

UNIVERSIDAD COMPLUTENSE DE MADRID

FACULTAD DE CIENCIAS MATEMÁTICAS

Departamento de Estadística e Investigación Operativa I



**APLICACIONES BAYESIANAS A PROBLEMAS NO
PARAMÉTRICOS**

MEMORIA PARA OPTAR AL GRADO DE DOCTOR

PRESENTADA POR

Beatriz González Pérez

Bajo la dirección del doctor

Miguel Ángel Gómez Villegas

Madrid, 2005

ISBN: 84-669-2804-9

APLICACIONES BAYESIANAS
A
PROBLEMAS NO PARAMÉTRICOS

Beatriz González Pérez

Madrid, enero de 2005

Departamento de Estadística e Investigación Operativa I

Facultad de CC. Matemáticas

UNIVERSIDAD COMPLUTENSE DE MADRID

UCM

Aplicaciones Bayesianas a Problemas No Paramétricos

Tesis Doctoral de Beatriz González Pérez

realizada bajo la dirección de D. Miguel Ángel Gómez-Villegas¹

Madrid, enero de 2005

Departamento de Estadística e Investigación Operativa I,

Facultad de CC. Matemáticas

UNIVERSIDAD COMPLUTENSE DE MADRID

¹Catedrático del Departamento de Estadística e Investigación Operativa I de la Universidad Complutense de Madrid

A José.

Índice

Prólogo	III
1. Distribuciones ε-Contaminadas en Tablas de Contingencia	1
1.1. Introducción	1
1.2. Formulación del Problema	3
1.3. Notación	6
1.4. Acotaciones de la Probabilidad A Posteriori	7
1.5. Ejemplos	10
1.6. Comparación con el Método Clásico	11
1.7. Comentarios	15
2. Análisis Bayesiano de Tablas de Contingencia	17
2.1. Introducción	17
2.2. Formulación del Problema y Cálculo de la Probabilidad A Posteriori	20
2.3. Comparación con el Método Clásico	22
2.4. Tablas $r \times s$ con \mathbf{p}_0 Conocido	29
2.5. Conclusiones	32
2.6. Tablas $r \times s$ con \mathbf{p}_0 Desconocido	33
2.6.1. Primer Método	34
2.6.2. Segundo Método	35
2.6.3. Tercer Método	36

2.6.4.	Comparación con el Método Clásico	36
2.7.	Tablas $r \times s$ con $\mathbf{p}_0 = \mathbf{p}(\omega)$	44
2.7.1.	Primer Método	45
2.7.2.	Segundo Método	46
2.7.3.	Tercer Método	47
2.7.4.	Comparación con el Método Clásico	47
2.8.	Observaciones	57
2.9.	Conclusiones y Comentarios	58
3.	Acuerdo entre la Aproximación Clásica y Bayesiana en	
	Tablas de Contingencia	60
3.1.	Introducción	60
3.2.	Preliminares	62
3.3.	Acuerdo entre la Aproximación Clásica y Bayesiana	66
3.4.	Comparación con el Método Clásico Usual	69
3.5.	Reparametrizaciones	71
3.6.	Comentarios	74
4.	Acuerdo entre la Aproximación Clásica y Bayesiana en	
	el Contraste de la Hipótesis Nula Puntual Multivariante	76
4.1.	Introducción	76
4.2.	Acuerdo entre la Aproximación Clásica y Bayesiana	80
4.3.	Aplicaciones	83
4.3.1.	Paradoja de Lindley	84
4.3.2.	Cotas Inferiores para Distribuciones Unimodales y Simétricas	88
4.3.3.	Cotas Inferiores para Mixturas de Normales con Parámetro de Escala	90
4.4.	Observaciones	93
4.5.	Conclusiones y Comentarios	93
	Referencias	95

Prólogo

La relación entre la respuesta clásica y bayesiana a un mismo problema de Inferencia Estadística ha sido objeto de estudio en numerosas ocasiones. En muchas situaciones se ha conseguido un acuerdo numérico entre ambas aproximaciones, aunque, en algunos casos, la discrepancia entre ambas posturas se mantiene. Se dice numérico porque la interpretación de los resultados es completamente distinta.

En el ámbito de los contrastes de hipótesis paramétricos, cuando el contraste es unilateral es posible reconciliar la evidencia clásica, expresada en términos del p-valor, con la bayesiana, expresada en términos de la probabilidad final de la hipótesis nula. Concretamente, Casella y Berger (1987) prueban que, utilizando diferentes clases de distribuciones iniciales, el ínfimo de la probabilidad final es igual, e incluso, en determinados casos, menor que el correspondiente p-valor, concluyendo que para contrastes unilaterales el p-valor se puede considerar como una aproximación a las medidas de evidencia bayesianas. Sin embargo, en los contrastes paramétricos de hipótesis nula puntual la discrepancia entre ambos métodos es bien conocida. Lindley (1957) presentó la discrepancia en el supuesto de que la verosimilitud fuese normal. En general hay coincidencia en señalar que, en este caso, el p-valor tiende a exagerar la evidencia en contra de la hipótesis nula. En este contexto, Berger y Sellke (1987) calculan el ínfimo de las probabilidades finales de la hipótesis nula puntual sobre diferentes clases de distribuciones a priori y obtienen que el ínfimo es sustancialmente mayor que el correspondiente p-valor, concluyendo que los p-valores puede ser medidas muy engañosas de la evidencia en contra de la hipótesis nula. Es necesario apuntar que la masa asignada al punto de la hipótesis nula es 0.5.

Gómez-Villegas y Gómez Sánchez-Manzano (1992) estudian la relación entre los contrastes de hipótesis nula puntual y de intervalo, poniendo de manifiesto que, cuando el intervalo es suficientemente pequeño, dan lugar a la misma decisión. Concretamente, en el problema del contraste de hipótesis nula puntual univariante $H_0 : \theta = \theta_0$, versus $H_1 : \theta \neq \theta_0$, con una densidad a priori dada para el parámetro θ , $\pi(\theta)$, introducen un procedimiento, que posteriormente es justificado por Gómez-Villegas y Sanz (1998), para determinar una distribución a priori de tipo mixto. La metodología consiste en fijar un intervalo de amplitud 2ϵ alrededor de θ_0 y asignar una masa a priori π_0 a H_0 , calculada integrando $\pi(\theta)$ sobre el intervalo $(\theta_0 - \epsilon, \theta_0 + \epsilon)$, repartiendo la probabilidad restante, $1 - \pi_0$, sobre H_1 mediante $\pi(\theta)$. Además, Gómez-Villegas y Sanz (1998) muestran que el ínfimo de la probabilidad a posteriori de H_0 puede estar próximo al p-valor en la clase de distribuciones a priori unimodales y simétricas.

Gómez-Villegas y Sanz (2000) calculan una cota inferior para la probabilidad a posteriori del punto de la hipótesis nula, cuando la distribución a priori pertenece a la clase de distribuciones ϵ -contaminadas, y muestran algunos ejemplos en los que dicha cota se puede hacer numéricamente próxima al p-valor.

En la mayoría de las contribuciones existentes se utiliza una clase de distribuciones a priori, sin embargo Gómez-Villegas, Maín y Sanz (2002) estudian el problema cuando se utiliza una única distribución a priori y muestran que para la *paradoja de Lindley* y el ejemplo de Darwin-Fisher, este procedimiento hace que las diferencias entre la probabilidad a posteriori de la hipótesis nula puntual y el p-valor no sean tan acusadas como cuando se utiliza $\pi_0 = 0,5$.

Recientemente, Gómez-Villegas, Maín y Sanz (2004) han desarrollado un test Bayesiano para el contraste de hipótesis nula simple frente a alternativa bilateral en el caso multivariante. Sugieren un procedimiento para construir una distribución a priori de tipo mixto y calculan cotas inferiores de las probabilidades a posteriori de la hipótesis nula puntual sobre algunas clases de distribuciones a priori, obteniendo una mejor aproximación debido a que el correspondiente p-valor se encuentra en el rango de las medidas de evidencia

bayesianas.

De la Horra (2005) considera el problema del contraste de la hipótesis nula puntual para un parámetro de localización y reconcilia los p-valores clásicos y predictivos a priori.

En esta memoria el problema del contraste de la hipótesis nula puntual es considerado en un contexto no paramétrico. Este problema ha dado lugar a una extensa literatura, desde que Karl Pearson introdujera su ya clásico test de la χ^2 para valorar la bondad del ajuste (Pearson, 1900). El objeto de interés es la relación entre el p-valor y las medidas bayesianas de evidencia en contra de la hipótesis nula de homogeneidad de r poblaciones multinomiales independientes. Concretamente, se utiliza la probabilidad a posteriori de que la hipótesis nula sea cierta. Existen muchas aproximaciones bayesianas al problema de homogeneidad, sin embargo la relación entre la respuesta clásica y bayesiana no ha sido muy estudiada en este contexto.

En el caso de tablas de contingencia 2×2 , Howard (1998) aboga por el uso más frecuente de contrastes unilaterales y aborda el problema desde un punto de vista bayesiano, proporcionando una medida cuantitativa de la fuerza de la evidencia en apoyo de la hipótesis más probable. Además, observa que el contraste clásico de la χ^2 puede considerarse como una aproximación de la probabilidad a posteriori de la hipótesis nula unilateral, para una distribución a priori particular.

Nuestra aproximación general al problema de homogeneidad en tablas de contingencia $r \times s$, consiste en trabajar directamente con la hipótesis nula simple calculando su probabilidad a posteriori. Para ello, se sigue el método utilizado por Gómez-Villegas, Maín y Sanz (2004). Con este procedimiento, en el contexto de la hipótesis nula puntual, es posible precisar cómo y cuándo es posible una reconciliación entre el p-valor y la probabilidad a posteriori de la hipótesis nula de homogeneidad.

En el capítulo 1 se desarrolla un procedimiento bayesiano para el contraste de homogeneidad de distribuciones multinomiales independientes en tablas de contingencia. Se calcula una cota inferior de la probabilidad a posteriori, cuando la distribución a priori de tipo mixto pertenece a la clase de distribuciones ε -contaminadas, y se compara con el

p-valor. En el capítulo 2, mediante el procedimiento bayesiano introducido en el capítulo 1 para el contraste de hipótesis nula simple frente a alternativa bilateral en tablas de contingencia, dadas muestras independientes de dos distribuciones binomiales y tomando una única distribución a priori de tipo mixto, se calcula la probabilidad a posteriori de que la proporción de éxitos en la primera población sea igual que en la segunda. Dicha probabilidad a posteriori se compara con el p-valor del procedimiento clásico usual, obteniéndose una reconciliación entre ambos resultados, en los términos de una condición suficiente. Los resultados obtenidos se generalizan para tablas $r \times s$, cuando el punto de la hipótesis nula, \mathbf{p}_0 , es conocido, desconocido o de forma funcional conocida, $\mathbf{p}_0 = \mathbf{p}(\omega)$. En el capítulo 3, para el análisis bayesiano introducido en el capítulo 1, se formalizan los resultados obtenidos en el capítulo 2 y se demuestra un teorema que pone de manifiesto los términos en los que es posible un acuerdo entre ambos métodos. En el capítulo 4, para el problema general del contraste de hipótesis nula puntual multivariante frente a alternativa bilateral, se desarrolla un procedimiento que permite concretar cómo y cuándo es posible un acuerdo entre los procedimientos clásicos y bayesianos.

