distribución de los test estadísticos es general, en el sentido que se permite que las variables tengan ciertas estructuras de dependencia entre si. En otras palabras, los procedimientos implementados en este programa permiten estimar la distribución conjunta de los test estadísticos. Lo anterior hace de este paquete un programa bastante interesante y útil.

Una de las particularidades del programa `multtest`, es que sigue la filosofía de R en cuanto se pueden extender las rutinas que vienen implementadas originalmente. Esto permite que el programa sea aplicable en una gran variedad de casos. Para instalarlo basta, una vez abierto R, escribir en su interfaz:

²Lo anterior es básicamente por razones de velocidad de conexión.

³Center for Biomolecular Science and Engineering, University of California, Santa Cruz.

⁴Division of Biostatistics, School of Public Health, University of California, Berkeley.

⁵Division of Biostatistics, School of Public Health, University of California, Berkeley.

```
> install.packages("multtest")
```

Después de elegir el servidor a través del cual descargar el paquete, se comienza a usar, previa ejecución de la instrucción:

```
> library("multtest")
```

tras lo cual este queda operativo. Es importante destacar que esta última secuencia debe ser ejecutada cada vez que se inicie R y se quiera ocupar el paquete `multtest`.

El objeto fundamental de este paquete es la función `MTP`. Para obtener ayuda acerca de las opciones de esta función se digita la secuencia `?MTP`. Para esto también puede ser útil el comando `args(MTP)` el cual entrega las opciones por defecto de la función `MTP`, que se detallan a continuación:

```
> args(MTP)
function (X, W = NULL, Y = NULL, Z = NULL, Z.incl = NULL, Z.test = NULL,
  na.rm = TRUE, test = "t.twosamp.unequalvar", robust = FALSE,
  standardize = TRUE, alternative = "two.sided", psi0 = 0,
  typeone = "fwer", k = 0, q = 0.1, fdr.method = "conservative",
  alpha = 0.05, smooth.null = FALSE, nulldist = "boot.cs",
  B = 1000, ic.quant.trans = FALSE, MVN.method = "mvrnorm",
  penalty = 1e-06, method = "ss.maxT", get.cr = FALSE, get.cutoff = FALSE,
  get.adj = TRUE, keep.nulldist = TRUE, keep.rawdist = FALSE,
  seed = NULL, cluster = 1, type = NULL, dispatch = NULL, marg.null = NULL,
  marg.par = NULL, keep.margpar = TRUE, ncp = NULL, perm.mat = NULL,
  keep.index = FALSE, keep.label = FALSE)
```

En lo que sigue se explica en qué consisten las opciones más importantes de la función anterior: `X` corresponde a la matriz de datos y `W` a una matriz de pesos que puede ser aplicada sobre cada observación. La variable `Y` se usa cuando se observan variables posiblemente censuradas. En algunos casos, covariables adicionales pueden ser medidas sobre las unidades muestrales. En tal caso, `Z` representa dichas covariables. Al respecto, `Z.incl` y `Z.test` sirven cuando se quiere generar modelos lineales usando como covariables a `Z`. Así, estos comandos dicen qué columnas de `Z` incluir en el modelo. En `test` se especifica el test de hipótesis a considerar. Por defecto se utiliza el test `t` de Welch para comparar dos poblaciones con varianzas posiblemente diferentes, sin embargo hay otras opciones. En este sentido, los test que se pueden efectuar son:

- `t.onesamp`: *Test t de medias para una muestra.*
- `t.twosamp.equalvar`: *Test de comparación de medias (2 muestras) con varianzas iguales.*
- `t.twosamp.unequalvar`: *Test de comparación de medias (2 muestras) con varianzas diferentes.*

- `t.pair`: *Test t de comparación de medias (2 muestras) para datos pareados.*
- `f`: *Test F para testear la igualdad de medias en una población.*
- `f.block`: *Test F para testear la igualdad de medias en una población con un diseño de bloques.*
- `lm.XvsZ`: *Test t para los coeficientes del modelo lineal X en función de Z.test con posible covariables adicionales Z.incl.*
- `lm.YvsXZ`: *Test t para los coeficientes del modelo lineal Y en función de X y Z.*
- `coxph.YvsXZ`: *Test t para los coeficientes del modelo hazard de Cox Y en función de X y Z.*

A través de `alternative` se especifica la hipótesis alternativa. Las opciones son `two.sided`, `less` y `greater`. El valor `psi0` especifica el valor nulo para los parámetros de interés. Su valor por defecto es cero. El tipo de tasa de error se describe a través de `typeone`. Las opción por defecto de FWER, pero también se pueden controlar `gFWER`, `TPFP` y `FDR`. Los valores `k` y `q` tienen que ver con el número y la proporción de falsos positivos respectivamente. Con `alpha` se determina el nivel nominal o de significación del test. La distribución nula del test estadístico es estimada por una versión no paramétrica del bootstrap. También es posible hacer esto a través de estadísticos de permutación. La especificación de lo anterior se realiza a través de `nulldist`. Las opciones son “`boot`” y “`perm`”. Finalmente, hay varias opciones para obtener regiones de confianza, parámetros estimados, puntos de corte, *p*-valores ajustados, etc. Algunas opciones para ver los resultados de la función MTP son `print`, `summary` y `plot`. De estos, el más útil es el comando `summary`, ya que, como es de esperar, entrega un resumen indicando la cantidad de hipótesis rechazadas, puntos de corte, *p*-valores ajustados, etc.

4.3 Aplicación: datos de microarray

En esta sección se usa el paquete `multtest` sobre el conjunto de datos de microarray denominado ALL (acute lymphoblastic leukemia). La idea es mostrar el tipo de procedimientos que es posible obtener en R, a través del paquete antes citado. Los datos se encuentran en el paquete ALL de R. Se comienza cargando los paquetes necesarios para efectuar el análisis:

```
> source("http://bioconductor.org/biocLite.R")
> biocLite("ALL")
> biocLite("hgu95av2.db")
> biocLite("genefilter")
> library("ALL")
> library("hgu95av2.db")
> library("genefilter")
> data(ALL)
```

Destacar que el paquete `ALL`, `hgu95av2` y `genefilter` se descargan de Bioconductor. Para esto se usa el comando `biocLite`. En lo que sigue la idea es identificar genes cuyas medidas de expresión están asociadas con posiblemente censuradas variables biológicas y clínicas. En lo que sigue se filtran los genes para que se cumplan los siguientes criterios: (i) Al menos el 20% de los sujetos tenga una medida de intensidad de al menos 100 y (ii) que el coeficiente de variación (es decir, el cociente de la desviación estándar por la media) de las intensidades a través de la muestra esté entre 0.7 y 10. Para esto se utiliza el paquete `genefilter`, como se muestra a continuación:

```
> ffun <- filterfun(pOverA(p=0.2,A=100),cv(a=0.7,b=10))
> filt <- genefilter(2^exprs(ALL),ffun)
> filtALL <- ALL[filt,]
> filtX <- exprs(filtALL)
> pheno <- pData(filtALL)
```

Así, el nuevo conjunto de datos `filtALL` contiene medidas de expresión sobre 431 genes, para 128 pacientes. Para identificar los genes con más alta medidas de expresión promedio en pacientes con células B, comparados con pacientes con células T, se crea la variable indicadora `Bcell` (1 para células B y 0 para células T) y calculamos para cada gen el test de Welch (para dos muestras y varianzas distintas). La tasa de error de tipo I a controlar es FWER, usando un procedimiento de Bootstrap con $B = 100$ iteraciones⁶ para estimar la distribución nula del test estadístico.

```
> table(pData(ALL)$BT)

  B B1 B2 B3 B4  T T1 T2 T3 T4
  5 19 36 23 12  5  1 15 10  2
> Bcell <- rep(0,length(pData(ALL)$BT))
> Bcell[grep("B",as.character(pData(ALL)$BT))] <- 1
> seed <- 99
> BT.boot <- MTP(X=filtX,Y=Bcell,alternative="greater",B=100,
                 method="sd.minP",seed=seed)
```

```
running bootstrap...
iteration = 100
```

Miremos los resultados que se obtienen en el objeto `BT.boot`:

```
> summary(BT.boot)

MTP:  sd.minP
Type I error rate:  fwer
```

⁶En la práctica es comendable hacer más.

```

Level Rejections
1  0.05          196

      Min. 1st Qu. Median   Mean 3rd Qu.  Max. NA's
adjp      0.000  0.0000 0.9900 0.5371  1.0000  1.000   0
rawp      0.000  0.0000 0.0300 0.3570  0.9400  1.000   0
statistic -34.420 -1.5690 2.0120 2.0590  5.3830 22.330   0
estimate  -4.655 -0.3168 0.3814 0.3258  0.9949  4.249   0

```

El método `summary` entrega varios detalles del MTP, por ejemplo, el nombre del procedimiento (`sd.minP` aquí), la tasa de error de tipo I (`fwer` aquí) y el número de rechazos (196 aquí), entre otros.

La siguiente línea de comandos puede ser usada para obtener una lista de los genes que están diferencialmente expresados en las células B vs las células T, entre todos los pacientes a un nivel $\alpha = 0.05$ y controlando la FWER, es decir, genes para los cuales los p -valores ajustados son menores o iguales a 0.05. Al respecto, los paquetes `annotate` y `annaffy` pueden ser usados para obtener información sobre esos genes.

```

> biocLite("annotate")
> biocLite("annaffy")
> library("annotate")
> library("annaffy")
> BT.diff <- BT.boot@adjp <= 0.05
> BT.AffyID <- featureNames(filtALL)[BT.diff]
> mget(BT.AffyID[1:3],env=hgu95av2GENENAME)
$'1065_at'
[1] "fms-related tyrosine kinase 3"

$'1096_g_at'
[1] "CD19 molecule"

$'1102_s_at'
[1] "nuclear receptor subfamily 3, group C, member 1
      (glucocorticoid receptor)"

```

Varios resúmenes gráficos pueden ser obtenidos con el comando `plot`, seleccionando el apropiado argumento para `which`:

```

> postscript("Figura1.eps")
> par(mfrow=c(2,2))
> plot(BT.boot)
> dev.off()

```

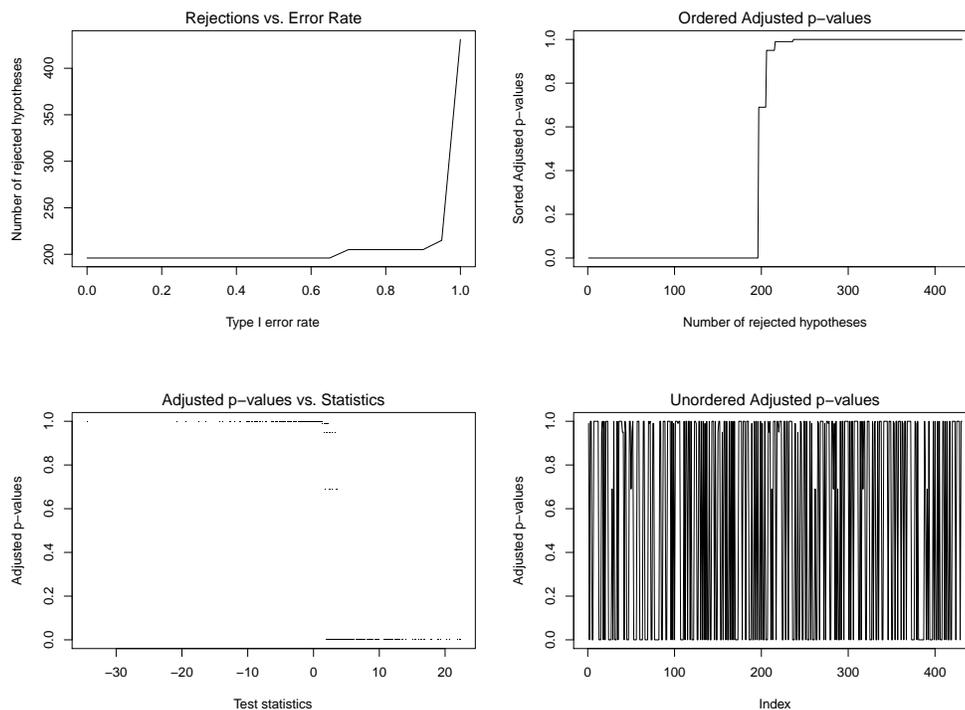


Figura 4.1: Células B vs células T - *FWER control, step-down minP*.