Quiero terminar esta introducción con mi más sincero agradecimiento a Miguel Ángel Gómez-Villegas, maestro y persona excepcional, por su apoyo y estímulo.

Capítulo 1

Distribuciones ε –Contaminadas en Tablas de Contingencia

Para llevar a cabo un análisis bayesiano sobre un parámetro desconocido, θ , se necesita modelizar la opinión a priori que se tiene sobre θ mediante una distribución de probabilidad. Se desarrolla un procedimiento bayesiano para el contraste de homogeneidad de distribuciones multinomiales independientes en tablas de contingencia. Se calcula una cota inferior de la probabilidad a posteriori, cuando la distribución a priori pertenece a la clase de distribuciones ε –contaminadas, y se compara dicha cota con el p -valor.

1.1. Introducción

Se supone que son extraídas muestras aleatorias independientes de dos poblaciones suficientemente grandes donde cada una de sus unidades se clasifica como éxito o fracaso. Una muestra es de tamaño n_1 y produce a éxitos y b fracasos y la otra es de tamaño n_2 y produce c éxitos y d fracasos. Los datos se visualizan en la Tabla 1.1.

Se precisa una medida cuantitativa de hasta qué punto los datos apoyan o no la hipótesis de que la proporción de éxitos en la primera población, p_1 , es igual a la proporción de éxitos en la segunda población, p_2 . Por lo tanto, el parámetro de interés para el problema

de homogeneidad planteado es $\theta = (p_1, p_2)$.

Tabla 1.1: datos en la tabla 2×2 .

	Éxitos	Fracasos	Total
Muestra 1	a	b	n_1
Muestra 2	c	d	n_2
Total	m_1	m_2	N

Para realizar un análisis bayesiano sobre un parámetro desconocido, θ , es necesario modelizar la opinión a priori que se tiene sobre θ mediante una distribución de probabilidad. Sin embargo, no es usual el caso en el que la información inicial pueda ser expresada en términos de una distribución de probabilidad concreta, debido a que, con frecuencia, esta información a priori es difusa. Generalmente, debido a esta ausencia de precisión, la información a priori se expresa en términos de una clase de distribuciones, Γ , que contenga a todas las distribuciones a priori posibles para θ . Además, para comparar la probabilidad a posteriori de la hipótesis nula del procedimiento bayesiano con el p-valor del método clásico parece razonable considerar una clase de distribuciones a priori en lugar de una distribución a priori concreta, puesto que el p-valor no utiliza información a priori.

Una forma interesante de describir opiniones a priori consiste en considerar la clase de distribuciones ε -contaminadas dada por

$$\Gamma = \{\pi = (1 - \varepsilon) q_0 + \varepsilon q, q \in Q\}, \quad (1.1)$$

donde q_0 es una distribución a priori particular, la que se utilizaría en un análisis Bayesiano con una sólo distribución a priori. Q es una clase de distribuciones de probabilidad que representan las posibles (y razonables) desviaciones de q_0 . Un ε fijo, con $0 < \varepsilon < 1$, representa el grado de contaminación que se quiere introducir en q_0 .

En lo referente a la clase \mathcal{Q} se pueden tener en cuenta varias posibilidades. Vamos a utilizar la clase de todas las distribuciones de probabilidad. Gómez-Villegas y Sanz (2000) utilizan esta clase de distribuciones para estudiar el problema de contraste de hipótesis nula simple desde una perspectiva bayesiana. Huber (1973) y Sivaganesan (1988) también la utilizan aunque en otro contexto. Berger y Berlinger (1986), Berger (1985, 1994), y Sivaganesan y Berger (1989) proporcionan información relevante sobre otras clases de contaminación.

En la sección 1.2 se formula de forma precisa el problema de contraste de hipótesis nula simple en tablas de contingencia y se justifica la utilización de una distribución a priori de tipo mixto. En la sección 1.3 se introduce la notación. En la sección 1.4 se calcula el ínfimo de la probabilidad a posteriori de que la proporción de éxitos en la primera población sea igual que en la segunda e igual a un valor común p_0 conocido, cuando la a priori pertenece a la clase de distribuciones ε -contaminadas. En la sección 1.5 se exponen algunos ejemplos. En la sección 1.6 se compara la cota inferior de la probabilidad a posteriori obtenida en la sección 1.4 con el p-valor. Finalmente, la sección 7 contiene algunos comentarios.

1.2. Formulación del Problema

Sean X_i , $i = 1, 2$, variables aleatorias independientes y distribuidas, respectivamente, $B(n_i, p_i)$, $i = 1, 2$, con $n_i \in \mathbb{N}$, $i = 1, 2$, fijos y conocidos.

Se pretende contrastar

$$H_0 : p_1 = p_2 = p_0 \text{ versus } H_1 : p_1 \neq p_2 , \quad (1.2)$$

donde p_0 es un valor conocido, y $p_1 \neq p_2$ significa que al menos uno de ellos es distinto de p_0 , es decir, $(p_1, p_2) \neq (p_0, p_0)$.

Además, se supone que la opinión a priori sobre $\theta = (p_1, p_2)$ viene dada por la densidad $\pi(p_1, p_2) \in \Gamma$, siendo Γ la clase de distribuciones ε -contaminadas dada en la expresión (1.1).

En consecuencia, para contrastar (1.2) se necesita una distribución a priori de tipo mixto. Se propone

$$\pi^*(p_1, p_2) = \pi_0 I_{H_0}(p_1, p_2) + (1 - \pi_0) \pi(p_1, p_2) I_{H_1}(p_1, p_2), \quad (1.3)$$

siendo π_0 la masa a priori asignada a la hipótesis nula.

Aunque no hay una regla para fijar el valor de π_0 , usualmente se emplea $\pi_0 = \frac{1}{2}$ (véase Robert, 1994, capítulo 5).

Ahora, se van a considerar las hipótesis más realistas

$$H_{0\delta} : d((p_0, p_0), (p_1, p_2)) \leq \delta, \text{ versus } H_{1\delta} : d((p_0, p_0), (p_1, p_2)) > \delta, \quad (1.4)$$

con una métrica d apropiada, por ejemplo la distancia euclídea, y un valor de $\delta > 0$ suficientemente pequeño, de tal forma que cualquier punto (p_1, p_2) tal que $d((p_0, p_0), (p_1, p_2)) \leq \delta$ pueda ser considerado indistinguible de (p_0, p_0) .

Se pueden considerar varias formas de especificar $d((p_1, p_2), (p_0, p_0))$, una de ellas sería considerar un valor arbitrario de δ y dividirlo en dos valores δ_1 y δ_2 , quizá $\delta_1 = \delta_2 = \frac{\delta}{2}$, entonces, construiríamos la distancia empezando por $|p_i - p_0| < \delta_i, i = 1, 2$. Otra forma sería considerar

$$B((p_0, p_0), \delta) = \left\{ (p_1, p_2) \in (0, 1) \times (0, 1), (p_1 - p_0)^2 + (p_2 - p_0)^2 \leq \delta^2 \right\},$$

la bola de centro (p_0, p_0) y radio δ .

Aplicando el método de Gómez-Villegas y Sanz (2000) y Gómez-Villegas, Maín y Sanz (2002), introducido por Gómez-Villegas y Gómez Sánchez-Manzano (1992) y justificado por Gómez-Villegas y Sanz (1998), se podría utilizar $\pi(p_1, p_2)$, la opinión sobre (p_1, p_2) , y calcular π_0 promediando,

$$\pi_0 = \int_{B((p_0, p_0), \delta)} \pi(p_1, p_2) dp_2 dp_1. \quad (1.5)$$

En particular, es posible calcular el valor de δ en (1.5) para que $\pi_0 = \frac{1}{2}$. Cabe destacar que si la opinión a priori sobre (p_1, p_2) viene dada por la densidad uniforme, $\pi(p_1, p_2) = 1$,

$p_1, p_2 \in (0, 1)$, entonces el valor de π_0 que se obtiene a partir de la expresión (1.5), para δ suficientemente pequeño, es $\pi_0 = \pi\delta^2$, el área de la circunferencia de radio δ .

Por lo tanto, la probabilidad a priori asignada a H_0 mediante $\pi^*(p_1, p_2)$ y la asignada a $H_{0\delta}$ mediante $\pi(p_1, p_2)$ es la misma, eligiendo un valor adecuado de δ .

Varias razones pueden justificar esta aproximación mediante la elección de π_0 dado en (1.5). Una discusión interesante y una posterior justificación de esta construcción mediante el uso de la medida de información de Kullback-Leibler sobre el cambio de (1.2) por (1.4) pueden encontrarse en Gómez-Villegas y Sanz (2000) y Gómez-Villegas, Maín y Sanz (2002).

Si $\pi(\theta)$ es la información a priori, entonces la distribución a priori mixta $\pi^*(\theta)$ debería estar próxima a $\pi(\theta)$ en algún sentido. Utilizando la medida de información de Kullback-Leibler,

$$\mu(\pi^*, \pi) = \int \pi(\theta) \operatorname{Ln} \frac{\pi(\theta)}{\pi^*(\theta)} d\theta,$$

como medida de la discrepancia entre π y π^* , se tiene $\mu(\pi^*, \pi) \rightarrow 0$, cuando $\delta \rightarrow 0$. Quizá la elección de δ sea más intuitiva que la selección de un valor arbitrario para π_0 , usualmente $\pi_0 = \frac{1}{2}$ en la literatura.

En la misma línea de Berger y Sellke (1987), se pretende minimizar $P(H_0|a, c)$ sobre la clase Γ de distribuciones a priori dada en la expresión (1.1). A partir de (1.5) se obtiene que

$$\pi_0 = (1 - \varepsilon) \pi_{q_0}^0 + \varepsilon \pi_q^0,$$

donde

$$\pi_{q_0}^0 = \int_{B((p_0, p_0), \delta)} q_0(p_1, p_2) dp_2 dp_1 \text{ y } \pi_q^0 = \int_{B((p_0, p_0), \delta)} q(p_1, p_2) dp_2 dp_1. \quad (1.6)$$

Un motivo que justifica la toma del ínfimo, es que para un valor pequeño de dicho ínfimo la hipótesis nula debe ser rechazada, con lo que se tiene la misma interpretación del p-valor. Además, este desarrollo es similar al que en Casella y Berger (1987) reconcilia la

evidencia bayesiana y frecuentista en el problema del contraste de hipótesis unilaterales y nuestro objetivo es poner de manifiesto los motivos de la discrepancia entre ambas aproximaciones para el problema del contraste de hipótesis nula simple dado en (1.2).

Hay una extensa literatura sobre la comparación entre medidas clásicas y bayesianas. Algunas referencias importantes, además de las ya mencionadas, son Edwards, Lindman y Savage (1963), Pratt (1965), Dickey y Lienz (1970), Cox y Hinckley (1974), Bernardo (1980), Spiegelhalter y Smith (1982), Rubin (1984), Ghosh y Mukerjee (1992), McCulloch y Rossi (1992), Mukhopadhyay y DasGupta (1997), Berger, Boukai y Wang (1997, 1999), Oh y DasGupta (1999), De la Horra y Rodríguez-Bernal (2003), Gómez-Villegas, Maín, Sanz y Navarro (2004) y De la Horra (2005).

1.3. Notación

Se va a denotar a la función de verosimilitud mediante

$$f(a, c | p_1, p_2) = \binom{n_1}{a} \binom{n_2}{c} p_1^a (1 - p_1)^{n_1 - a} p_2^c (1 - p_2)^{n_2 - c},$$

que se considera como una función de $\theta = (p_1, p_2)$ para el valor observado de $(X_1, X_2) = (a, c)$, $a = 0, 1, \dots, n_1$, $c = 0, 1, \dots, n_2$. Si la distribución marginal de (X_1, X_2) , con respecto a la distribución a priori $\pi \in \Gamma$, se denota mediante $m(a, c | \pi)$, entonces

$$m(a, c | \pi) = (1 - \varepsilon) m(a, c | q_0) + \varepsilon m(a, c | q).$$

Por lo tanto, suponiendo que existan las distribuciones a posteriori $q_0(p_1, p_2 | a, c)$ y $q(p_1, p_2 | a, c)$, la distribución a posteriori de (p_1, p_2) dado (a, c) con respecto a $\pi \in \Gamma$ es

$$\pi(p_1, p_2 | a, c) = \lambda(a, c) q_0(p_1, p_2 | a, c) + (1 - \lambda(a, c)) q(p_1, p_2 | a, c),$$

donde $\lambda(a, c) = \frac{(1 - \varepsilon)m(a, c | q_0)}{m(a, c | \pi)}$.

Una medida clásica de la evidencia en contra de la hipótesis nula, que depende de las observaciones, es el p-valor. Si $T = T(X_1, X_2)$ es un estadístico adecuado para contrastar

(1.4), por ejemplo un *estadístico suficiente*, el p-valor correspondiente al punto (a, c) del espacio muestral es

$$p(a, c) = \sup_{(p_1, p_2) \in H_{0\delta}} P(|T(X_1, X_2)| > |T(a, c)| | (p_1, p_2)).$$

En particular, para contrastar (1.2), el p-valor es de la forma

$$p(a, c) = P(|T(X_1, X_2)| > |T(a, c)| | (p_0, p_0)).$$

En la sección 1.6 se consideran dos estadísticos diferentes para contrastar (1.2).

Con este procedimiento, la decisión de aceptar o rechazar H_0 depende de lo grande o pequeño que sea el p-valor, es decir, se rechaza H_0 si $p < p^*$, siendo $p^* \in (0, 1)$ un valor suficientemente pequeño.

1.4. Acotaciones de la Probabilidad A Posteriori

En esta sección se obtiene una cota inferior de la probabilidad a posteriori de la hipótesis nula del contraste (1.2), para la distribución a priori π^* dada en la expresión (1.3) y la probabilidad a priori de la hipótesis nula π_0 dada en (1.5).

Si se considera el contraste de hipótesis introducido en (1.2), una distribución a priori arbitraria $\pi \in \Gamma$ como en (1.1) y una distribución a priori de tipo mixto π^* dada en (1.3) con masa asignada a la hipótesis nula π_0 según (1.5), entonces la probabilidad a posteriori de $H_0 : p_1 = p_2 = p_0$ es

$$P(H_0|a, c) = \left[1 + \frac{1 - \pi_0 \int_0^1 \int_0^1 p_1^a (1 - p_1)^b p_2^c (1 - p_2)^d \pi(p_1, p_2) dp_2 dp_1}{\pi_0 p_0^{a+c} (1 - p_0)^{b+d}} \right]^{-1}.$$

Para calcular una cota inferior de la probabilidad a posteriori de H_0 es suficiente calcular una cota superior de $\frac{1 - \pi_0}{\pi_0} m(a, c|\pi)$, cuando $\pi \in \Gamma$. Por la construcción de $\pi^*(p_1, p_2)$, π_0 depende de $q \in Q$ a través del valor π_q^0 dado en la expresión (1.6). Por lo tanto, dicha cota inferior se puede calcular evaluando el supremo cuando $q \in Q$ de

$$\frac{1 - \pi_0}{\pi_0} m(a, c|\pi) = \left[\frac{1}{(1 - \varepsilon) \pi_{q_0}^0 + \varepsilon \pi_q^0} - 1 \right] [(1 - \varepsilon) m(a, c|q_0) + \varepsilon m(a, c|q)]. \quad (1.7)$$

Como el supremo de (1.7) cuando $q \in Q$ es siempre menor o igual que el producto de

$$\sup_{q \in Q} \left[\frac{1}{(1 - \varepsilon) \pi_{q_0}^0 + \varepsilon \pi_q^0} - 1 \right] = \frac{1}{(1 - \varepsilon) \pi_{q_0}^0}$$

y

$$\sup_{q \in Q} [(1 - \varepsilon) m(a, c|q_0) + \varepsilon m(a, c|q)],$$

siendo

$$m(a, c|q) \propto \int_0^1 \int_0^1 p_1^a (1 - p_1)^b p_2^c (1 - p_2)^d q(p_1, p_2) dp_2 dp_1 \leq$$

$$\sup_{(p_1, p_2) \neq (p_0, p_0)} p_1^a (1 - p_1)^b p_2^c (1 - p_2)^d,$$

entonces

$$P(H_0|a, c) \geq \left[1 + \frac{1 - (1 - \varepsilon) \pi_{q_0}^0}{(1 - \varepsilon) \pi_{q_0}^0} \eta_\varepsilon(a, c) \right]^{-1}, \quad (1.8)$$

donde

$$\eta_\varepsilon(a, c) = (1 - \varepsilon) \frac{\int_0^1 \int_0^1 p_1^a (1 - p_1)^b p_2^c (1 - p_2)^d q_0(p_1, p_2) dp_2 dp_1}{p_0^{a+c} (1 - p_0)^{b+d}} +$$

$$\varepsilon \frac{\left(\frac{a}{a+b}\right)^a \left(\frac{b}{a+b}\right)^b \left(\frac{c}{c+d}\right)^c \left(\frac{d}{c+d}\right)^d}{p_0^{a+c} (1 - p_0)^{b+d}}.$$

La expresión anterior proporciona una cota inferior para la probabilidad a posteriori de la hipótesis nula, por lo tanto la primera cuestión a plantearse es si es posible alcanzar dicho ínfimo mediante una distribución de la clase Γ dada en (1.1). La respuesta a dicha pregunta viene dada en el siguiente teorema.

Teorema 1.4.1 Sea (\hat{p}_1, \hat{p}_2) estimador de máxima verosimilitud de (p_1, p_2) cuando $(p_1, p_2) \in H_1$. Si $(\hat{p}_1, \hat{p}_2) \notin B((p_0, p_0), \delta)$ y, para ρ fijo, $\int_{B((p_0, p_0), \rho)} f(a, c|p_1, p_2) dp_2 dp_1$ se aproxima mediante $\pi \rho^2 f(a, c|\hat{p}_1, \hat{p}_2)$, entonces la distribución dada por

$$\tilde{\pi}(p_1, p_2) = (1 - \varepsilon) q_0(p_1, p_2) + \varepsilon \tilde{q}(p_1, p_2),$$

donde $\tilde{q}(p_1, p_2)$ es uniforme en $B((\hat{p}_1, \hat{p}_2), \rho)$, verifica que

$$\inf_{\pi \in \Gamma} P_\pi(H_0|a, c) = P_\pi(H_0|a, c) = \left[1 + \frac{1 - (1 - \varepsilon) \pi_{q_0}^0}{(1 - \varepsilon) \pi_{q_0}^0} \eta_\varepsilon(a, c) \right]^{-1}, \quad (1.9)$$

donde $\pi_{q_0}^0$ y $\eta_\varepsilon(a, c)$ están dados en (1.6) y (1.8), respectivamente.

Demostración. Por (1.7), se necesita calcular π_0 y $m(a, c|\tilde{\pi})$. Puesto que para $\tilde{\pi}$,

$$\begin{aligned} \pi_0 &= \int_{B((p_0, p_0), \delta)} \tilde{\pi}(p_1, p_2) dp_2 dp_1 \\ &= (1 - \varepsilon) \int_{B((p_0, p_0), \delta)} q_0(p_1, p_2) dp_2 dp_1 + \varepsilon \int_{B((p_0, p_0), \delta)} \tilde{q}(p_1, p_2) dp_2 dp_1 = (1 - \varepsilon) \pi_{q_0}^0 \end{aligned}$$

y

$$m(a, c|\tilde{\pi}) = (1 - \varepsilon) m(a, c|q_0) + \varepsilon m(a, c|\tilde{q}),$$

donde

$$\begin{aligned} m(a, c|\tilde{q}) &= \int_0^1 \int_0^1 f(a, c|p_1, p_2) \tilde{q}(p_1, p_2) dp_2 dp_1 \\ &= \frac{1}{\pi \rho^2} \int_{B((\hat{p}_1, \hat{p}_2), \rho)} f(a, c|p_1, p_2) dp_2 dp_1 \approx f(a, c|\hat{p}_1, \hat{p}_2) \\ &= \binom{n_1}{a} \binom{n_2}{c} \left(\frac{a}{a+b} \right)^a \left(\frac{b}{a+b} \right)^b \left(\frac{c}{c+d} \right)^c \left(\frac{d}{c+d} \right)^d, \end{aligned}$$

con lo que se obtiene (1.9).

Es interesante observar que la restricción real en este teorema es $(\hat{p}_1, \hat{p}_2) \notin B((p_0, p_0), \delta)$, puesto que en este caso, eligiendo un valor suficientemente pequeño de ρ , la aproximación de la integral es siempre posible y $B((p_0, p_0), \delta) \cap B((\hat{p}_1, \hat{p}_2), \rho)$ es vacía. Si $(\hat{p}_1, \hat{p}_2) \in B((p_0, p_0), \delta)$, la desigualdad dada en la expresión (1.9) es estricta.