En la figura (4.1) se presentan los resultados del comando `plot`. En el extremo superior izquierdo vemos que el número de rechazos aumenta levemente cuando la FWER nominal es más grande que 0.6 y entonces aumenta rápidamente cuando FWER se acerca a 1. Similarmente, los p -valores ajustados para muchos genes son cercanos a cero y a uno (extremo superior derecho), y los test estadísticos para genes con pequeños p -valores no sobrepasan aquellos genes con p -valores cercanos a uno. De lo anterior se desprende una clara separación entre las hipótesis rechazadas y las hipótesis aceptadas, es decir, entre genes que están diferencialmente expresados y aquellos que no.

Dado un vector de p -valores no ajustados, la función `mt.rawp2adjp` calcula los p -valores ajustados para el marginal control del FWER, que es la tasa de tipo I que se está analizando. Este control marginal puede ser llevado a cabo por los procedimientos de Bonferroni, Holm, Hochberg y Sidak. Más aún, la función `mt.plot` puede ser usada para comparar dichos procedimientos en términos de sus p -valores ajustados.

```
> marg <- c("Bonferroni", "Holm", "Hochberg", "SidakSS", "SidakSD", "BH",
           "TSBH", "ABH")
> BT.marg <- mt.rawp2adjp(rawp=BT.boot@rawp, proc=marg)
> comp.marg <- cbind(BT.boot@adjp, BT.marg$adjp[order(BT.marg$index), -1])
> postscript("Figura2.eps")
> par(mfrow=c(1,1))
> mt.plot(adjp=comp.marg, teststat=BT.boot$statistic, proc=c("SD minP", marg),
```

```

leg=c(0.1,400),col=1:9,lty=1:9,lwd=3)
> tmp <- title("Comparison of FWER and FDR controlling marginal MTPs
and \n step-down minP MTP")
> dev.off()

```

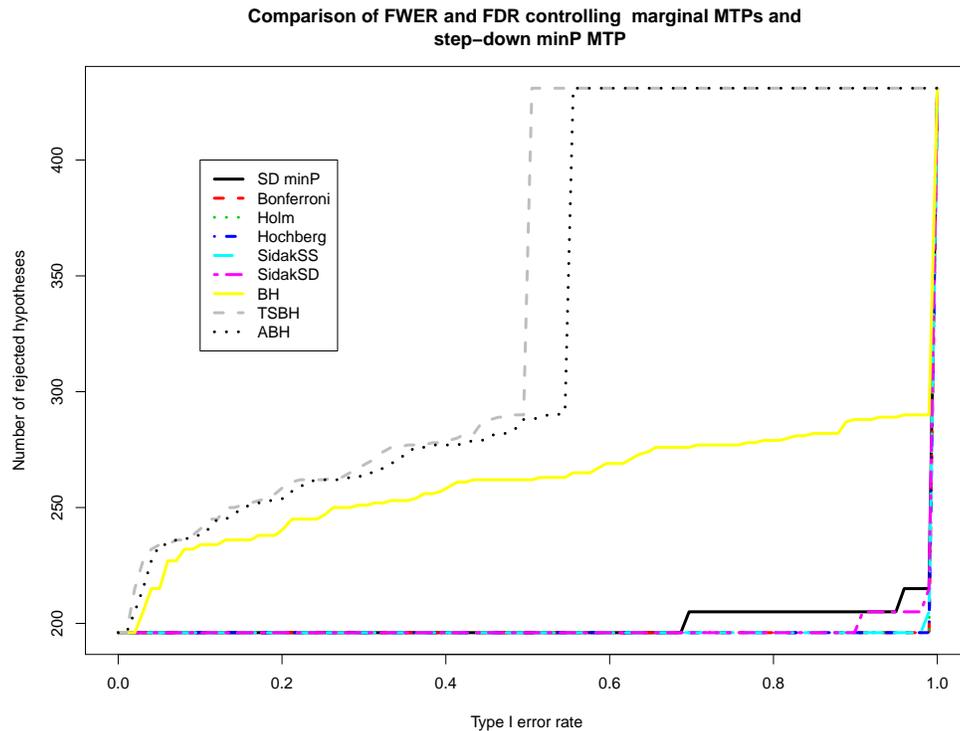


Figura 4.2: Células B vs células T - *Marginal vs conjunto control de FWER.*

En relación a la figura (4.2), todos los procedimientos concuerdan en rechazar pocas hipótesis para un control nominal de la tasa de error de tipo I cercano a cero. Por otro lado, el procedimiento `SD minP`, que toma en cuenta la distribución conjunta del test estadístico asociado, lleva a rechazar más hipótesis que los métodos marginales.

Debido a que los tamaños muestrales asociados a los pacientes con células B y células T son diferentes, y que las medidas de expresión de los genes pueden tener diferentes estructuras de covarianzas en ambas subpoblaciones, se espera que las distribuciones nulas estimadas por Bootstrap y por estadísticos de permutaciones sean distintas. Para obtener la distribución basadas en permutaciones se ocupa la función `mt.minP`. Las líneas de comandos para esto son las siguientes:

```

> set.seed(99)
> BT.perm.old <- mt.minP(X=filtX,classlabel=Bcell,side="upper",B=100)
> names(BT.perm.old)
[1] "index" "teststat" "rawp" "adjp" "plower"

```

En última línea anterior, se ven los elementos que retorna la función `mt.minP`.

```
> set.seed(99)
> BT.perm.new <- MTP(X=filtX,Y=Bcell,alternative="greater",
                    nulldist="perm",B=100,method="sd.minP")
> sum(BT.perm.old$adjp <= 0.05)
[1] 0
> sum(BT.perm.new@adjp <= 0.05)
[1] 0
> sum(BT.perm.new@adjp <= 0.05 & BT.boot@adjp <= 0.05)
[1] 0
```

A un nivel nominal de $\alpha = 0.05$ para el control de FWER, el procedimiento basado en permutaciones identifica cero genes como expresados diferencialmente en pacientes con células B y T respectivamente. En cambio con Bootstrap ese número era 196.

El estadístico de suma de rango de Wilcoxon (también conocido como estadístico de Mann-Whitney) es una alternativa robusta al usual estadístico de dos muestras. En lo que sigue se describe el código para usar este procedimiento:

```
> BT.wilcox <- MTP(X=filtX,Y=Bcell,robust=TRUE,alternative="greater",
                  B=100,method="sd.minP",seed=seed)
running bootstrap...
iteration = 100
> sum(BT.wilcox@adjp <= 0.05)
[1] 196
> sum(BT.wilcox@adjp <= 0.05 & BT.boot@adjp <= 0.05)
[1] 186
```

Al comparar ambos procedimientos (sin y con estadístico robusto), ambos concuerdan en identificar 186 genes.

En lo que sigue, se considera la siguiente situación. En el contexto de análisis de microarray, a menudo se tolera una cierta cantidad de falsos positivos, esperando que dicho número sea pequeño con el número total de hipótesis rechazadas. En tal situación, controlar el FWER es inapropiado y se recurren a técnicas de aumentación para usar el control inicial del FWER para controlar otras tasas de error de tipo I (por ejemplo, $gFWER$ ⁷, $TPFP$ ⁸ y FDR). Destacar que es posible controlar directamente las tasas de error indicadas en el paréntesis, sin embargo, la motivación de lo viene es mostrar como aplicar procedimientos aumentadores.

⁷Forma genérica de denominar a la k -FWER.

⁸Forma genérica de denominar a la $TPFP(q) = P_\theta \left(\frac{V}{R} > q \right)$ con $q \in (0, 1)$.

Controlar el gFWER:

```

> k <- c(5,10,50,100)
> BT.gfwer <- fwer2gfwer(adjp=BT.boot@adjp,k=k)
> comp.gfwer <- cbind(BT.boot@adjp,BT.gfwer)
> mtps <- paste("gFWER(",c(0,k),")",sep="")
> postscript("Figura3.eps")
> mt.plot(adjp=comp.gfwer,teststat=BT.boot@statistic,proc=mtps,
          leg=c(0.1,430),col=1:5,lty=1:5,lwd=3)
> tmp <- title("Comparison of gFWER(k)-controlling
              AMTPs \n based on SD minP MTP")
> dev.off()

```

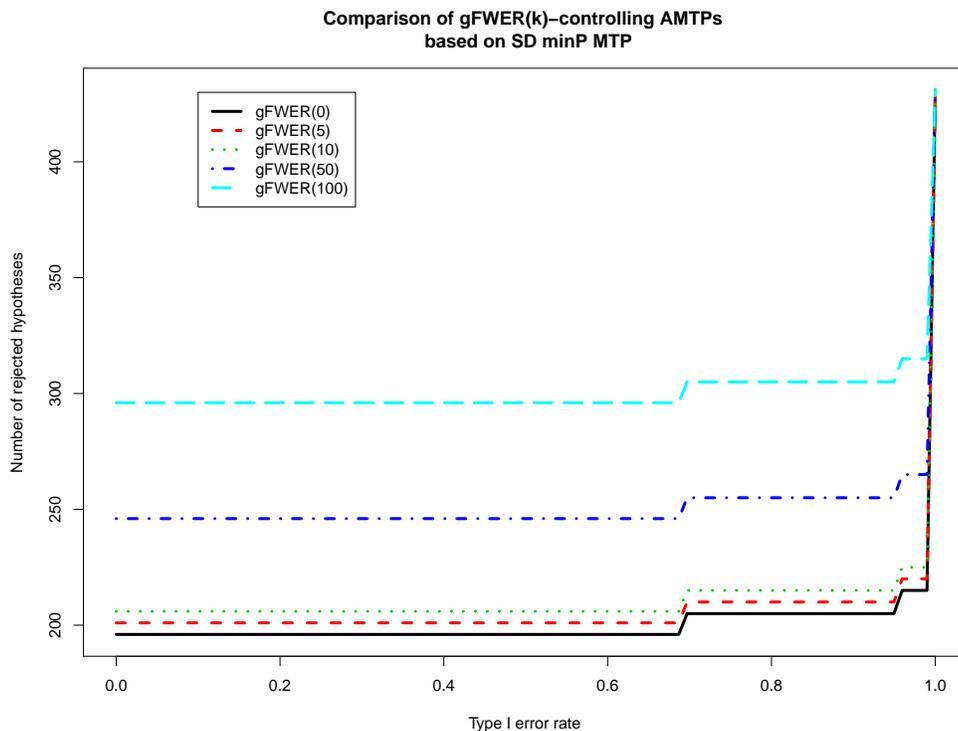


Figura 4.3: Células B vs células T - *Control del gFWER*.

La figura (4.3) muestra el resultado de los comandos anteriores. Ahí se ve como el número de hipótesis rechazadas aumenta linealmente con el número de falsos positivos permitidos, para niveles nominales de α tales que el control inicial del FWER no rechace más de $M - k$ hipótesis.

Controlar el TPPFP:

```

> q <- c(0.05,0.1,0.25)

```

```

> BT.tppfp <- fwer2tppfp(adjp=BT.boot@adjp,q=q)
> comp.tppfp <- cbind(BT.boot@adjp,BT.tppfp)
> mtps <- c("FWER",paste("TPPFP(",q,")",sep=""))
> postscript("Figura4.eps")
> mt.plot(adjp=comp.tppfp,teststat=BT.boot@statistic,proc=mtps,
          leg=c(0.1,430),col=1:4,lty=1:4,lwd=3)
> tmp <- title("Comparison of TPPFP(q)-controlling
              AMTPs \n based on SD minP MTP")
> dev.off()

```

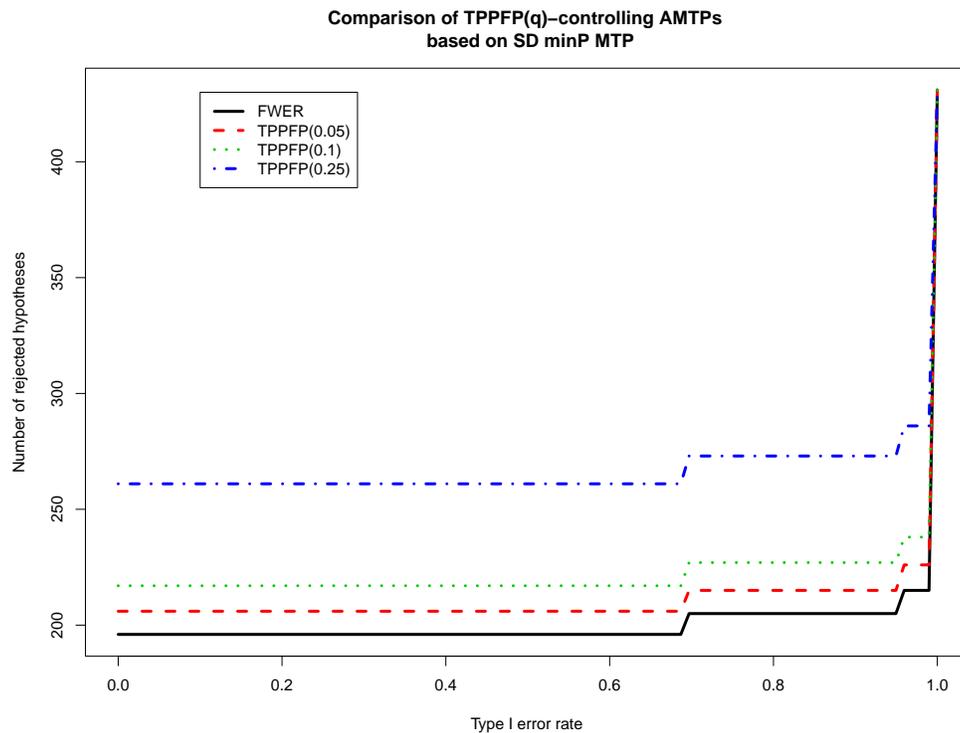


Figura 4.4: Células B vs células T - *Control del TPPFP*.

En la figura (4.4) como el número de hipótesis rechazadas aumenta con la proporción permitida q de falsos positivos.

Controlar el FDR:

Para controlar el FDR hay dos procedimientos posibles debidos a Benjamini y Hochberg (1995), y a Benjamini y Yekutieli (2001).

```

> BT.fdr <- fwer2fdr(adjp=BT.boot@adjp,method="both")$adjp
> BT.marg.fdr <- mt.rawp2adjp(rawp=BT.boot@rawp,proc=c("BY","BH"))
> comp.fdr <- cbind(BT.fdr, BT.marg.fdr$adjp[order(BT.marg.fdr$index),-1])
> mtps <- c("AMTP Cons","AMTP Rest","BY","BH")

```

```

> postscript("Figura5.eps")
> mt.plot(adjp=comp.fdr,teststat=BT.boot@statistic,proc=mtps,
          leg=c(0.1,430),col=c(2,2,3,3),lty=rep(1:2,2),lwd=3)
> tmp <- title("Comparison of FDR-controlling MTPs")
> dev.off()

```

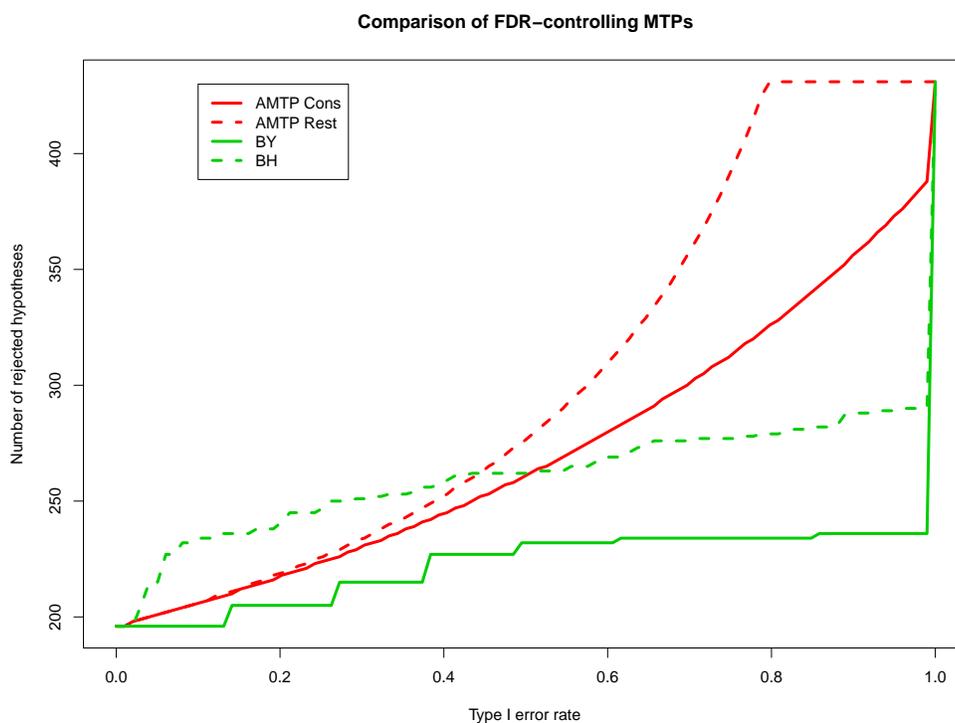


Figura 4.5: Células B vs células T - *Control del FDR*.

La figura (4.5) muestra que las técnicas aumentadoras basadas en cotas conservadoras del FDR (“AMTP Cons” y “AMTP Rest”) son más conservadoras que el procedimiento de Benjamini y Hochberg (“BH”) para valores nominales del FDR menores que 0.4, pero menos conservadoras que “BH” para valores grandes del FDR. Por otro lado, el procedimiento de Benjamini y Yekutieli (una versión más conservadora del procedimiento de “BH”) es el que produce menos hipótesis rechazadas.

Finalmente, se presenta la función EBMP, la cual utiliza métodos bayesianos empíricos para controlar las distintas tasas de error.

```

> ebFdrIc <- EBMP(X=filtX,Y=Bcell,nullldist="ic",typeone="fdr",B=100)
calculating vector influence curve...
sampling null test statistics...
counting guessed false positives...
100

```

```
counting guessed true positives...
```

```
100
```

```
> summary(ebFdrIc)
```

```
EBMTP: common.cutoff
```

```
Type I error rate: fdr
```

```
prior: conservative
```

```
Level Rejections
```

```
1 0.05      341
```

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's
adjp	0.000	0.0000	0.0000463	0.04235	0.03376	0.3338	0
rawp	0.000	0.0000	0.0000000	0.10890	0.06000	1.0000	0
statistic	-34.420	-1.5690	2.0120000	2.05900	5.38300	22.3300	0
estimate	-4.655	-0.3168	0.3814000	0.32580	0.99490	4.2490	0

```
> postscript("Figura6.eps")
```

```
> par(mfrow=c(2,2))
```

```
> plot(ebFdrIc)
```

```
> dev.off()
```

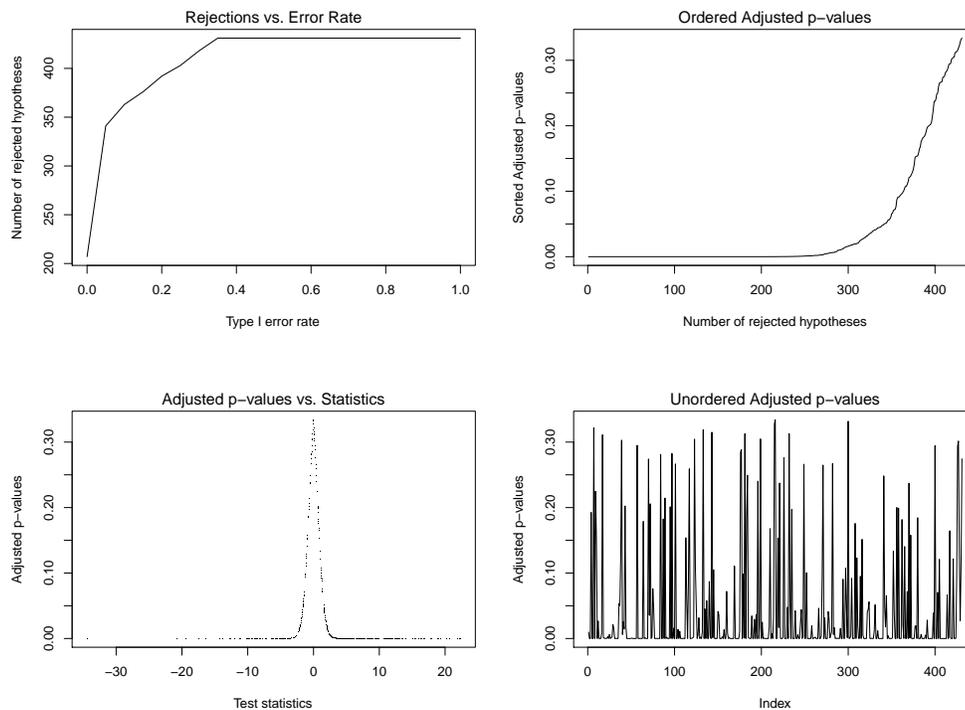


Figura 4.6: Células B vs células T - Control del FDR usando EBMP.

En figura (4.6) se muestran los resultados del código anterior.

Existen varios procedimientos alternativos que se pueden realizar con el paquete `multtest` y que no fueron discutidos en esta sección.

4.4 Comentarios finales y recomendaciones

Los métodos de BH y sus variantes tienden a rechazar muchas más hipótesis que los procedimientos clásicos de Bonferroni y similares. Por lo mismo, se recomienda usar los procedimientos de BH cuando el número de hipótesis a testear sea muy grande, para así obtener procedimientos que sean más potentes.

En caso que, por razones propias del problema en estudio, sea más natural controlar la FWER, existen versiones generalizadas de esta tasa, la $gFWER$, y versiones aumentadas, que proveen procedimientos más potentes, los cuales debieran ser usados en reemplazo de los métodos asociados a la FWER.

Si bien es cierto que la recomendación en general es usar procedimientos asociados a BH, para valores nominales de la FDR mayores que 0.5, existen métodos denominados aumentadores que proveen un mayor número de hipótesis nulas rechazadas y que resultan ser más potentes. En tales casos, la recomendación es usar estos procedimientos aumentadores.

Análisis Convexo aplicado a los Test de Hipótesis Múltiples

En este capítulo se plantea una forma distinta de enfrentar el problema de los test de hipótesis múltiples, utilizando ideas de dualidad provenientes del análisis convexo. A pesar que la idea de usar métodos provenientes del análisis convexo en los problemas de test de hipótesis múltiple es nueva, la aplicación de este tipo de técnicas al estudio de los test de hipótesis no múltiples no es reciente. Los primeros trabajos al respecto son de Witting (ver [85]), y se destacan los aportes de Cvitanic y Karatzas (ver [18]) y Rudloff y Karatzas (ver [55]). Sin perjuicio de lo anterior, esta es la primera vez que se usan herramientas de dualidad convexa para el análisis de contrastes múltiples. Por otro lado, se manifiesta en la literatura la existencia de no muchos resultados de optimalidad estadística, lo que hace que el resultado propuesto en la sección 5.1 sea, al menos, interesante.

La composición de este capítulo es la siguiente: en la sección 5.1 se formula el teorema mencionado en el párrafo anterior. En la sección 5.2 se demuestra dicho teorema. Finalmente, en la sección 5.3 se discuten algunas limitaciones y conclusiones del resultado propuesto.