1.5. Ejemplos

Una posible distribución inicial consiste en asignar distribuciones a priori uniformes o de Laplace independientes, es decir,

$$q_0(p_1, p_2) = I_{(0, 1)}(p_1) I_{(0, 1)}(p_2).$$

En este caso, la cota inferior de la probabilidad a posteriori de la hipótesis nula se obtiene evaluando la expresión (1.8) en

$$\eta_\varepsilon(a, c) = (1 - \varepsilon) \eta(a, c) + \varepsilon \frac{\left(\frac{a}{a+b}\right)^a \left(\frac{b}{a+b}\right)^b \left(\frac{c}{c+d}\right)^c \left(\frac{d}{c+d}\right)^d}{p_0^{a+c} (1 - p_0)^{b+d}}, \quad (1.10)$$

siendo $\eta(a, c) = p_0^{-m_1} (1 - p_0)^{-m_2} \frac{\Gamma(a+1)\Gamma(b+1)}{\Gamma(a+b+2)} \frac{\Gamma(c+1)\Gamma(d+1)}{\Gamma(c+d+2)}$.

Se puede observar que fijados n_1, n_2 y κ , a lo sumo existen cuatro tablas 2×2 pertenecientes al conjunto $A_\kappa = \{(a, c), \eta(a, c) = \kappa\}$. Además, la función

$$g(a, c) = \left(\frac{a}{a+b}\right)^a \left(\frac{b}{a+b}\right)^b \left(\frac{c}{c+d}\right)^c \left(\frac{d}{c+d}\right)^d$$

es constante sobre A_κ y para p_0 fijo, la función $p_0^{a+c} (1 - p_0)^{b+d}$ toma a lo sumo dos valores distintos sobre A_κ . En esta situación,

$$P(H_0|\kappa) \geq \left[1 + \frac{1 - (1 - \varepsilon) \pi_{q_0}^0}{(1 - \varepsilon) \pi_{q_0}^0} \eta_\varepsilon(\kappa)\right]^{-1}, \quad (1.11)$$

donde $\eta_\varepsilon(\kappa) = (1 - \varepsilon) \kappa + \varepsilon \frac{g(\kappa)}{\min_{A_\kappa} p_0^{a+c} (1 - p_0)^{b+d}}$.

Una asignación más general consiste en utilizar distribuciones a priori betas independientes, es decir,

$$q_0(p_1, p_2) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(\gamma + \delta)}{\Gamma(\gamma)\Gamma(\delta)} p_1^{\alpha-1} (1 - p_1)^{\beta-1} p_2^{\gamma-1} (1 - p_2)^{\delta-1},$$

para $p_1, p_2 \in (0, 1)$, $(\alpha, \beta, \gamma, \delta > 0)$. En este caso, la cota de la probabilidad a posteriori de la hipótesis nula se obtiene evaluando la expresión (1.8) en

$$\eta(a, c) = p_0^{-m_1} (1 - p_0)^{-m_2} \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(\gamma + \delta)}{\Gamma(\gamma)\Gamma(\delta)} \frac{\Gamma(a + \alpha)\Gamma(b + \beta)}{\Gamma(a + b + \alpha + \beta)} \frac{\Gamma(c + \gamma)\Gamma(d + \delta)}{\Gamma(c + d + \gamma + \delta)}.$$

1.6. Comparación con el Método Clásico

Desde el punto de vista clásico, en lugar de considerar los valores observados (a, c) fijos y permitir que (p_1, p_2) varíe, se fija el punto (p_0, p_0) de la hipótesis nula y después se calcula la probabilidad de observar un punto en alguna *región extrema* de la hipótesis alternativa en la que esté incluido (a, c) , es decir, en lugar de calcular la probabilidad a posteriori de la hipótesis nula, se calcula el p-valor. (La idea es básicamente que o H_0 es falsa, o ha ocurrido un suceso con probabilidad muy baja.)

En los contrastes paramétricos de hipótesis nula simple es sabido que los métodos clásicos y bayesianos pueden dar lugar a diferentes decisiones, véase Lindley (1957), Berger y Selke (1987) y Berger y Delampady (1987), entre otros. En la mayoría de las aproximaciones bayesianas se considera el ínfimo de la probabilidad a posteriori del punto de la hipótesis nula o del factor Bayes sobre una amplia clase de distribuciones a priori, y se obtiene que el ínfimo es sustancialmente mayor que el correspondiente p-valor. Es necesario apuntar que en todos estos casos la masa asignada al punto de la hipótesis nula es $\frac{1}{2}$. Por otro lado, Casella y Berger (1987) muestran que no hay discrepancia en los problemas de contraste unilaterales.

En la mayoría de las contribuciones existentes se utiliza una clase de distribuciones a priori. Gómez-Villegas y Sanz (2004) utilizan la clase de las distribuciones a priori unimodales y simétricas para mostrar que los p-valores y las probabilidades a posteriori se pueden igualar para el contraste de hipótesis nula simple en el caso multivariante. Nuestro objetivo es comprobar que para contrastar (1.2) no hay discrepancia entre las aproximaciones clásica y bayesiana cuando la distribución a priori pertenece a la clase de las distribuciones ε -contaminadas.

Se puede observar que la cota inferior de la probabilidad a posteriori dada en la expresión (1.9) depende del estadístico $\eta(a, c)$, que se puede utilizar como *estadístico de contraste* para construir una *región crítica* y calcular el p-valor, $p(a, c)$, correspondiente al punto (a, c) observado del espacio muestral.

Si observado un punto (a_0, c_0) , κ_0 denota el valor de η en dicho punto, es decir,

$\eta(a_0, c_0) = \kappa_0$, entonces se puede calcular la probabilidad de que al repetirse el experimento, para (p_0, p_0) fijo, se pueda obtener un nuevo valor de η mayor o igual que κ_0 . Por lo tanto, $\{\eta \geq \kappa_0\}$ es una posible *región crítica*, y

$$p(a_0, c_0) = P\{\eta \geq \kappa_0 | (p_0, p_0)\} \quad (1.12)$$

$$= \sum_{\eta(a,c) \geq \kappa_0} f(a, c | p_0, p_0) = \sum_{\eta(a,c) \geq \kappa_0} \binom{n_1}{a} \binom{n_2}{c} p_0^{a+c} (1-p_0)^{b+d} = p(\kappa_0)$$

es el p-valor.

Sin embargo, el método clásico usual χ^2 de Pearson utiliza la variable aleatoria

$$\Lambda = \frac{a^2}{n_1 p_0} + \frac{b^2}{n_1 (1-p_0)} + \frac{c^2}{n_2 p_0} + \frac{d^2}{n_2 (1-p_0)} - N \quad (1.13)$$

como *estadístico de contraste*. La distribución asintótica de Λ cuando H_0 es cierta es una χ_2^2 . En este caso, si observado (a_0, c_0) , $\Lambda(a_0, c_0) = \lambda_0$, entonces la evidencia utilizada es el p-valor

$$p(a_0, c_0) = P\{\Lambda \geq \lambda_0 | (p_0, p_0)\} = P\left(\chi_2^2 \geq \lambda_0\right) = e^{-\frac{\lambda_0}{2}}. \quad (1.14)$$

En cualquier caso, con los procedimientos clásicos descritos anteriormente se rechaza H_0 si $p < p^*$, siendo $p^* \in (0, 1)$ un valor suficientemente pequeño.

El objetivo es determinar un valor adecuado de δ tal que los valores de la cota inferior de la probabilidad a posteriori dada en la expresión (1.11) estén próximos a los respectivos p-valores.

Una posibilidad es elegir un valor δ que verifique la ecuación

$$p(a, c) = \inf_{\pi \in \Gamma} P_{\pi}(H_0 | a, c).$$

Por ejemplo, si se utiliza η como *estadístico de contraste*, para la cota inferior dada en la expresión (1.11), se obtiene

$$\pi_{q_0}^0 = \frac{1}{1 - \varepsilon p(\kappa) \eta_{\varepsilon}(\kappa) + 1 - p(\kappa)} \frac{p(\kappa) \eta_{\varepsilon}(\kappa)}{p(\kappa) \eta_{\varepsilon}(\kappa)}. \quad (1.15)$$

Por lo tanto, la masa a priori que se asigna a la hipótesis nula dependería de los datos.

Quizá sería más conveniente considerar el valor

$$\pi_{q_0}^0 = \frac{1}{1 - \varepsilon} \frac{p(\kappa) \eta_\varepsilon(\kappa)}{p(\kappa) \eta_\varepsilon(\kappa) + 2p^* - p(\kappa)},$$

para el que se verifica que

$$\frac{p(\kappa)}{2p^*} = \inf_{\pi \in \Gamma} P_\pi(H_0 | a, c).$$

Esta posibilidad se contempla en el capítulo 2.

Una posibilidad para eliminar esta dependencia de los datos es reemplazar $p(\kappa)$ por el nivel de significación α . Además, si el valor elegido para δ está próximo a un valor obtenido mediante (1.15), entonces se consigue que el p-valor y el ínfimo de la probabilidad a posteriori de la hipótesis nula estén próximos, puesto que el ínfimo es una función continua de δ .

Tabla 1.2: cotas inferiores de la probabilidad a posteriori de $H_0 : p_1 = p_2 = \frac{1}{2}$, con $q_0(p_1, p_2) = I_{(0,1)}(p_1) I_{(0,1)}(p_2)$ y $\varepsilon = 0,2$, para tablas (a, c) con $P\{\eta \geq \kappa_0 | (p_0, p_0)\}$ próximo a $0,1, 0,05$ y $0,01$.

$P(\eta \geq \eta(a, c))$	0,1063	0,0905	0,0533	0,0442	0,0118	0,0097
$\pi_{q_0}^0 = 0,5$	0,1677	0,1594	0,1013	0,0825	0,0222	0,0182
$\pi_{q_0}^0 = 0,2$	0,1052	0,0996	0,0617	0,0499	0,0131	0,0107
$\pi_{q_0}^0 = 0,12$	0,1014	0,0961	0,0594	0,048	0,0126	0,0103
$\pi_{q_0}^0 = 0,11$	0,0978	0,0926	0,0572	0,0462	0,0121	0,0099
$\pi_{q_0}^0 = 0,1$	0,0942	0,0891	0,0549	0,0444	0,0116	0,0095
$\pi_{q_0}^0 = 0,09$	0,0906	0,0858	0,0528	0,0426	0,0111	0,0091
$\pi_{q_0}^0 = 0,08$	0,0871	0,0824	0,0507	0,0408	0,0106	0,0087
$\eta(a, c)$	1,025	1,064	1,794	2,07	7,68	8,28

La Tabla 1.2 muestra los valores de la cota inferior de la probabilidad a posteriori de H_0 dada en la expresión (1.11), para algunos valores específicos de $\eta(a, c)$ y de $\pi_{q_0}^0$, cuando

$p_0 = \frac{1}{2}$ y la opinión inicial $q_0(p_1, p_2) = I_{(0,1)}(p_1)I_{(0,1)}(p_2)$ está contaminada con $\varepsilon = 0,2$. Se puede observar que si se toma un valor adecuado de $\pi_{q_0}^0 = \pi_{q_0}^0(\delta) = \frac{\pi\delta^2}{1-\varepsilon}$, los valores de la cota inferior están próximos a los respectivos p-valores dados en la expresión (1.12). Por ejemplo, si se toma $\pi_{q_0}^0 \in (0,3, 0,35)$, entonces para $\delta \in (0,27, 0,3)$ las cotas inferiores de la probabilidad a posteriori son aproximadamente iguales a los p-valores. También se puede observar que cuando $\pi_{q_0}^0 = \frac{1}{2}$ la discrepancia entre las dos medidas es mayor.