5.1 Enunciado del Teorema

Teorema 5.1. Sean $\mathcal{P}_i, \mathcal{Q}_i$, con $i = 1, \dots, m$, familias de medidas de probabilidad sobre $(\mathcal{X}_i, \mathcal{A}_i)$ tales que

$$P \ll R_i, \forall P \in \mathcal{P}_i \cup \mathcal{Q}_i, \quad (5.1)$$

donde R_i es una medida de probabilidad sobre $(\mathcal{X}_i, \mathcal{A}_i)$. Se define $\forall i = 1, \dots, m$,

$$\xi_{\mathcal{H}} := \left\{ \frac{dP}{dR_i} : P \in \mathcal{H} \right\}, \quad (5.2)$$

conjunto que se supondrá débil compacto si $\mathcal{H} \in \{\mathcal{P}_i, \mathcal{Q}_i\}$ y convexo si $\mathcal{H} = \mathcal{Q}_i$. Se designa por $\Phi_i := \{\phi : \mathcal{X}_i \rightarrow [0, 1] \text{ medibles}\}$ a la familia de los test de hipótesis aleatorizados asociados a \mathcal{X}_i y por $\Phi^m := \prod_{i=1}^m \Phi_i$ a la familia de los test de hipótesis múltiples aleatorizados.

Para $\alpha \in [0, 1]$ se define

$$\Phi_\alpha^m := \left\{ \varphi \in \Phi^m : \sup_{P \in \mathcal{P}^m} G(P, \varphi) \leq \alpha \right\} \quad (5.3)$$

como la familia de los test de hipótesis múltiples aleatorizados de nivel α , donde

$$G(P, \varphi) := \sum_{i=1}^m \delta_i E^{P_i}(\varphi_i), \quad (5.4)$$

$\delta_1, \dots, \delta_m$ son constantes no negativas tales que

$$\sum_{i=1}^m \delta_i = 1 \quad (5.5)$$

y $P = (P_1, \dots, P_m) \in \mathcal{P}^m := \prod_{i=1}^m \mathcal{P}_i$. Entonces $\exists \tilde{\varphi} \in \Phi_\alpha^m$ tal que

$$\inf_{Q \in \mathcal{Q}^m} H(Q, \tilde{\varphi}) = \sup_{\varphi \in \Phi_\alpha^m} \inf_{Q \in \mathcal{Q}^m} H(Q, \varphi), \quad (5.6)$$

donde

$$H(Q, \varphi) := \sum_{i=1}^m \delta_i E^{Q_i}(\varphi_i), \quad (5.7)$$

y $Q = (Q_1, \dots, Q_m) \in \mathcal{Q}^m := \prod_{i=1}^m \mathcal{Q}_i$. Más aún, $\exists (\tilde{Q}, \tilde{\lambda}) \in \mathcal{Q}^m \times \Lambda_+$, con

$$\Lambda_+ = \left\{ \lambda : \lambda \text{ es medida finita positiva sobre } \prod_{i=1}^m \xi_{\mathcal{P}_i} \right\}, \quad (5.8)$$

que verifica que

$$D(\tilde{Q}, \tilde{\lambda}) = \inf_{\substack{Q \in \mathcal{Q}^m \\ \lambda \in \Lambda_+}} D(Q, \lambda), \quad (5.9)$$

donde

$$D(Q, \lambda) = \sum_{i=1}^m \delta_i E^{R_i} \left(\left[Z_{Q_i} - \int_{\prod_{j=1}^m \xi_{\mathcal{P}_j}} Z_{P_i} d\lambda \right]^+ \right) + \alpha \lambda \left(\prod_{j=1}^m \xi_{\mathcal{P}_j} \right), \quad (5.10)$$

con $Z_{H_i} := \frac{dH_i}{dR_i}$ para $H_i \in \mathcal{P}_i \cup \mathcal{Q}_i$. Finalmente, el test múltiple óptimo $\tilde{\varphi}$ posee la siguiente estructura: $\forall i = 1, \dots, m$,

$$\tilde{\varphi}_i(x_i) = \begin{cases} 1 & \text{si } Z_{\tilde{Q}_i}(x_i) > \int_{\prod_{j=1}^m \xi_{\mathcal{P}_j}} Z_{P_i}(x_i) d\tilde{\lambda}, \\ 0 & \text{si } Z_{\tilde{Q}_i}(x_i) < \int_{\prod_{j=1}^m \xi_{\mathcal{P}_j}} Z_{P_i}(x_i) d\tilde{\lambda}, \end{cases} \quad R_i - \text{c.s. } x_i \in \mathcal{X}_i, \quad (5.11)$$

$E^{R_i}(\tilde{\varphi}_i Z_{P_i}) = \alpha_i$ para $\tilde{\lambda} - \text{c.s. } Z_{P_i} \in \xi_{\mathcal{P}_i}$ y $\sum_{i=1}^m \delta_i \alpha_i = \alpha$.

Observación 5.1. Antes de pasar a la sección siguiente, algunos comentarios relevantes relativos al teorema anterior.

- (a) $\forall i = 1, \dots, m$, las familias de medidas de probabilidad \mathcal{P}_i y \mathcal{Q}_i están asociadas a las hipótesis nula y alternativa respectivamente.
- (b) Es importante destacar que el concepto de regla de decisión usado en este teorema es un poco diferente del dado en las ecuaciones (1.28) y (1.29) de la sección 1.7 del capítulo 1. En cualquier caso, $\varphi \in \Phi^m$ se interpreta como $\varphi(x) = (\varphi_1(x_1), \dots, \varphi_m(x_m))$ donde $x = (x_1, \dots, x_m) \in \prod_{i=1}^m \mathcal{X}_i := \mathcal{X}$.
- (c) La medida $\lambda \in \Lambda_+$ corresponde a una distribución a priori sobre $\prod_{i=1}^m \xi_{\mathcal{P}_i}$ y $\tilde{\lambda} \in \Lambda_+$ se interpreta como una medida menos favorable. Para más detalles y referencias ver pags. 83-86 de [44].
- (d) La regla de decisión óptima resultante tiene la estructura de un test de hipótesis no aleatorizado.
- (e) La integral $\int_{\prod_{j=1}^m \xi_{\mathcal{P}_j}} Z_{\mathcal{P}_i}(x_i) d\lambda$ se interpreta como la integral con respecto a la medida λ del funcional continuo $\prod_{j=1}^m \xi_{\mathcal{P}_j} \ni Z = (Z_1, \dots, Z_m) \rightarrow l(Z, x_i) := \mathfrak{T}_i(Z(x)) \in \mathbb{R}$ con $x_i \in \mathcal{X}_i$ fijo, $Z(x) = (Z_1(x_1), \dots, Z_m(x_m))$ y $\mathfrak{T}_i : \mathbb{R}^m \rightarrow \mathbb{R}$ la función de proyección. Además, se supone que $\mathcal{X} \times \prod_{j=1}^m \xi_{\mathcal{P}_j} \ni (x, Z) \rightarrow Z(x) \in [0, +\infty)^m$ es medible con respecto a la σ -álgebra producto $\mathcal{A} \otimes \mathfrak{B}$, donde $\mathcal{A} = \prod_{i=1}^m \mathcal{A}_i$ (σ -álgebra producto) y \mathfrak{B} es la σ -álgebra asociada a los borelianos de $\prod_{j=1}^m \xi_{\mathcal{P}_j}$.

5.2 Demostración del Teorema

Demostración (del Teorema 5.1). Notar que $\forall i = 1, \dots, m$, $\Phi_i \subseteq B_i$ donde B_i es la bola unitaria cerrada en $L^\infty(\mathcal{X}_i, \mathcal{A}_i, R_i)$. Por el teorema de Banach-Alaoglu-Bourbaki resulta que B_i es débil-* compacta, recordando que $L^\infty(\mathcal{X}_i, \mathcal{A}_i, R_i) = L^1(\mathcal{X}_i, \mathcal{A}_i, R_i)^*$. Para referencias, ver pags. 39-43 de [15]. Se designa por τ_i a la topología débil-* definida sobre $L^\infty(\mathcal{X}_i, \mathcal{A}_i, R_i)$, $i = 1, \dots, m$.

Lema 5.1. $\forall i = 1, \dots, m$, Φ_i es τ_i -compacto.

Demostración (del Lema 5.1). Sea $i \in \{1, \dots, m\}$. Del párrafo anterior se desprende que basta probar que Φ_i es τ_i -cerrado. En efecto, sea $\{\phi_\alpha\}_{\alpha \in D} \subseteq \Phi_i$ una red τ_i -convergente, es decir, $\forall h \in L^1(\mathcal{X}_i, \mathcal{A}_i, R_i)$, $E^{R_i}(\phi_\alpha h) \rightarrow E^{R_i}(\phi h)$. Por demostrar que ϕ toma valores en $[0, 1]$, R_i -c.s. Por contradicción, se supone que $\exists A \in \mathcal{A}_i$ tal que $\{\phi > 1\} \subseteq A$ y $R_i(A) > 0$, y se toma $h(x_i) = \mathbf{1}_A(x_i)$, que pertenece a $L^1(\mathcal{X}_i, \mathcal{A}_i, R_i)$. Entonces, $E^{R_i}(\phi h) > R_i(A)$, sin embargo, $E^{R_i}(\phi h) = \lim_\alpha E^{R_i}(\phi_\alpha h) \leq R_i(A)$, lo cual es una contradicción. Se concluye entonces que $\phi \leq 1$ R_i -c.s. y de manera similar se prueba que R_i -c.s., $\phi \geq 0$. \square

De la afirmación anterior sigue que $\forall i = 1, \dots, m$, (Φ_i, τ_i) es un espacio topológico compacto. Luego, por el teorema de Tychonoff, lo mismo ocurre para el espacio topológico (Φ^m, τ^m) con $\tau^m := \prod_{i=1}^m \tau_i$, la topología producto.

Lema 5.2. Φ_α^m es τ^m -compacto.

Demostración (del Lema 5.2). El argumento clave de esta demostración es que los conjuntos de subnivel de una función s.c.i. son cerrados. Así, basta probar que $\forall P \in \mathcal{P}^m$ el funcional $G(P, \cdot)$ es τ^m -semicontinuo inferior en Φ^m . Dado que el supremo y las combinaciones lineales de funciones s.c.i. son s.c.i., basta con probar que $E^{P_i}(\varphi_i)$ es s.c.i. Pero esto es consecuencia del lema de Fatou. En efecto, si $\varphi_n \rightarrow \varphi_0$ entonces

$$\liminf_{n \rightarrow \infty} E^{P_i}(\varphi_n) \geq \int \liminf_{n \rightarrow \infty} \varphi_n dP_i = E^{P_i}(\varphi_0), \quad (5.12)$$

es decir, $\varphi \rightarrow E^{P_i}(\varphi)$ es s.c.i. □

De manera similar se prueba que el funcional $\Phi_\alpha^m \ni \varphi \rightarrow \inf_{Q \in \mathcal{Q}^m} \sum_{i=1}^m \delta_i E^{Q_i}(\varphi_i) \in \mathbb{R}$ es τ^m -semicontinuo superior, lo que junto a la τ^m -compacidad de Φ_α^m permite concluir, gracias a un resultado general de optimización, la existencia de $\tilde{\varphi} \in \Phi_\alpha^m$ tal que

$$\inf_{Q \in \mathcal{Q}^m} H(Q, \tilde{\varphi}) = \sup_{\varphi \in \Phi_\alpha^m} \inf_{Q \in \mathcal{Q}^m} H(Q, \varphi).$$

Así, queda demostrada la primera parte del teorema, esto es, la condición (5.6).

Lema 5.3.

$$\underline{V} := \sup_{\varphi \in \Phi_\alpha^m} \inf_{Q \in \mathcal{Q}^m} H(Q, \varphi) = \inf_{Q \in \mathcal{Q}^m} \sup_{\varphi \in \Phi_\alpha^m} H(Q, \varphi) := \overline{V}. \quad (5.13)$$

Observación 5.2. Antes de continuar con la demostración el lema anterior, algunas observaciones importantes.

(1) Notar que la desigualdad directa es $\underline{V} \leq \overline{V}$.

(2) Para probar la otra desigualdad se hará uso del siguiente teorema:

Teorema 5.2 (minimax de Sion). Sean (X, τ_X) e (Y, τ_Y) espacios topológicos y una función $f : X \times Y \rightarrow \mathbb{R} \cup \{+\infty\}$ tales que

(i) X e Y son conjuntos convexos y compactos.