Tabla 1.3: cotas inferiores de la probabilidad a posteriori de $H_0 : p_1 = p_2 = \frac{1}{2}$, con $q_0(p_1, p_2) = I_{(0,1)}(p_1)I_{(0,1)}(p_2)$ y $\varepsilon = 0$, para tablas (a, c) con $P\{\eta \geq \kappa_0 | (p_0, p_0)\}$ próxima a 0,1, 0,05 y 0,01.

$P(\eta \geq \eta(a, c))$	0,1063	0,0905	0,0533	0,0442	0,0118	0,0097
$\pi_{q_0}^0 = 0,5$	0,4938	0,4844	0,358	0,3257	0,1151	0,1077
$\pi_{q_0}^0 = 0,2$	0,1961	0,1902	0,1223	0,1077	0,0315	0,0293
$\pi_{q_0}^0 = 0,12$	0,1174	0,1135	0,0706	0,0618	0,0174	0,0162
$\pi_{q_0}^0 = 0,11$	0,107	0,104	0,0644	0,0563	0,0158	0,0147
$\pi_{q_0}^0 = 0,1$	0,0978	0,094	0,0583	0,0509	0,0142	0,0132
$\pi_{q_0}^0 = 0,09$	0,0879	0,085	0,0522	0,0456	0,0127	0,0118
$\pi_{q_0}^0 = 0,08$	0,0782	0,075	0,0462	0,0403	0,0111	0,0104
$\pi_{q_0}^0 = 0,07$	0,0684	0,066	0,0403	0,0351	0,0096	0,009
$\eta(a, c)$	1,025	1,064	1,794	2,07	7,68	8,28

El mismo estudio para $\varepsilon = 0$ se muestra en la Tabla 1.3. En este caso, se puede observar que si se toma $\pi_{q_0}^0 \in (0,09, 0,11)$, entonces para $\delta \in (0,17, 0,19)$ las cotas inferiores de la probabilidad a posteriori de H_0 son aproximadamente iguales a los p-valores.

La Tabla 1.4 muestra los valores de las cotas inferiores de la probabilidad a posteriori de H_0 obtenidas mediante la expresión (1.11), cuando $p_0 = \frac{1}{2}$, $q_0(p_1, p_2) = I_{(0,1)}(p_1)I_{(0,1)}(p_2)$ y $\varepsilon = 0$, para tablas (a, c) tales que el p-valor $P(\Lambda \geq \Lambda(a, c))$ da-

do en la expresión (1.14) esté próximo a los valores usuales, 0.1, 0.05 y 0.01. Se puede observar que si se toma $\pi_{q_0}^0 \in (0,09, 0,11)$, entonces para $\delta \in (0,17, 0,19)$ las cotas inferiores de la probabilidad a posteriori de H_0 son aproximadamente iguales a los p-valores. Cabe destacar que el rango de valores de $\pi_{q_0}^0$ que estabiliza la cota inferior de la probabilidad a posteriori de $H_0 : p_1 = p_2 = \frac{1}{2}$ en torno al p-valor del método clásico es el mismo para los estadísticos η y Λ . Esto pone de manifiesto que la discrepancia entre los p-valores correspondientes a ambos estadísticos no es muy grande.

Tabla 1.4: cotas inferiores de la probabilidad a posteriori de $H_0 : p_1 = p_2 = \frac{1}{2}$, con $q_0(p_1, p_2) = I_{(0,1)}(p_1) I_{(0,1)}(p_2)$ y $\varepsilon = 0$, para tablas (a, c) con $P\{\Lambda \geq \lambda_0 | (p_0, p_0)\}$ próxima a 0,1, 0,05 y 0,01.

$P(\Lambda \geq \Lambda(a, c))$	0,143	0,0868	0,052	0,0445	0,0138	0,0094
$\pi_{q_0}^0 = 0,5$	0,6095	0,4938	0,358	0,3257	0,1077	0,085
$\pi_{q_0}^0 = 0,2$	0,2807	0,1961	0,1223	0,1077	0,2931	0,0227
$\pi_{q_0}^0 = 0,12$	0,1755	0,1174	0,0706	0,0618	0,0162	0,0125
$\pi_{q_0}^0 = 0,11$	0,1617	0,107	0,0644	0,0563	0,0147	0,0113
$\pi_{q_0}^0 = 0,1$	0,1478	0,0978	0,0583	0,0509	0,0132	0,0102
$\pi_{q_0}^0 = 0,09$	0,1337	0,0879	0,0522	0,0456	0,0118	0,0091
$\pi_{q_0}^0 = 0,08$	0,1195	0,0782	0,0462	0,0403	0,0104	0,008
$\pi_{q_0}^0 = 0,07$	0,1051	0,0684	0,0403	0,0351	0,009	0,0069
$\Lambda(a, c)$	3,89	4,89	5,89	6,22	8,56	9,33

1.7. Comentarios

Los resultados obtenidos son una consecuencia de la metodología basada en la relación entre la hipótesis nula puntual del contraste (1.2) y la hipótesis nula más realista del contraste (1.4). En los términos de la medida de información de Kullback-Leibler, la

discrepancia entre $\pi(p_1, p_2) \in \Gamma$, siendo Γ la clase de distribuciones ε -contaminadas, dada en (1.1), y la distribución a priori de tipo mixto $\pi^*(p_1, p_2)$, dada en (1.3), justifica la elección de π_0 como en (1.5), con un valor adecuado de δ . Según este procedimiento, $\pi^*(p_1, p_2)$, utilizada para el contraste (1.2), está próxima a la distribución a priori continua $\pi(p_1, p_2)$, utilizada para el contraste (1.4), como puede verse en en Gómez-Villegas y Sanz (2000) y Gómez-Villegas, Maín y Sanz (2002).

Cuando $\pi(p_1, p_2)$ pertenece a la clase de distribuciones ε -contaminadas, la cota inferior de la probabilidad a posteriori del punto de la hipótesis nula se puede hacer próxima al p-valor, eligiendo un valor adecuado de δ , como se muestra en la sección 1.6. Estos resultados parecen indicar que en el problema del contraste de homogeneidad de distribuciones multinomiales independientes, la discrepancia observada entre las aproximaciones clásicas y Bayesianas es más acusada usando $\pi_0 = \frac{1}{2}$ en la distribución mixta. Gómez-Villegas y Sanz (2000) obtienen resultados similares en un contexto diferente.

Finalmente, siguiendo un razonamiento similar al desarrollado para tablas 2×2 , es posible la generalización de los resultados obtenidos al caso de tablas $r \times s$, cuando el punto de la hipótesis nula, \mathbf{p}_0 , es conocido, desconocido, o de forma funcional conocida, $\mathbf{p}_0 = \mathbf{p}(\omega)$, en el problema del contraste de homogeneidad de r poblaciones multinomiales independientes. Un estudio detallado de estos casos se realiza en el capítulo 2.

Capítulo 2

Análisis Bayesiano de Tablas de Contingencia

La visualización de los datos mediante tablas de contingencia se utiliza en diferentes aproximaciones de la inferencia estadística, por ejemplo, para abordar el contraste de homogeneidad de distribuciones multinomiales independientes. Mediante el procedimiento bayesiano introducido en el capítulo 1 para el contraste de hipótesis nula simple frente a alternativa bilateral en tablas de contingencia, dadas muestras independientes de dos distribuciones binomiales y tomando una distribución a priori de tipo mixto, se calcula la probabilidad a posteriori de que la proporción de éxitos en la primera población sea igual que en la segunda. Dicha probabilidad a posteriori se compara con el p-valor del procedimiento clásico, obteniéndose una reconciliación entre ambos resultados. Los resultados obtenidos se generalizan para tablas $r \times s$, cuando \mathbf{p}_0 es conocido, desconocido o de forma funcional conocida, $\mathbf{p}_0 = \mathbf{p}(\omega)$.

2.1. Introducción

En el mismo contexto de la sección 1.1 se supone que se dispone de muestras aleatorias independientes de dos poblaciones suficientemente grandes, siendo cada uno de sus ele-

mentos clasificado como éxito o fracaso. La primera muestra es de tamaño n_1 y produce a éxitos y b fracasos, la segunda es de tamaño n_2 y produce c éxitos y d fracasos. Los datos se visualizan en la Tabla 1.1. En esta situación, se precisa una medida cuantitativa de hasta qué punto los datos apoyan o no la hipótesis de que la proporción de éxitos en la primera población, p_1 , es igual a la proporción de éxitos en la segunda población, p_2 . Este problema, aparentemente sencillo, ha dado lugar a una extensa literatura, desde que Karl Pearson introdujera su ya clásico test de la χ^2 para valorar la bondad del ajuste (Pearson, 1900). Además es uno de los problemas en los que existen discrepancias manifiestas entre los métodos clásicos y Bayesianos y entre distintos tipos de análisis clásicos. Naturalmente, hay un número de variaciones a este problema. Algunas referencias bayesianas importantes se citan a continuación.

Howard (1998) aboga por el uso más frecuente de contrastes unilaterales y aborda el problema desde un punto de vista bayesiano, considerando como hipótesis de interés $H_1 : p_2 < p_1$ y $H_2 : p_1 < p_2$. Proporciona una medida cuantitativa de la fuerza de la evidencia en apoyo de la hipótesis más probable, en el supuesto de que se esté considerando que p_1 y p_2 no son ninguno 0 ó 1 y que $p_1 \neq p_2$, y observa que el contraste clásico de la χ^2 puede considerarse como una aproximación de una probabilidad a posteriori de $p_2 < p_1$, tomando distribuciones a priori de Jeffreys independientes, es decir,

$$\pi(p_1, p_2) \propto p_1^{-1/2} (1 - p_1)^{-1/2} p_2^{-1/2} (1 - p_2)^{-1/2} .$$

Además introduce una familia conjugada de distribuciones a priori que incorpora dependencia entre las opiniones iniciales sobre las dos poblaciones.