(ii) $\forall y \in Y$, la aplicación $x \rightarrow f(x, y)$ es cuasi-convexa y semicontinua inferior.

(iii) $\forall x \in X$, la aplicación $y \rightarrow f(x, y)$ es cuasi-cóncava y semicontinua superior.

Entonces existe (\bar{x}, \bar{y}) tal que

$$\sup_{y \in Y} f(\bar{x}, y) = f(\bar{x}, \bar{y}) = \inf_{x \in X} f(x, \bar{y}). \quad (5.14)$$

El lector interesado en ver detalles acerca del teorema minimax¹ de Sion, así como las definiciones de los conceptos usados en este resultado, puede consultar las pags. 44 y 218 de [2]. Para una demostración, ver [74]. Para más bibliografía acerca de teoremas tipo minimax ver [73] y sus referencias.

¹Una buena referencia para teoremas tipo minimax es el sitio: <http://www.math.ucsb.edu/~simons/>.

Demostración (del Lema 5.3). Para la demostración se va a utilizar el teorema minimax de Sion con $X = \prod_{i=1}^m \xi_{Q_i}$ e $Y = \Phi_\alpha^m$. En lo que sigue, se verifican las hipótesis de este teorema. En efecto, $\forall i = 1, \dots, m$, ξ_{Q_i} es ζ_i -compacto, donde ζ_i es la topología débil en $L^1(\mathcal{X}_i, \mathcal{A}_i, R_i)$, y convexo, de donde se deduce inmediatamente que $\prod_{i=1}^m \xi_{Q_i}$ es ζ^m -compacto, por el teorema de Tychonoff con $\zeta^m = \prod_{i=1}^m \zeta_i$ la topología producto, y convexo. Gracias al lema 5.2 se sabe que Φ_α^m es τ^m -compacto. En lo que sigue se probará que, además, este conjunto es convexo. En efecto, sean $\varphi_1 = (\varphi_{11}, \dots, \varphi_{1m})$, $\varphi_2 = (\varphi_{21}, \dots, \varphi_{2m}) \in \Phi_\alpha^m$ y $\rho \in [0, 1]$. Claramente $\rho\varphi + (1 - \rho)\varphi_2 \in \Phi^m$, ya que $\forall i = 1, \dots, m$, $0 \leq \rho\varphi_{1i} + (1 - \rho)\varphi_{2i} \leq 1$. Además, $\forall P \in \mathcal{P}^m$ se tiene que

$$G(P, \rho\varphi_1 + (1 - \rho)\varphi_2) = \sum_{i=1}^m \delta_i E^{P_i}(\rho\varphi_{1i} + (1 - \rho)\varphi_{2i}) \quad (5.15)$$

$$= \rho G(P, \varphi_1) + (1 - \rho)G(P, \varphi_2), \quad (5.16)$$

y entonces

$$\sup_{P \in \mathcal{P}^m} G(P, \rho\varphi_1 + (1 - \rho)\varphi_2) \leq \rho \sup_{P \in \mathcal{P}^m} G(P, \varphi_1) + (1 - \rho) \sup_{P \in \mathcal{P}^m} G(P, \varphi_2) \quad (5.17)$$

$$\leq \rho\alpha + (1 - \rho)\alpha = \alpha. \quad (5.18)$$

Así, se concluye que Φ_α^m es un conjunto convexo, lo que prueba la hipótesis (i) del teorema 5.2. En la relación a la hipótesis (iii), se sabe que $\forall Q \in \mathcal{Q}^m$, $\Phi_\alpha^m \ni \varphi \rightarrow H(Q, \varphi)$ es τ^m -semicontinua superior, de manera que sólo falta probar que la aplicación anterior es cuasi-cóncava, lo que equivale a demostrar que el conjunto $S_1 := \{\varphi \in \Phi_\alpha^m : H(Q, \varphi) \geq m\}$ es convexo (ver pag. 44 de Aubin [2]). En efecto, sean $Q \in \mathcal{Q}^m$, $m \in \mathbb{R}$, $\varphi_1, \varphi_2 \in S_1$ y $\rho \in [0, 1]$. Entonces

$$H(Q, \rho\varphi_1 + (1 - \rho)\varphi_2) = \rho H(Q, \varphi_1) + (1 - \rho)H(Q, \varphi_2) \quad (5.19)$$

$$\geq \rho m + (1 - \rho)m = m, \quad (5.20)$$

lo que verifica el punto (iii) del teorema minimax de Sion. Finalmente, se prueba la hipótesis (ii). En efecto, sean $\varphi \in \Phi_\alpha^m$, $m \in \mathbb{R}$ y el conjunto $S_2 := \{Z_Q \in \prod_{i=1}^m \xi_{Q_i} : H(Q, \varphi) \leq m\}$. Dada la linealidad en Z_Q del funcional $H(Q, \varphi)$, es directo que el conjunto S_2 es convexo, lo que prueba la cuasi-convexidad de la aplicación $\prod_{i=1}^m \xi_{Q_i} \ni Z_Q \rightarrow H(Q, \varphi)$, $\forall \varphi \in \Phi_\alpha^m$. Por último, se prueba que la aplicación anterior es ζ^m -semicontinua inferior. En efecto, sea $\varphi \in \Phi_\alpha^m$. Claramente la aplicación $\xi_{Q_i} \ni Z_{Q_i} \rightarrow E^{R_i}(\varphi_i Z_{Q_i})$ es ζ_i -semicontinua inferior, para todo $i = 1, \dots, m$, y por lo tanto el funcional $\prod_{i=1}^m \xi_{Q_i} \ni Z_Q \rightarrow \sum_{i=1}^m \delta_i E^{R_i}(\varphi_i Z_{Q_i})$ es ζ^m -semicontinuo inferior (para detalles revisar la demostración del lema 5.2). Se cumplen así todas las hipótesis para aplicar el teorema minimax de Sion y concluir que $\underline{V} = \overline{V}$. \square

Para lo que sigue, es muy importante explicitar el siguiente cálculo:

$$\begin{aligned} E^{Q_i}(\varphi_i) &= E^{R_i}(\varphi_i Z_{Q_i}) = E^{R_i} \left(\varphi_i \left[Z_{Q_i} - \int_{\prod_{j=1}^m \xi_{P_j}} Z_{P_i} d\lambda \right] \right) + E^{R_i} \left(\varphi_i \int_{\prod_{j=1}^m \xi_{P_j}} Z_{P_i} d\lambda \right) \\ &\leq E^{R_i} \left(\left[Z_{Q_i} - \int_{\prod_{j=1}^m \xi_{P_j}} Z_{P_i} d\lambda \right]^+ \right) + \int_{\prod_{j=1}^m \xi_{P_j}} \left(\int_{\mathcal{X}_i} \varphi_i(x_i) Z_{P_i}(x_i) dR_i(x_i) \right) d\lambda, \end{aligned} \quad (5.21)$$

donde en el último término de la desigualdad anterior se usa el teorema de Fubini para poder intercambiar el orden de las integrales. Además, el término dentro del paréntesis corresponde a $E^{P_i}(\varphi_i)$, de manera que si se supone que esta esperanza está acotada por α_i , entonces se concluye que

$$E^{Q_i}(\varphi_i) \leq E^{R_i} \left(\left[Z_{Q_i} - \int_{\prod_{j=1}^m \xi_{P_j}} Z_{P_i} d\lambda \right]^+ \right) + \alpha_i \lambda \left(\prod_{j=1}^m \xi_{P_j} \right), \quad (5.22)$$

y por consiguiente,

$$H(Q, \varphi) \leq \sum_{i=1}^m \delta_i E^{R_i} \left(\left[Z_{Q_i} - \int_{\prod_{j=1}^m \xi_{P_j}} Z_{P_i} d\lambda \right]^+ \right) + \lambda \left(\prod_{j=1}^m \xi_{P_j} \right) \sum_{i=1}^m \delta_i \alpha_i, \quad (5.23)$$

donde el último término en la desigualdad anterior corresponde a $D(Q, \lambda)$, si es que se supone que $\sum_{i=1}^m \delta_i \alpha_i = \alpha$. Para $Q \in \mathcal{Q}^m$ fijo, se define el problema

$$(P) \quad \left\{ \mathfrak{p}(Q) = \sup_{\varphi \in \Phi_\alpha^m} H(Q, \varphi), \right.$$

denominado en lo que sigue problema primal. Notar que gracias a la τ^m -compacidad del conjunto Φ_α^m , (P) tiene solución. Por otro lado, considerar el siguiente problema:

$$(D) \quad \left\{ \mathfrak{d}(Q) = \inf_{\lambda \in \Lambda_+} D(Q, \lambda) = \inf_{\lambda \in \Lambda_+} \left[\sum_{i=1}^m \delta_i E^{R_i} \left(\left[Z_{Q_i} - \int_{\prod_{j=1}^m \xi_{P_j}} Z_{P_i} d\lambda \right]^+ \right) + \alpha \lambda \left(\prod_{j=1}^m \xi_{P_j} \right) \right]. \right.$$

Notar que de la desigualdad (5.23) es directo que $\mathfrak{p}(Q) \leq \mathfrak{d}(Q)$, $\forall Q \in \mathcal{Q}^m$.

Lema 5.4. (D) *corresponde al problema dual de Fenchel de (P) y $\forall Q \in \mathcal{Q}^m$, $\mathfrak{p}(Q) = \mathfrak{d}(Q)$. Más aún, $\forall Q \in \mathcal{Q}^m$, $\exists \tilde{\lambda} \in \Lambda_+$ que alcanza el ínfimo en (D) y $\exists \tilde{\varphi} \in \Phi_\alpha^m$ que alcanza el supremo en (P), que tiene la estructura dada en (5.11).*

La demostración del lema anterior, en el caso de test de hipótesis no múltiples, se encuentra en las pags. 1231-1234 de [55]. En este sentido, la demostración que se muestra a continuación es una extensión del argumento de Rudloff y Karatzas al contexto de hipótesis múltiples.

Demostración (del Lema 5.4). Para comenzar, se distinguen los siguientes espacios vectoriales sobre \mathbb{R} :

$$\mathcal{L} := \left\{ l : \prod_{i=1}^m \xi_{P_i} \rightarrow \mathbb{R} : l \text{ es un funcional continuo} \right\} \quad (5.24)$$

y

$$\mathcal{L}_+ := \{ l \in \mathcal{L} : l(Z_P) \geq 0, \forall P \in \mathcal{P}^m \}, \quad (5.25)$$

donde \mathcal{L} está dotado de las operaciones usuales de suma de funciones y ponderación por escalar, y además, satisface el siguiente orden parcial: $l_1 \leq l_2$ ssi $l_2 - l_1 \in \mathcal{L}_+$. Es conocido que si se dota a \mathcal{L} de la norma del supremo $\|l\|_{\mathcal{L}} = \sup_{P \in \mathcal{P}^m} |l(Z_P)|$, entonces $(\mathcal{L}, \|\cdot\|_{\mathcal{L}})$ es un espacio de Banach. Sea Λ , el espacio de las medidas con signo finitas y regulares sobre $(\prod_{i=1}^m \xi_{\mathcal{P}_i}, \mathfrak{B})$. Dada la ζ^m -compacidad de $\prod_{i=1}^m \xi_{\mathcal{P}_i}$ y que este espacio es metrizable (ver propiedad (c) en pag. 63 de [54]), entonces aplicando el corolario 14.15 de la pag. 498 de [1], se concluye que Λ se puede identificar con el dual topológico de $(\mathcal{L}, \|\cdot\|_{\mathcal{L}})$, vía la siguiente forma bilineal:

$$\langle l, \lambda \rangle := \int_{\prod_{i=1}^m \xi_{\mathcal{P}_i}} l d\lambda, \quad (5.26)$$

con $l \in \mathcal{L}$ y $\lambda \in \Lambda$. Se denota por $L^{\infty, m} := \prod_{i=1}^m L^{\infty}(\mathcal{X}_i, \mathcal{A}_i, R_i)$ y $\|\cdot\|_{L^{\infty, m}}$ alguna norma definida sobre él. Considerar el siguiente operador:

$$\begin{aligned} A : (L^{\infty, m}, \|\cdot\|_{L^{\infty, m}}) &\rightarrow (\mathcal{L}, \|\cdot\|_{\mathcal{L}}) \\ \varphi &\rightarrow (A\varphi)(Z_P) := -\sum_{i=1}^m \delta_i E^{R_i}(\varphi_i Z_{P_i}). \end{aligned}$$

Claramente este operador es lineal y acotado, y por lo tanto continuo. Sean $\vec{1}, \vec{0} \in \mathcal{L}$ tales que $\forall Z_P \in \prod_{i=1}^m \xi_{\mathcal{P}_i}$, $\vec{1}(Z_P) = 1 \in \mathbb{R}$ y $\vec{0}(Z_P) = 0 \in \mathbb{R}$. En términos de lo anterior se puede reescribir la restricción

$$\sup_{P \in \mathcal{P}^m} G(P, \varphi) \leq \alpha \quad (5.27)$$

de la siguiente manera:

$$\alpha \vec{1} + A\varphi \geq \vec{0} \quad \Leftrightarrow \quad A\varphi \in \mathcal{L}_+ - \alpha \vec{1}. \quad (5.28)$$

Con la notación anterior, para $Q \in \mathcal{Q}^m$, el problema primal (P) puede replantearse como

$$-\mathfrak{p}(-Q) = \inf_{\varphi \in L^{\infty, m}} \left[H(Q, \varphi) + \mathcal{I}_{\Phi^m}(\varphi) + \mathcal{I}_{\mathcal{L}_+ - \alpha \vec{1}}(A\varphi) \right] \quad (5.29)$$

$$= \inf_{\varphi \in L^{\infty, m}} [f(\varphi) + g(A\varphi)], \quad (5.30)$$

donde $-Q$ se interpreta como una medida finita con signo sobre $(\prod_{i=1}^m \mathcal{X}_i, \prod_{i=1}^m \mathcal{A}_i)$,

$$\mathcal{I}_C(\varphi) = \begin{cases} 0 & \text{si } \varphi \in C, \\ +\infty & \text{si } \varphi \notin C, \end{cases} \quad (5.31)$$

$f(\varphi) := H(Q, \varphi) + \mathcal{I}_{\Phi^m}(\varphi)$ y $g(A\varphi) := \mathcal{I}_{\mathcal{L}_+ - \alpha \vec{1}}(A\varphi)$. En lo que sigue se demuestra que el dual de Fenchel del problema primal, dado por (P), tiene la forma expresada en (D). Gracias a la proposición III.1.1, el teorema III.4.1 y la observación III.4.2 encontradas en las pags. 48 y 59-61 de [24], se cumple que el dual de Fenchel del problema (5.30) está dado por

$$-\mathfrak{d}(-Q) = \sup_{\lambda \in \Lambda} [-f^*(A^*\lambda) - g^*(-\lambda)], \quad (5.32)$$

donde $A^* : \Lambda \rightarrow (\mathcal{L}, \|\cdot\|_{\mathcal{L}})^*$ es el operador adjunto de A y g^*, f^* son las conjugadas de las funciones g, f respectivamente, cuyo cálculo se realiza a continuación:

$$g^*(\lambda) = \sup_{\tilde{l} \in \mathcal{L}} \left[\langle \tilde{l}, \lambda \rangle - \mathcal{I}_{\mathcal{L}_+ - \alpha \vec{1}}(\tilde{l}) \right] = \sup_{\tilde{l} \in \mathcal{L}_+ - \alpha \vec{1}} \langle \tilde{l}, \lambda \rangle = \sup_{l \in \mathcal{L}_+} \langle l - \alpha \vec{1}, \lambda \rangle \quad (5.33)$$

$$= \sup_{l \in \mathcal{L}_+} \langle l, \lambda \rangle - \alpha \int_{\prod_{i=1}^m \xi_{\mathcal{P}_i}} d\lambda = \mathcal{I}_{\mathcal{L}_+^*}(\lambda) - \alpha \lambda \left(\prod_{i=1}^m \xi_{\mathcal{P}_i} \right), \quad (5.34)$$

donde $\mathcal{L}_+^* := \{\lambda \in \Lambda : \langle l, \lambda \rangle \leq 0, \forall l \in \mathcal{L}_+\}$ y la última igualdad se tiene debido a que \mathcal{L}_+ es un cono que contiene al origen $\vec{0} \in \mathcal{L}$. Por otro lado:

$$f^*(A^*\lambda) = \sup_{\varphi \in L^{\infty, m}} [\langle A^*\lambda, \varphi \rangle - H(Q, \varphi) - \mathcal{I}_{\Phi^m}(\varphi)], \quad (5.35)$$

donde por definición A^* satisface la ecuación $\langle A^*\lambda, \varphi \rangle = \langle \lambda, A\varphi \rangle, \forall \varphi \in L^{\infty, m}, \forall \lambda \in \Lambda$. Así, de (5.26), de la definición de A y lo anterior resulta que

$$\langle A^*\lambda, \varphi \rangle = - \int_{\prod_{j=1}^m \xi_{\mathcal{P}_j}} \sum_{i=1}^m \delta_i E^{R_i}(Z_{P_i} \varphi_i) d\lambda, \quad (5.36)$$

y entonces

$$f^*(A^*\lambda) = \sup_{\varphi \in \Phi^m} \left[- \int_{\prod_{j=1}^m \xi_{\mathcal{P}_j}} \sum_{i=1}^m \delta_i E^{R_i}(Z_{P_i} \varphi_i) d\lambda - H(Q, \varphi) \right]. \quad (5.37)$$

Por lo tanto, el dual del problema (5.30) es

$$\begin{aligned} -\mathfrak{d}(-Q) &= \sup_{\lambda \in \Lambda} \left[- \sup_{\varphi \in \Phi^m} \left(- \int_{\prod_{j=1}^m \xi_{\mathcal{P}_j}} \sum_{i=1}^m \delta_i E^{R_i}(Z_{P_i} \varphi_i) d\lambda - H(Q, \varphi) \right) - \mathcal{I}_{-\mathcal{L}_+^*}(\lambda) - \alpha \lambda \left(\prod_{j=1}^m \xi_{\mathcal{P}_j} \right) \right] \\ &= \sup_{\lambda \in -\mathcal{L}_+^*} \left[- \sup_{\varphi \in \Phi^m} \left(- \int_{\prod_{j=1}^m \xi_{\mathcal{P}_j}} \sum_{i=1}^m \delta_i E^{R_i}(Z_{P_i} \varphi_i) d\lambda - H(Q, \varphi) \right) - \alpha \lambda \left(\prod_{j=1}^m \xi_{\mathcal{P}_j} \right) \right]. \end{aligned}$$

Notando que $-\mathcal{L}_+^* = \Lambda_+$, la expresión anterior toma la forma:

$$-\mathfrak{d}(-Q) = \sup_{\lambda \in \Lambda_+} \left[- \sup_{\varphi \in \Phi^m} \left(- \int_{\prod_{j=1}^m \xi_{\mathcal{P}_j}} \sum_{i=1}^m \delta_i E^{R_i}(Z_{P_i} \varphi_i) d\lambda - H(Q, \varphi) \right) - \alpha \lambda \left(\prod_{j=1}^m \xi_{\mathcal{P}_j} \right) \right]. \quad (5.38)$$

Notar que $\forall i = 1, \dots, m$, $(\mathcal{X}_i, \mathcal{A}_i, R_i)$ y $(\prod_{i=1}^m \xi_{\mathcal{P}_i}, \mathfrak{B}, \lambda)$ son espacios de medida finita y positiva, y la aplicación $\mathcal{X}_i \times \prod_{i=1}^m \xi_{\mathcal{P}_i} \ni (x_i, Z_P) \rightarrow Z_{P_i}(x_i) \varphi_i(x_i)$ es medible con respecto a la σ -álgebra producto $\mathcal{A}_i \otimes \mathfrak{B}$ (ver observación 5.1, parte (c)). Además, $\forall i = 1, \dots, m$ se tiene que

$$\int_{\prod_{j=1}^m \xi_{\mathcal{P}_j}} \int_{\mathcal{X}_i} |Z_{P_i} \varphi_i(x_i)| dR_i(x_i) d\lambda \leq \left(\sup_{P_i \in \mathcal{P}_i} \|Z_{P_i}\|_{L^1(\mathcal{X}_i, \mathcal{A}_i, R_i)} \right) \lambda \left(\prod_{i=1}^m \xi_{\mathcal{P}_i} \right) = \lambda \left(\prod_{i=1}^m \xi_{\mathcal{P}_i} \right) < \infty,$$

para todo $\lambda \in \Lambda_+$ y $\varphi \in \Phi^m$. Luego, por el teorema de Fubini-Tonelli se puede intercambiar el orden de las integrales, es decir,

$$\int_{\prod_{j=1}^m \xi_{\mathcal{P}_j}} \int_{\mathcal{X}_i} Z_{P_i}(x_i) \varphi_i(x_i) dR_i(x_i) d\lambda = \int_{\mathcal{X}_i} \int_{\prod_{j=1}^m \xi_{\mathcal{P}_j}} Z_{P_i}(x_i) \varphi_i(x_i) d\lambda dR_i(x_i). \quad (5.39)$$

Gracias a lo anterior,

$$-\mathfrak{d}(-Q) = \sup_{\lambda \in \Lambda_+} \left[- \sup_{\varphi \in \Phi^m} \sum_{i=1}^m \delta_i E^{R_i} \left(\varphi_i \left[-Z_{Q_i} - \int_{\prod_{j=1}^m \xi_{\mathcal{P}_j}} Z_{P_i} d\lambda \right] \right) - \alpha \lambda \left(\prod_{j=1}^m \xi_{\mathcal{P}_j} \right) \right]. \quad (5.40)$$

De (5.21) y dado que $\varphi \in \Phi^m$ es un test múltiple aleatorizado, el supremo sobre Φ^m en (5.40) es alcanzado por algún $\bar{\varphi} \in \Phi^m$ tal que $\forall i = 1, \dots, m$

$$\bar{\varphi}_i(x_i) = \begin{cases} 1 & \text{si } -Z_{Q_i}(x_i) > \int_{\prod_{j=1}^m \xi_{\mathcal{P}_j}} Z_{P_i}(x_i) d\lambda, \\ 0 & \text{si } -Z_{Q_i}(x_i) < \int_{\prod_{j=1}^m \xi_{\mathcal{P}_j}} Z_{P_i}(x_i) d\lambda, \end{cases} \quad R_i - c.s. \ x_i \in \mathcal{X}_i. \quad (5.41)$$

Dada cualquier medida finita con signo $\Pi_i = \Pi_i^+ - \Pi_i^-$ sobre $(\mathcal{X}_i, \mathcal{A}_i)$ tal que $\Pi_i^\pm \ll R_i$, la derivada de Radon-Nikodym correspondiente se escribe $Z_{\Pi_i} = Z_{\Pi_i^+} - Z_{\Pi_i^-}$. Definiendo

$$\gamma_{\lambda, \Pi_i} := Z_{\Pi_i} - \int_{\prod_{j=1}^m \xi_{\mathcal{P}_j}} Z_{P_i} d\lambda \in L^1(\mathcal{X}_i, \mathcal{A}_i, R_i) \quad (5.42)$$

y $\gamma_{\lambda, \Pi_i}^+$ (resp. $\gamma_{\lambda, \Pi_i}^-$) como la parte positiva (resp. como la parte negativa) de la variable aleatoria en (5.42), más (5.41), se tiene que el valor del problema (5.40) es

$$-\mathfrak{d}(-Q) = \sup_{\lambda \in \Lambda_+} \left[- \sum_{i=1}^m \delta_i E^{R_i}(\gamma_{\lambda, -Q_i}^+) - \alpha \lambda \left(\prod_{i=1}^m \xi_{\mathcal{P}_i} \right) \right] \quad (5.43)$$

y así

$$\mathfrak{d}(Q) = \inf_{\lambda \in \Lambda_+} \left[\sum_{i=1}^m \delta_i E^{R_i}(\gamma_{\lambda, Q_i}^+) + \alpha \lambda \left(\prod_{i=1}^m \xi_{\mathcal{P}_i} \right) \right]. \quad (5.44)$$