En esta misma línea de trabajo con hipótesis unilaterales del tipo $p_1 > p_2$, se pueden mencionar otras aproximaciones bayesianas al problema de comparación de dos proporciones para una tabla 2×2 : métodos *log-odds-ratio* y métodos *inverse-root-sine*, que calculan la probabilidad a posteriori, para distribuciones a priori beta, de que $\Lambda_1 - \Lambda_2 > 0$, siendo $\Lambda_i = \log p_i (1 - p_i)^{-1}$, y $\Lambda_i = \arcsen \sqrt{p_i}$, $i = 1, 2$, respectivamente, como medidas del grado en el que dos poblaciones son homogéneas. (véase Lee (1997), págs. 152-154).

Quintana (1998) desarrolla un modelo bayesiano no paramétrico para valorar la homogeneidad de r poblaciones multinomiales independientes, en tablas de contingencia $r \times s$ con tamaños muestrales fijos. Supone que el vector de clasificación de probabilidades es una muestra de una distribución F que se ajusta a un proceso Dirichlet, centrado en una medida de probabilidad α y con peso c . También asigna a c una distribución a priori y propone un factor Bayes.

Lindley (1988) expone un modelo de probabilidad para la formación de genotipos, AA, Aa y aa, utilizando dos parámetros, $\alpha = \frac{1}{2} \log \frac{4p_1p_3}{p_2^2}$ y $\beta = \frac{1}{2} \log \frac{p_1}{p_3}$. Considera un test bayesiano para el contraste $\alpha = 0$, frente a $\alpha \neq 0$, basado en el factor Bayes, donde $\alpha = 0$ corresponde a la hipótesis nula de equilibrio Hardy-Weinberg, $H_0 : p^2, 2p(1-p), (1-p)^2$, siendo p la proporción de alelos de tipo A en la población.

Se considera el contraste de igualdad de proporciones de poblaciones multinomiales independientes, cuando las proporciones comunes son conocidas. Nuestra aproximación general al problema de homogeneidad, consiste en trabajar directamente con la hipótesis nula simple, calculando su probabilidad a posteriori. Para ello, se sigue el método utilizado por Gómez-Villegas y Sanz (2000) y Gómez-Villegas, Maín y Sanz (2002), que consiste en asignar una probabilidad inicial π_0 a la hipótesis nula, y repartir la probabilidad restante en los puntos de la alternativa, mediante una densidad a priori $\pi(p_1, p_2)$. Con este procedimiento, en el contexto de la hipótesis nula puntual, se consigue una reconciliación entre el p-valor y la probabilidad a posteriori de la hipótesis nula.

En la sección 2.2 se formula de forma precisa el problema y se calcula una expresión exacta de la probabilidad a posteriori de que la proporción de éxitos en la primera población sea igual que en la segunda, e igual a un valor común p_0 conocido. En la sección 2.3 se llega a una reconciliación entre el resultado clásico y bayesiano, y se consideran los datos de Pearson (1947) (véase la Tabla 2.2) para ilustrar el procedimiento. En la sección 2.4 se generalizan los resultados de las secciones 2.2 y 2.3 para tablas $r \times s$. En la sección 2.6 se expone un resumen de conclusiones. En la sección 2.7 se desarrollan tres métodos bayesianos que permiten abordar el problema de homogeneidad de poblaciones

multinomiales para tablas de contingencia $r \times s$ cuando \mathbf{p}_0 sea desconocido, y se compara el resultado clásico con los resultados obtenidos con dichos métodos Bayesianos para los datos del ejemplo de Pearson. En la sección 2.8 se generalizan los resultados de la sección 2.7 en el supuesto de que la forma funcional de \mathbf{p}_0 sea conocida a falta de q parámetros por determinar, $q < s$, y se consideran los datos del ejemplo de Lindley (1998) (véase la Tabla 2.7) para ilustrar dichos procedimientos. En la sección 2.8 se concluye con algunos comentarios.

Tabla 2.1: ejemplo de Pearson.

	Éxitos	Fracasos	Total
Muestra 1	3	15	18
Muestra 2	7	5	12
Total	10	20	30

2.2. Formulación del Problema y Cálculo de la Probabilidad A Posteriori

Al igual que en la sección 1.2, sean X_i , $i = 1, 2$, variables aleatorias binomiales, $B(n_i, p_i)$, $i = 1, 2$, independientes, con $n_i \in \mathbb{N}$, $i = 1, 2$, fijos y conocidos.

Se pretende contrastar (1.2) con p_0 conocido. En el capítulo 1 se utilizó la clase de distribuciones ε -contaminadas para describir la opinión que se tiene a priori sobre $\theta = (p_1, p_2)$. En este capítulo se va a suponer que $\varepsilon = 0$, es decir, la opinión a priori viene dada por una densidad particular $\pi(\theta) = \pi(p_1, p_2)$.

En consecuencia, para contrastar (1.2) se necesita una distribución a priori de tipo mixto. Se propone (1.3), en cuyo caso, la probabilidad a posteriori de la hipótesis nula, observados los datos de la Tabla 1.1, es

$$P(H_0|a, c) = \frac{p_0^{a+c} (1-p_0)^{b+d} \pi_0}{p_0^{a+c} (1-p_0)^{b+d} \pi_0 + (1-\pi_0) \int_0^1 \int_0^1 p_1^a (1-p_1)^b p_2^c (1-p_2)^d \pi(p_1, p_2) dp_2 dp_1}$$

donde π_0 es la probabilidad a priori inicial asignada a $H_0 : p_1 = p_2 = p_0$.

Una posible distribución inicial consiste en asignar distribuciones a priori uniformes o de Laplace independientes, es decir,

$$\pi(p_1, p_2) = I_{(0, 1)}(p_1) I_{(0, 1)}(p_2).$$

En este caso, la probabilidad a posteriori de la hipótesis nula es

$$P(H_0|a, c) = \left[1 + \frac{1 - \pi_0}{\pi_0} \eta\right]^{-1}, \quad (2.1)$$

donde $\eta = p_0^{-m_1} (1 - p_0)^{-m_2} \frac{\Gamma(a+1)\Gamma(b+1)}{\Gamma(a+b+2)} \frac{\Gamma(c+1)\Gamma(d+1)}{\Gamma(c+d+2)}$.

Una asignación más general consiste en utilizar distribuciones a priori betas independientes, es decir,

$$\pi(p_1, p_2) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(\gamma + \delta)}{\Gamma(\gamma)\Gamma(\delta)} p_1^{\alpha-1} (1 - p_1)^{\beta-1} p_2^{\gamma-1} (1 - p_2)^{\delta-1},$$

siendo $p_1, p_2 \in (0, 1), (\alpha, \beta, \gamma, \delta > 0)$.

En este caso, la probabilidad a posteriori de la hipótesis nula se obtiene evaluando la expresión (2.1) en el estadístico

$$\eta = p_0^{-m_1} (1 - p_0)^{-m_2} \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(\gamma + \delta)}{\Gamma(\gamma)\Gamma(\delta)} \frac{\Gamma(a + \alpha)\Gamma(b + \beta)}{\Gamma(a + b + \alpha + \beta)} \frac{\Gamma(c + \gamma)\Gamma(d + \delta)}{\Gamma(c + d + \gamma + \delta)}.$$

La probabilidad a posteriori calculada en la expresión (2.1) depende de π_0 , la masa a priori inicial asignada a la hipótesis nula. Ahora, considerando las hipótesis de la expresión (1.4), para una métrica d apropiada y un valor de $\delta > 0$ suficientemente pequeño, se puede aplicar el método de Gómez-Villegas y Sanz (2000) y Gómez-Villegas, Maín y Sanz (2002) y calcular π_0 mediante (1.5). Con este segundo procedimiento, la probabilidad a posteriori obtenida en (2.1) puede ser expresada en términos de δ . Aquí, los resultados se obtienen en función de π_0 y posteriormente se especifican en términos de δ empleando la expresión (1.5).

Por otro lado, se puede observar que valores de δ que se corresponden con valores de $\pi_0 > \frac{\eta}{\eta+1}$, hacen que $P(H_0|a, c) > \frac{1}{2}$. Además, si δ es tal que $\pi_0 = \frac{1}{2}$, entonces $P(H_0|a, c) = \frac{1}{\eta+1}$.

En cualquier caso, cuando $p_0 = \frac{1}{2}$ en (1.2), cualquiera que sea el valor de π_0 elegido, la probabilidad a posteriori de la hipótesis nula se obtiene para

$$\eta = 2^N \frac{\Gamma(a+1)\Gamma(b+1)\Gamma(c+1)\Gamma(d+1)}{\Gamma(a+b+2)\Gamma(c+d+2)}. \quad (2.2)$$

Para los datos de la Tabla 2.1 se obtiene $\eta = 6,7265$ y, si $\delta = \frac{1}{\sqrt{2\pi}}$, entonces $\pi_0 = \frac{1}{2}$ y $P(H_0|a, c) = 0,1294$, con lo cual se rechaza H_0 . Además, para aceptar H_0 con los datos del ejemplo de Pearson, debe ser $\delta > 0,53905$ o equivalentemente $\pi_0 > 0,8706$. Por lo tanto, para dichos datos, se observa que existe un amplio rango de valores de δ , $\delta < 0,53905$, donde se rechaza H_0 .

2.3. Comparación con el Método Clásico

En la sección 1.6, siguiendo un razonamiento similar al de Gómez-Villegas y Sanz (2000), se comprobó que para contrastar (1.2), eligiendo un valor adecuado de δ , el ínfimo de la probabilidad a posteriori del punto de la hipótesis nula se puede hacer próximo al p-valor, cuando la distribución a priori pertenece a la clase de distribuciones ε -contaminadas. La metodología empleada se basaba en la búsqueda de un intervalo de valores de δ en el que ambas medidas, la clásica y la bayesiana, fueran aproximadamente iguales para valores de dicha cota inferior próximos a los p-valores usuales.

Ahora, el objetivo es comprobar que para contrastar (1.2) no hay discrepancia entre la aproximación clásica, expresada en términos del p-valor, y la bayesiana, en términos de la probabilidad a posteriori del punto de la hipótesis nula calculada en la sección anterior, cuando se utiliza una única distribución a priori.

Como es usual, si se considera, como medida de la evidencia a favor de H_1 , la discrepancia entre los valores observados y los valores esperados bajo H_0 cierta, en los términos del estadístico χ^2 de Pearson, el *estadístico de contraste* es la variable aleatoria Λ dada en la expresión (1.13). Entonces, observado un punto (a_0, c_0) , si λ_0 denota el valor de Λ en dicho punto, $\{\Lambda \geq \lambda_0\}$ es una posible *región crítica* y el valor dado en la expresión (1.14) es el p-valor.