De esta representación y (5.42) se deduce que (D) corresponde al problema dual de Fenchel de (P). Por otro lado, la existencia de $\bar{\varphi} \in \Phi_\alpha^m$ que alcance el supremo en (P) se tiene debido a la compacidad débil-* de Φ_α^m y además, este tiene la estructura (5.11). Más aún, la propiedad que $\forall Q \in \mathcal{Q}^m$, $\mathfrak{p}(Q) = \mathfrak{d}(Q)$ sigue del teorema III.4.1 y la observación III.4.2 de [24], gracias a lo cual se verifica también la existencia de una solución del problema dual que alcance el ínfimo en (D), quedando con esto probado el lema 5.4. \square

Para concluir con la demostración del teorema 5.1, notar que el lema 5.4 garantiza que

$$\min_{Q \in \mathcal{Q}^m} \left(\max_{\varphi \in \Phi_\alpha^m} H(Q, \varphi) \right) = \min_{Q \in \mathcal{Q}^m} \mathfrak{p}(Q) = \min_{Q \in \mathcal{Q}^m} \mathfrak{d}(Q) = \min_{\substack{Q \in \mathcal{Q}^m \\ \lambda \in \Lambda_+}} \mathcal{D}(Q, \lambda). \quad (5.45)$$

Del lema 5.3 sigue que existe $\tilde{Q} \in \mathcal{Q}^m$ que alcanza el ínfimo en

$$\sup_{\varphi \in \Phi_\alpha^m} H(Q, \varphi).$$

Para este \tilde{Q} , el lema 5.4 muestra la existencia de un elemento $\tilde{\lambda} \in \Lambda_+$ que alcanza el ínfimo en (D). Así, existe un par $(\tilde{Q}, \tilde{\lambda})$ que alcanza el ínfimo en (5.9) y el lema 5.4 entrega los requerimientos estructurales del test óptimo, concluyendo con esto la demostración del teorema 5.1. ■

5.3 Limitaciones y alcances del Teorema

Una de las primeras observaciones que puede realizarse del teorema propuesto en este capítulo, es que la restricción (5.5) puede cambiarse a

$$\sum_{i=1}^m \delta_i = k, \quad (5.46)$$

con $k \in \mathbb{N}$ arbitrario. Lo anterior es bastante interesante, ya que si se supone que los parámetros $\alpha_1 = \dots = \alpha_m = cte$, entonces la condición $\sum_{i=1}^m \delta_i \alpha_i = \alpha$ permite concluir que $\alpha_i = \frac{\alpha}{k}$, recuperando así el control que se realiza a través del procedimiento de Bonferroni, que como ya se ha mencionado, es uno de los procedimientos clásicos que se pueden realizar sobre la FWER.

Un problema del resultado propuesto, y que le quita aplicabilidad, es que no hay una caracterización para la medida $\tilde{\lambda} \in \Lambda_+$, que como se mencionó anteriormente, tiene la interpretación de una distribución a priori sobre $\prod_{i=1}^m \xi_{\mathcal{P}_i}$. Al respecto, quedan algunas preguntas pendientes, como por ejemplo, si la medida óptima $\tilde{\lambda}$ podría tener una estructura de medida producto. Lo anterior es interesante porque permitiría analizar el problema múltiple de manera individual, esto es, hipótesis por hipótesis, obteniendo aún así, un procedimiento optimal global. No obstante, se cree a modo de conjetura que no hay razón para que $\tilde{\lambda}$ sea una medida producto. Con respecto a la caracterización de $\tilde{\lambda} \in \Lambda_+$, se podrían aplicar técnicas provenientes de la estadística bayesiana no paramétrica para definir una distribución a priori sobre $\prod_{i=1}^m \xi_{\mathcal{P}_i}$ (o sobre $\prod_{i=1}^m \mathcal{P}_i$), que tuviera un soporte grande y fuese no subjetiva, y simular a través de algoritmos de MCMC dicha distribución. Con lo anterior se lograría, en principio, construir los test de hipótesis dados por el teorema.

La definición del funcional $G(P, \varphi)$ puede ser muy simple a la hora de querer capturar posibles relaciones entre los test estadísticos, lo que podría llevar a cuestionar su uso. El

problema en este punto es que la demostración presentada en este capítulo usa fuertemente propiedades que surgen de la linealidad de los funcionales G y H , lo que dificulta la tarea de extender, por este lado, este resultado. Un intento por modificar los funcionales G y H fue el siguiente:

$$G(P, \varphi) = 1 - \prod_{i=1}^m (1 - E^{P_i}(\varphi_i)), \quad (5.47)$$

$$H(Q, \varphi) = 1 - \prod_{i=1}^m (1 - E^{Q_i}(\varphi_i)). \quad (5.48)$$

En este caso, se pudo probar los lemas 5.1 y 5.2, sin embargo, no se pudo demostrar el lema 5.3, debido a que en este caso el teorema minimax de Sion no se cumple. No se encontró en la literatura algún resultado de tipo minimax que se ajustara a este contexto, lo que no significa que no exista. En algún momento se pensó en usar funcionales G y H que no tuvieran la misma forma, sin embargo, esto no es posible, en el escenario de la demostración dada en este capítulo, debido a (5.40).

Un aspecto relevante del Teorema 5.1 es que este entrega una manera de enfrentar el problema de los test de hipótesis múltiples cuando tanto la hipótesis nula como alternativa son compuestas. Lo anterior se distingue de los resultados existentes en la literatura, en donde la situación típica es contrastar hipótesis nulas y alternativas simples, y en el caso más general, hipótesis nula simple y alternativa compuesta. Finalmente, se destaca la idea original de aplicar técnicas de análisis convexo en la resolución de test de hipótesis múltiples.

Conclusiones y Trabajo Futuro

Del presente trabajo se desprenden varias conclusiones que se detallan a continuación:

- (1) Uno de los principales objetivos de esta memoria fue realizar un estado del arte de la teoría de pruebas de múltiples hipótesis estadísticas. Lo anterior fue parcialmente realizado debido a que hay muchos artículos que sólo fueron mencionados y que contribuyen con ideas novedosas al tema de los THM. En este sentido este trabajo es, antes que todo, una introducción a la teoría de los THM.
- (2) A partir de 1995 ha habido literalmente una explosión de publicaciones centradas en definir nuevas tasas de error y procedimientos más potentes, sin embargo, se detectan la falta de textos de estudios donde se exponga la teoría de los THM moderna (posterior a 1995). Los libros clásicos, esto es, [48], [38], [84] y [39], se enfocan presentar procedimientos que controlan la FWER, los cuales se sabe no tienen muy buenas propiedades de potencia. Existen solamente dos textos que presentan la teoría moderna: [19], en el cual se enfoca en presentar métodos basados en remuestreo, y [22], en el cual se expone la teoría de los THM desde la perspectiva de los métodos bayesianos empíricos. Tras el estudio bibliográfico realizado durante el desarrollo de esta memoria, se recomienda al lector interesado en iniciarse en la temática de los THM comenzar leyendo [19] y continuar con [22].
- (3) Una vez terminado este trabajo de memoria se puede afirmar con toda propiedad, que el tema de los THM está a la vanguardia de la investigación estadística actual, lo cual queda demostrado por los diversos y activos grupos de investigación desarrollados en importantes universidades en el mundo como Berkeley y Stanford, entre otras, además de la reiterada aparición de artículos de esta área en revistas de gran impacto como “The Annals of Statistics”.
- (4) Un objetivo secundario de esta memoria era desarrollar algún aporte de tipo teórico, tarea que se da por cumplida. Si bien es cierto que el resultado propuesto dista de ser del todo aplicable, es un aporte en el sentido que una de las debilidades detectadas en la teoría de los THM fue la ausencia de resultados de optimalidad como si ocurre en la estadística clásica no múltiple. En esta dirección, una conclusión final de este trabajo es que el tema de los THM está todavía en una etapa de maduración de ideas

y que todavía hay mucho por hacer en cuanto a darle más sustentos matemáticos a los fundamentos de esta teoría.

En relación a trabajos futuros se destacan los siguientes:

- (i) Sobre el teorema propuesto en el Capítulo 5, queda abierta la interrogante de si es posible encontrar un funcional que permita controlar la k -FWER, la FDR o algunas de sus variantes. Lo anterior no es para nada trivial, debido a que en general no existen fórmulas explícitas para calcular la FDR. Otra situación que queda pendiente es intentar caracterizar la medida $\tilde{\lambda} \in \Lambda_+$. Desde un punto de vista técnico y pensando en la demostración planteada en la sección 5.2, se destaca la necesidad de tener que trabajar con métodos no lineales, si se quiere definir funcionales (G y H) que sean más apropiados a las tasas de error ya mencionadas. Por otro lado, queda propuesto como trabajo futuro también, analizar la incorporación de técnicas provenientes de la estadística bayesiana no paramétrica para encontrar, vía simulación a través de algoritmos de MCMC, la medida $\tilde{\lambda} \in \Lambda_+$.
- (ii) Uno de los temas aún no resueltos satisfactoriamente en la teoría presentada de los THM, es qué hacer en caso que exista dependencia de los test estadísticos. Como ya se manifestó en el capítulo 2, este es un tema de mucho interés debido a las repercusiones negativas que tiene la dependencia en los THM, en cuanto a la obtención de procedimientos extremadamente conservadores. El autor de este trabajo cree que para resolver este tipo de situaciones se hace necesaria la utilización de metodologías matemáticas más avanzadas. Algunas propuestas que quedan para una investigación futura son: los campos aleatorios markovianos, las cadenas de Markov ocultas y las cópulas. Por otro lado, el trabajo de Hall y Wang (ver [37]) da luces de que bajo ciertas condiciones entre el tamaño muestral n y el número de hipótesis m , esto es, que n diverja más rápido que $\ln(m)$, tal vez la hipótesis de independencia de los test estadísticos no es tan grave. Queda pendiente analizar dicho artículo y buscar generalizaciones.

Bibliografía

- [1] ALIPRANTIS, D. C., AND BORDER, K. C. *Infinite dimensional analysis: a hitchhiker's guide*, third ed. Springer-Verlag, 2006.
- [2] AUBIN, J.-P. *Mathematical methods of game and economic theory*. North-Holland Publishing Company, 1979.
- [3] BENJAMINI, Y., AND HELLER, R. False discovery rates for spatial signals. *J. Amer. Statist. Assoc.* *102*, 480 (2007), 1272–1281.
- [4] BENJAMINI, Y., AND HOCHBERG, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. Roy. Statist. Soc. Ser. B* *57* (1995), 289–300.
- [5] BENJAMINI, Y., AND HOCHBERG, Y. On the adaptive control of the false discovery rate in multiple testing with independent statistics. *J. Behav. Educ. Statist.* *25* (2000), 60–83.
- [6] BENJAMINI, Y., HOCHBERG, Y., AND KLING, Y. False discovery rate control in multiple hypotheses testing using dependent test statistics. *Research Paper 97-1 Dept. Statistics and O.R.*, Tel Aviv Univ (1997).
- [7] BENJAMINI, Y., KRIEGER, A. M., AND YEKUTIELI, D. Adaptive linear step-up false discovery rate controlling procedures. *Biometrika* *93* (2006), 491–507.
- [8] BENJAMINI, Y., AND YEKUTIELI, D. The control of the false discovery rate in multiple testing under dependency. *Annals of Statistics* *29* (2001), 1165–1188.
- [9] BERGER, J., AND PERICCHI, L. Model selection. *Institute of Mathematical Statistics, Lecture Notes-Monograph Series* *38* (2001), 135–207.
- [10] BERGER, J., PERICCHI, L., AND VARSHAVSKY, J. Bayes factors and marginal distributions in invariant situations. *Sankhya* *60* (1998), 307–321.
- [11] BERGER, J. O. *Statistical decision theory and bayesian analysis*, second ed. Springer-Verlag, 1985.

- [12] BERNARDO, J. M., AND SMITH, A. F. M. *Bayesian theory*. John Wiley & Sons, 1994.
- [13] BLACK, M. A. A note on adaptive control of false discovery rates. *J. R. Stat. Soc. Ser. B* 66 (2004), 297–304.
- [14] BLACKWELL, D., AND MACQUEEN, J. B. Ferguson distribution via Polya urns schemes. *Annals of Statistics* 1 (1973), 353–355.
- [15] BRÉZIS, H. *Análisis funcional: teoría y aplicaciones*, versión española de juan ramón esteban ed. Alianza Editorial, S.A., 1984.
- [16] CASELLA, G., AND BERGER, R. L. *Statistical inference*, second ed. Duxbury Thomson Learning, 2002.
- [17] CHEN, J., AND SARKAR, S. K. Multiple testing of response rates with a control: a bayesian stepwise approach. *J. Statist. Plann. Inf.* 125 (2004), 3–16.
- [18] CVITANIC, J., AND KARATZAS, I. Generalized Neyman-Pearson lemma via convex duality. *Bernoulli*, 7 (2001), 79–97.
- [19] DUDOIT, S., AND VAN DER LAAN, M. J. *Multiple testing procedures with applications to genomics*. Springer Science+Business Media, New York, 2008.
- [20] EFRON, B. Local false discovery rates. Available at <http://www.stat.stanford.edu/brad/papers/False.pdf> (2005).
- [21] EFRON, B. The future of statistics. Available at <http://stat.stanford.edu/~ckirby/brad/other/2009Future.pdf> (2009).
- [22] EFRON, B. *Large-scale inference: empirical Bayes methods for estimation, testing, and prediction*. Cambridge University Press, 2010.
- [23] EFRON, B., TIBSHIRANI, R., STOREY, J. D., AND TUSHER, V. Empirical Bayes analysis of a microarray experiment. *J. Amer. Statist. Assoc.* 96 (2001), 1151–1160.
- [24] EKELAND, I., AND TÉMAM, R. *Convex analysis and variational problems*. Society for industrial and applied mathematics, 1999.
- [25] ESCOBAR, M., AND WEST, M. Bayesian density estimation and inference using mixtures. *Journal of American Statistical Association* 90 (1995), 577–588.
- [26] FERGUSON, T. S. A bayesian analysis of some nonparametric problems. *Annals of Statistics* 1 (1973), 209–230.
- [27] FINNER, H. *Testing multiple hypotheses: general theory, specific problems, and relationships to other multiple decision procedures*. PhD thesis, Habilitationsschrift, FB IV Mathematik, Univ. Trier, 1994.

- [28] FINNER, H., AND ROTERS, M. Multiple hypothesis testing and expected number of type I errors. *Annals of Statistics* 30 (2002), 220–238.
- [29] FINNER, H., AND STRASSBURGER, K. The partitioning principle: a powerful tool in multiple decision theory. *The Annals of Statistics* 30, 4 (2002), 1194–1213.
- [30] FLORENS, J.-P., MOUCHART, M., AND ROLIN, J.-M. *Elements of bayesian statistics*. Chapman and Hall, 1990.
- [31] GAVRILOV, Y., BENJAMINI, Y., AND SARKAR, S. K. An adaptive step-down procedure with proven FDR control under independence. *Annals of Statistics* 37, 2 (2009), 619–629.
- [32] GENOVESE, C., AND WASSERMAN, L. Operating characteristics and extensions of the false discovery rate procedure. *J. R. Stat. Soc. Ser. B* 64 (2002), 499–517.
- [33] GENOVESE, C. R., ROEDER, K., AND WASSERMAN, L. False discovery control with p-value weighting. *Biometrika* 93 (2006), 509–524.
- [34] GENOVESE, C. R., AND WASSERMAN, L. A stochastic process approach to false discovery rate. *Annals of Statistics* 32 (2004), 1035–1063.
- [35] GHOSAL, S., ROY, A., AND TANG, Y. Posterior consistency of Dirichlet mixtures of beta densities in estimating positive false discovery rates. In *Beyond Parametrics in Interdisciplinary Research: Festschrift in Honor of Professor Pranab K. Sen (E. Pena et al., eds.)*, Institute of Mathematical Statistics Collection 1 (2008), 105–115.
- [36] GHOSH, J. K., AND RAMAMOORTHI, R. V. *Bayesian nonparametrics*. Springer-Verlag, 2003.
- [37] HALL, P., AND WANG, Q. Strong approximations of level exceedences related to multiple hypothesis testing. *Bernoulli* 16, 2 (2010), 418–434.
- [38] HOCHBERG, Y., AND TAMHANE, A. C. *Multiple comparison procedures*. John Wiley & Sons, 1987.
- [39] HSU, J. *Multiple comparison: theory and methods*. Chapman and Hall, 1996.
- [40] JEFFREYS, H. *Theory of probability*. Oxford University Press, 1961.
- [41] LEHMANN, E. L. A theory of some multiple decision problems, I. *Ann. Math. Statist.* 28 (1957), 1–25.
- [42] LEHMANN, E. L. A theory of some multiple decision problems, II. *Ann. Math. Statist.* 28 (1957), 547–572.
- [43] LEHMANN, E. L., AND ROMANO, J. Generalizations of the familywise error rate. *Annals of Statistics* 33 (2005), 1138–1154.

- [44] LEHMANN, E. L., AND ROMANO, J. P. *Testing statistical hypotheses*, third ed. Springer Science+Business Media, 2005.
- [45] LEWIS, C., AND THAYER, D. T. Bayesian decision theory for multiple comparisons. *IMS Lecture Notes-Monograph Series 57* (2009), 326–332.
- [46] MACEACHERN, S. N., AND MULLER, P. Estimating mixture of Dirichlet process models. *Journal of Computational and Graphical Statistics 7* (1998), 295–315.
- [47] MILLER, C. J., GENOVESE, C., NICHOL, R. C., WASSERMAN, L., CONNOLLY, A., REICHART, D., HOPKINS, A., SCHNEIDER, J., AND MOORE, A. Controlling the false discovery rate in astrophysical data analysis. *The Astronomical Journal 122*, 6 (2001), 3492–3505.
- [48] MILLER, R. G. *Simultaneous statistical inference*, second ed. Springer-Verlag, New York, 1981.
- [49] MIWA, T., AND HAYTER, A. J. Combining the advantages of one-sided and two-sided test procedures for comparing several treatment effects. *J. Amer. Statist. Assoc. 94* (1999), 302–307.
- [50] PANTAZIS, D., NICHOLS, T. E., BAILLET, S., AND LEAHY, R. M. A comparison of random field theory and permutation methods for statistical analysis of meg data. *NeuroImage 25* (2005), 383–394.
- [51] PARKER, R. A., AND ROTHENBERG, R. B. Identifying important results from multiple statistical tests. *Statistics in Medicine 7* (1988), 1031–1043.
- [52] ROBERT, C. *The bayesian choice: from decision-theoretic foundations to computational implementation*, second ed. Springer Science+Business Media, 2007.
- [53] ROY, A., AND GHOSAL, S. Estimating false discovery rate under dependence: a mixture model approach. *Preprint* (2008).
- [54] RUDIN, W. *Functional analysis*, second ed. McGraw-Hill, Inc., 1991.
- [55] RUDLOFF, B., AND KARATZAS, I. Testing composite hypotheses via convex duality. *Bernoulli 16*, 4 (2010), 1224–1239.
- [56] SARKAR, S. Some probability inequalities for ordered MTP2 random variables: a proof of the simes conjecture. *Annals of Statistics 26* (1998), 494–504.
- [57] SARKAR, S., AND CHANG, C. The Simes method for multiple hypothesis testing with positively dependent test statistics. *J. Amer. Statist. Assoc. 92* (1997), 1601–1608.
- [58] SARKAR, S. K. Some results on false discovery rate in stepwise multiple testing procedures. *Annals of Statistics 30* (2002), 239–257.

- [59] SARKAR, S. K. False discovery and false non-discovery rates in single-step multiple testing procedure. *Annals of Statistics* 34 (2006), 394–415.
- [60] SARKAR, S. K. Step-up procedures controlling generalized FWER and generalized FDR. *Annals of Statistics* 35 (2007), 2405–2420.
- [61] SARKAR, S. K. Generalizing Simes’ test and Hochberg’s step-up procedure. *Annals of Statistics* 36 (2008), 337–363.
- [62] SARKAR, S. K. Two-stage step-up procedures controlling FDR. *J. Statist. Plann. Inf.* 138 (2008), 1072–1084.
- [63] SARKAR, S. K., AND GUO, W. On a generalized false discovery rate. *Annals of Statistics* 37, 3 (2009), 1545–1565.
- [64] SARKAR, S. K., AND GUO, W. Procedures controlling the k-FDR using bivariate distributions of the null p-values. *Statist. Sinica* 20, 3 (2010), 1227–1238.
- [65] SARKAR, T. K. Some lower bounds of reliability. *Technical Report, 124, Dept. Operation Research and Statistics, Stanford Univ* (1969).
- [66] SAVILLE, D. J. Multiple comparison procedures: the practical solution. *American Statistician* 44 (1990), 174–180.
- [67] SCHERVISH, M. J. *Theory of statistics*. Springer-Verlag, New York, 1995.
- [68] SCHWEDER, T., AND SPJOTVOLL, E. Plots of p-values to evaluate many test simultaneously. *Biometrika* 69 (1982), 493–502.
- [69] SCOTT, J. G., AND BERGER, J. O. An exploration of aspects of bayesian multiple testing. *Journal of Statistical Planning and Inference* 136 (2006), 2144–2162.
- [70] SHAO, J. *Mathematical statistics*, second ed. Springer Science+Business Media, 2003.
- [71] SILVEY, S. *Statistical inference*. Chapman and Hall, 1975.
- [72] SIMES, R. J. An improved Bonferroni procedure for multiple tests of significance. *Biometrika* 73 (1986), 751–754.
- [73] SIMONS, S. *Minimax and monotonicity*. Springer-Verlag, 1998.
- [74] SION, M. On general minimax theorems. *Pacific Journal of Mathematics* 8, 1 (1958), 171–176.
- [75] SONNEMANN, E. General solutions to multiple testing problems. *Biometrical* 50, 5 (2008), 641–656.
- [76] STOREY, J. D. A direct approach to false discovery rates. *J. Roy. Statist. Soc. Ser. B* 64 (2002), 479–498.

-
- [77] STOREY, J. D. The positive false discovery rate: a bayesian interpretation and the q -value. *Annals of Statistics* 31 (2003), 2013–2035.
- [78] STOREY, J. D. The optimal discovery procedure: a new approach to simultaneous significance testing. *J. Roy. Statist. Soc. Ser. B* 69, 3 (2007), 347–368.
- [79] STOREY, J. D., TAYLOR, J. E., AND SIEGMUND, D. Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: a unified approach. *J. Roy. Statist. Soc. Ser. B* 66 (2004), 187–205.
- [80] SUN, W., AND CAI, T. T. Large-scale multiple testing under dependence. *J. R. Stat. Soc. Ser. B* 71, 2 (2009), 393–424.
- [81] TANG, Y., GHOSAL, S., AND ROY, A. Nonparametric bayesian estimation of positive false discovery rates. *Biometrics* 63 (2007), 1126–1134.
- [82] VAN DE WIEL, M. A., AND VAN WIERINGEN, W. N. CGHregions: dimension reduction for array CGH data with minimal information loss. *Cancer Inform* 3 (2007), 55–63.
- [83] WASSERMAN, L. *All of nonparametric statistics*. Springer Science+Business Media, 2006.
- [84] WESTFALL, P. H., AND YOUNG, S. S. *Resampling-based multiple testing, examples and methods for p -value adjustment*. John Wiley & Sons, 1993.
- [85] WITTING, H. *Mathematische statistik I*. B. G. Teubner, Stuttgart, 1985.