Con este procedimiento, la decisión de aceptar o rechazar H_0 depende de lo grande o pequeño que sea el p-valor, es decir, se rechaza H_0 si $p < p^*$, siendo $p^* \in (0, 1)$ un valor suficientemente pequeño.

Si se pretende contrastar (1.2) con $p_0 = \frac{1}{2}$, mediante el método clásico descrito anteriormente, entonces el *estadístico de contraste* es la variable aleatoria

$$\Lambda = 2 \left[\frac{a^2 + b^2}{n_1} + \frac{c^2 + d^2}{n_2} \right] - N, \quad (2.3)$$

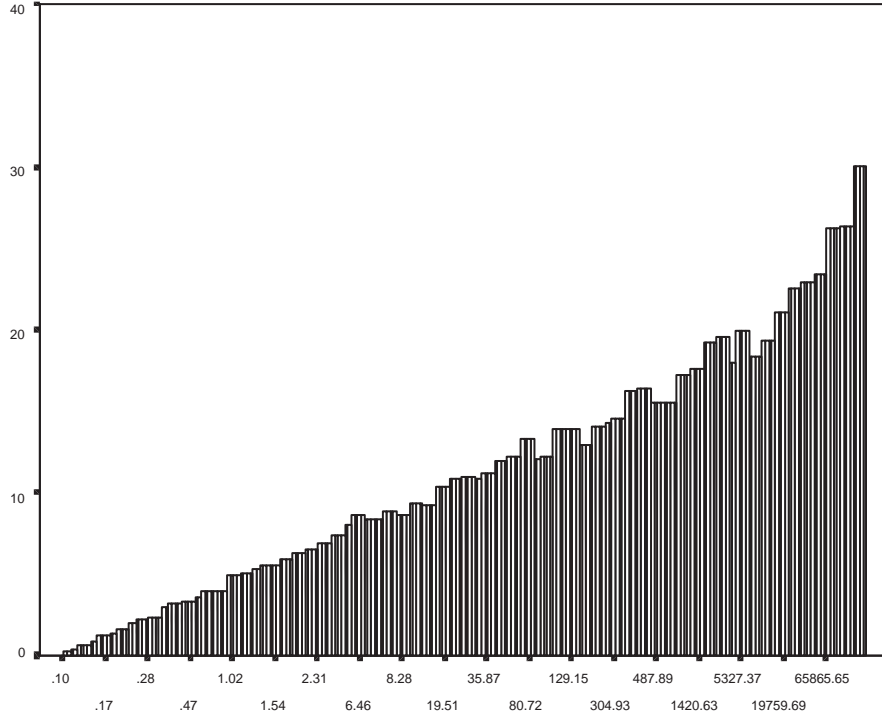
y la evidencia utilizada es el p-valor,

$$p = e^{\frac{N}{2} - \frac{a_0^2 + b_0^2}{n_1} - \frac{c_0^2 + d_0^2}{n_2}}. \quad (2.4)$$

Para los datos de la Tabla 2.1 se obtiene $\Lambda = 8,33333$, y un p-valor $p = 0,015504$. Se observa que para $p^* = 0,05$ se rechaza H_0 , pero para $p^* = 0,01$ no hay evidencia estadística suficiente para rechazar, y en ese sentido se acepta H_0 .

La probabilidad a posteriori del punto de la hipótesis nula, dada en la expresión (2.1), depende del estadístico η . Para comparar el método bayesiano propuesto con el método clásico χ^2 de Pearson, que utiliza como *estadístico de contraste* la variable aleatoria Λ dada en la expresión (1.13), sería interesante que existiese una dependencia funcional entre ambos estadísticos, η y Λ , (o entre la probabilidad a posteriori y el p-valor, p), es decir, $\eta = g(\Lambda)$, para cierta función $g : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ creciente. Sin embargo, en el caso de tablas 2×2 , si $n_1 = 18$ y $n_2 = 12$, se puede observar que para los datos del ejemplo de Pearson, $(a, c) = (3, 7)$, el valor de η que se obtiene a partir de la expresión (2.2) es 6,72 y el valor de Λ obtenido de la expresión (2.3) es 8,33333, mientras que si $(a, c) = (9, 1)$, se obtiene $\eta = 7,45$ y $\Lambda = 8,33333$, lo que pone de manifiesto que dicha dependencia funcional no es posible. Además, se puede observar que a diferencia de Λ , η si distingue entre las dos situaciones anteriores. No obstante, se puede comprobar que existe una función $h : \mathbb{R}^+ \rightarrow \mathbb{R}^+$, no monótona, tal que $\Lambda = h(\eta)$ (veáse la Figura 2.1). Por lo tanto, el test clásico admite una representación en términos del estadístico η . Cabe destacar, que ninguno de ellos es un *estadístico suficiente* para el modelo.

Figura 2.1: Diagrama de Barras $(\eta(a, c), \Lambda(a, c))$, para tablas 2×2 con $n_1 = 18$ y $n_2 = 12$. Existe una función no monótona, $h : \mathbb{R}^+ \rightarrow \mathbb{R}^+$, tal que $\Lambda = h(\eta)$.



En los contrastes paramétricos de hipótesis nula puntual, normalmente, existe una discrepancia entre el p-valor y la probabilidad a posteriori que puede dar lugar a controversias entre los métodos clásicos y bayesianos, debido a que, con frecuencia, la probabilidad a posteriori es considerablemente mayor que el p-valor, cuando se utiliza una distribución a priori particular. Para llegar a algún tipo de reconciliación entre la aproximación clásica y la bayesiana sería conveniente la equivalencia de ambas medidas, bien en cuanto a la decisión que se deriva de ellas, o bien que un mismo número tuviera ambas interpretaciones. A partir de ahora, se hará referencia a estas dos situaciones de equivalencia, como acuerdo cualitativo o cuantitativo, respectivamente.

Ahora, se va a considerar la siguiente ecuación,

$$\left[1 + \frac{1 - \pi_0}{\pi_0} \eta\right]^{-1} = \frac{p}{2p^*},$$

a partir de la cual se puede despejar el valor de π_0 ,

$$\pi_0 = \left[1 + \frac{1}{\eta} \left(\frac{2p^*}{p} - 1 \right) \right]^{-1} = \frac{\eta p}{\eta p + 2p^* - p}, \quad (2.5)$$

que verifica que, si (a, c) es tal que $P(H_0|a, c) > \frac{1}{2}$, entonces $p(a, c) > p^*$ y recíprocamente. Por lo tanto, utilizando el valor de π_0 obtenido en la expresión (2.5) se llegaría, en términos cualitativos, a la misma conclusión con los dos métodos, clásico y bayesiano. Además, si $p^* = \frac{1}{2}$, para dicho valor de π_0 se consigue que la probabilidad a posteriori sea igual al p-valor.

Se puede observar que el valor de π_0 dado en la expresión (2.5) verifica que $0 < \pi_0 < 1$ para tablas (a, c) tales que $2p^*(a, c) > p$. Por lo tanto, fijado p^* , $0 < p^* < 1$, para (a, c) tal que $0 < p(a, c) < p^*$ ó $p^* < p(a, c) < 2p^*$, existe una probabilidad a priori inicial π_0 , $0 < \pi_0 < 1$, que asignada a la hipótesis nula del contraste (1.2) hace que la decisión que se tome con ambos métodos sea la misma. Se observa también, que si $p^* = \frac{1}{2}$ entonces, cualquiera que sea el punto observado (a, c) , dicho π_0 siempre existe y verifica que $P(H_0|a, c) = p(a, c)$.

Tabla 2.2: resumen de resultados para los datos del ejemplo de Pearson, $p_0 = \frac{1}{2}$ y

$$\pi(p_1, p_2) = I_{(0, 1)}(p_1) I_{(0, 1)}(p_2).$$

κ	p	η	$(\eta + 1)^{-1}$	$\eta(\eta + 1)^{-1}$	δ
8,333333	0,015504	6,7265	0,1294	0,8706	0,53905

Tabla 2.1	$p^* = 0,5$	$p^* = 0,1$	$p^* = 0,05$	$p^* = 0,01$
$\eta p (\eta p + 2p^* - p)^{-1}$	0,09578	0,36113	0,5524	0,9587
δ	0,17461	0,33904	0,41933	0,6085
$p(2p^*)^{-1}$	0,015504	0,07752	0,15503	0,77523

Para los datos de la Tabla 2.1, si $p^* = \frac{1}{2}$ y $\pi(p_1, p_2) = I_{(0, 1)}(p_1) I_{(0, 1)}(p_2)$, el valor π_0 que reconcilia el p-valor, $p = 0,015504$, con la probabilidad a posteriori es $\pi_0 = 0,09578$. Si $p^* = 0,1$ se obtiene $\pi_0 = 0,36113$ y se rechaza con una probabilidad a posteriori de $0,07752$. Si $p^* = 0,05$ se obtiene $\pi_0 = 0,5524$ y se rechaza con $0,15503$. Para $p^* = 0,01$ se obtiene $\pi_0 = 0,9587$ y se acepta con $0,77533$. Los resultados obtenidos se resumen en la Tabla 2.2. Además, se puede observar cómo el valor de π_0 , y por lo tanto el valor de δ , para el que se consigue acuerdo entre el resultado clásico y Bayesiano, en los términos expuestos anteriormente, disminuye cuando p^* aumenta.

No obstante, dicha reconciliación es demasiado estricta, ya que el valor obtenido en la expresión (2.5) depende de los datos a través del estadístico η , sin embargo da una idea de cómo debería ser el valor de δ para que la reconciliación entre ambos métodos sea posible. En este sentido, no se afirma que haya que igualar ambas expresiones, sino que la utilización de un valor de π_0 próximo al resultado de esa igualación puede proporcionar, cuando esto sea posible, un valor numérico cualitativamente o cuantitativamente igual desde ambos puntos de vista, clásico y bayesiano.

Para eliminar la dependencia de los datos, una posibilidad es formular el acuerdo en términos cuantitativos utilizando $p^* = \frac{1}{2}$ en (2.5). En este caso se intenta encontrar un intervalo de valores de $\pi_0 = \pi_0(\delta)$, tal que en dicho intervalo, las probabilidades a posteriori y los p-valores sean aproximadamente iguales, al menos para los p-valores usuales, $p \in (0,01, 0,1)$. De esta forma la probabilidad a posteriori no se comparara con el valor $\frac{1}{2}$ sino con p^* . Esta fue la metodología utilizada en la sección 1.6 donde numéricamente se comprobó que, para contrastar (1.2) con $p_0 = \frac{1}{2}$ y distribuciones a priori uniformes independientes, basta con elegir $\pi_0 \in (0,09, 0,11)$ o equivalentemente $\delta \in (0,17, 0,19)$ para que las probabilidades a posteriori obtenidas evaluando la expresión (2.1) en (2.2) estén próximas a los p-valores usuales obtenidos a partir de la expresión (2.4).

Sin embargo, como se demostrará en el capítulo 3, se puede desarrollar un procedimiento que permite formalizar dicha reconciliación cualitativamente, mediante un teorema, en

los términos de una condición suficiente. Por todo ello, de ahora en adelante los esfuerzos se centrarán en el método cualitativo.

Desde un punto de vista cualitativo lo deseable sería formular el acuerdo de manera que fijado , $p^* \in (0, 1)$, usualmente $p^* = 0,01, 0,05$ o $0,1$, existiera un intervalo de valores de $\pi_0 = \pi_0(\delta, p^*)$, $\pi_0 \in (\ell_1, \ell_2)$, para ciertos $\ell_1, \ell_2 \in (0, 1)$, $\ell_1 < \ell_2$, tal que, para contrastar (1.2), un clásico que actuara utilizando p^* para ponderar el p-valor, llegara al mismo resultado, en los términos de aceptar o rechazar, que un bayesiano utilizando un valor $\pi_0 \in (\ell_1, \ell_2)$ como masa a priori del punto de la hipótesis nula.

En este caso, para eliminar la dependencia de los datos en (2.5), se han generado todas las posibles tablas 2×2 , para n_1 y n_2 fijos y conocidos. Así, si $n_1 = 18$ y $n_2 = 12$, se generan las 247 tablas posibles. Los datos de Pearson se corresponden con la tabla que ocupa la posición 95 en la ordenación en sentido ascendente realizada según los valores de η .

Para cada una de estas tablas se efectúa el mismo estudio que se ha efectuado previamente para los datos del ejemplo de Pearson. Mediante un análisis de datos sencillo se puede comprobar que existen valores de p^* , como por ejemplo $p^* = 0,5$, $p^* = 0,1$, $p^* = 0,05$ o $p^* = 0,01$, para los que existe un intervalo de valores de π_0 , $I = I(p^*, n_1 = 18, n_2 = 12)$, tal que el resultado que se obtiene con el método bayesiano propuesto para el contraste (1.2), con $p_0 = \frac{1}{2}$ y $\pi(p_1, p_2) = 1$, $p_1, p_2 \in (0, 1)$, utilizando un valor de $\pi_0 \in I$, es el mismo que se obtiene mediante el método clásico χ^2 de Pearson, (véase la Tabla 2.3). Por lo tanto existe acuerdo entre ambos métodos. Por ejemplo, si un clásico utilizara $p^* = 0,05$ para ponderar el p-valor, un bayesiano sólo tendría que usar $\pi_0 \in (0,643, 0,673)$ para llegar a la misma conclusión que aquel. No obstante, también existen valores de p^* , por ejemplo, $p^* = 0,015$, para los que esto no es posible. A modo de comprobación, con los datos de la Tabla 2.1, se obtuvo que $P(H_0|a, c) > \frac{1}{2}$ cuando $\pi_0 > 0,8706$, y $p = 0,015504$ (véase la Tabla 2.2). Así, cuando $p^* = 0,05$, eligiendo $\pi_0 \in (0,643, 0,673)$ los dos rechazarían H_0 , mientras que si $p^* = 0,01$, escogiendo $\pi_0 \in (0,893, 0,914)$ los dos aceptarían H_0 . Por lo tanto existe acuerdo entre ambos métodos.

Tabla 2.3: resumen de resultados para tablas 2×2 con $n_1 = 18$ y $n_2 = 12$, $p_0 = \frac{1}{2}$ y

$$\pi(p_1, p_2) = I_{(0, 1)}(p_1) I_{(0, 1)}(p_2).$$

$p^* \in$	(0,46, 0,513)	(0,087, 0,143)	(0,045, 0,052)	(0,0095, 0,0138)
$\delta \in$	(0,221, 0,23)	(0,353, 0,4)	(0,453, 0,462)	(0,5528, 0,5675)
$\pi_0 \in$	(0,153, 0,167)	(0,391, 0,506)	(0,643, 0,673)	(0,893, 0,914)

De nuevo, se puede observar como el valor de π_0 , y por lo tanto el valor de δ , para el que es posible la reconciliación anterior entre ambos métodos, disminuye cuando p^* aumenta. También se puede comprobar que el valor de π_0 calculado mediante la expresión (2.5) no siempre existe, y cuando existe, no siempre pertenece al intervalo de valores para los que se ha obtenido la reconciliación entre ambos métodos. En el capítulo 3 se estudia esta cuestión con más detalle.

En general, fijados n_1 , n_2 y p^* , si se denotan mediante

$$\ell_1 = \ell_1(p^*, n_1, n_2) = \max_{(a, c), p > p^*} \eta(\eta + 1)^{-1},$$

$$\ell_2 = \ell_2(p^*, n_1, n_2) = \min_{(a, c), p \leq p^*} \eta(\eta + 1)^{-1},$$

y p^* verifica que $\ell_1 < \ell_2$, entonces, existe un intervalo de valores de $\pi_0 = \pi_0(\delta, p^*)$, $I = I(p^*, n_1, n_2) = (\ell_1, \ell_2)$, tal que si $\pi_0 \in I$, el resultado que se obtiene con el método bayesiano propuesto para el contraste (1.2), utilizando un valor de $\pi_0 \in I$, es el mismo que se obtiene mediante el método clásico χ^2 de Pearson, utilizando p^* . Este resultado se formaliza en el capítulo 3.

Es evidente que, en el ejemplo estudiado, la existencia de valores p^* para los cuales se verifica la condición suficiente que garantiza el acuerdo entre ambos métodos depende de la *tendencia creciente* que se puede observar (véase la Figura 2.1) en la relación funcional que existe entre ambos estadísticos, $\Lambda = h(\eta)$, aunque no sea estrictamente monótona.

2.4. Tablas $r \times s$ con \mathbf{p}_0 Conocido

A continuación se generalizan los resultados obtenidos previamente al caso de tablas $r \times s$. Para ello, se dispone de muestras aleatorias independientes de r poblaciones suficientemente grandes, de forma que cada uno de sus individuos pertenezca a una y sólo una de las s clases excluyentes C_1, \dots, C_s . La muestra i -ésima, $i = 1, \dots, r$, es de tamaño n_i y produce n_{ij} individuos en la categoría C_j , $j = 1, \dots, s$. Los datos se visualizan en la Tabla 2.4.

Tabla 2.4: datos en la tabla $r \times s$.

	Clase 1	Clase 2	...	Clase s	Total
Muestra 1	n_{11}	n_{12}	...	n_{1s}	n_1
Muestra 2	n_{21}	n_{22}	...	n_{2s}	n_2
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
Muestra r	n_{r1}	n_{r2}	...	n_{rs}	n_r
Total	m_1	m_2	...	m_s	N

Sean X_i , $i = 1, \dots, r$, variables aleatorias independientes, multinomiales, $MB(n_i, \mathbf{p}_i)$, con $\mathbf{p}_i = (p_{i1}, \dots, p_{is}) \in \Theta$, siendo $\Theta = \{\mathbf{p} = (p_1, \dots, p_s) \in (0, 1)^s, \sum_{j=1}^s p_j = 1\} \subset \mathbb{R}^{s-1}$, para $n_i \in \mathbb{N}$, $i = 1, \dots, r$, valores fijos y conocidos. En esta situación, se pretende contrastar

$$H_0 : \mathbf{p}_1 = \dots = \mathbf{p}_r = \mathbf{p}_0 \text{ versus } H_1 : \exists i \neq j, \mathbf{p}_i \neq \mathbf{p}_j, \quad (2.6)$$

donde $\mathbf{p}_0 = (p_{01}, \dots, p_{0s}) \in \Theta$ es un valor desconocido y $H_1 : \exists i \neq j, \mathbf{p}_i \neq \mathbf{p}_j$ significa que al menos uno de ellos es distinto de \mathbf{p}_0 . Además, supongamos que la opinión a priori sobre $(\mathbf{p}_1, \dots, \mathbf{p}_r)$ viene dada por la densidad $\pi(\mathbf{p}_1, \dots, \mathbf{p}_r) = \prod_{i=1}^r \pi(\mathbf{p}_i)$. Por lo tanto, para contrastar (2.6) se precisa una distribución a priori de tipo mixto. Se propone,

$$\pi^*(\mathbf{p}_1, \dots, \mathbf{p}_r) = \pi_0 I_{H_0}(\mathbf{p}_1, \dots, \mathbf{p}_r) + (1 - \pi_0) \pi(\mathbf{p}_1, \dots, \mathbf{p}_r) I_{H_1}(\mathbf{p}_1, \dots, \mathbf{p}_r),$$

siendo π_0 la masa a priori asignada a la hipótesis nula.

Por consiguiente, la probabilidad a posteriori de la hipótesis nula, observados los datos de la Tabla 2.4, es

$$\frac{\prod_{j=1}^s p_{0j}^{\sum_{i=1}^r n_{ij}} \pi_0}{\prod_{j=1}^s p_{0j}^{\sum_{i=1}^r n_{ij}} \pi_0 + (1 - \pi_0) \prod_{i=1}^r \int_{\Theta} \prod_{j=1}^s p_{ij}^{n_{ij}} \pi(\mathbf{p}_i) d\mathbf{p}_i}.$$

Sea $\alpha_i = (\alpha_{i1}, \dots, \alpha_{is})$, con $\alpha_{ij} > 0$, para todo $j = 1, \dots, s$ y para todo $i = 1, \dots, r$. Si se asigna a cada \mathbf{p}_i una distribución a priori de Dirichlet de parámetro α_i , $D(\alpha_i)$, $i = 1, \dots, r$, (véase Ghosh y Ramamoorthi (2003), capítulo 3), es decir

$$\pi(\mathbf{p}_i) = \frac{\Gamma\left(\sum_{j=1}^s \alpha_{ij}\right)}{\prod_{j=1}^s \Gamma(\alpha_{ij})} \prod_{j=1}^s p_{ij}^{\alpha_{ij}-1}, \quad \mathbf{p}_i = (p_{i1}, \dots, p_{is}) \in \Theta, \quad i = 1, \dots, r,$$

entonces, dicha probabilidad a posteriori es

$$\left[1 + \prod_{j=1}^s p_{0j}^{-m_j} \frac{1 - \pi_0}{\pi_0} \prod_{i=1}^r \left\{ \frac{\Gamma\left(\sum_{j=1}^s \alpha_{ij}\right)}{\prod_{j=1}^s \Gamma(\alpha_{ij})} \int_{\Theta} \prod_{j=1}^s p_{ij}^{n_{ij} + \alpha_{ij} - 1} d\mathbf{p}_i \right\} \right]^{-1}.$$

Por lo tanto, la probabilidad a posteriori de la hipótesis nula, observados los datos de la Tabla 2.4, se puede expresar de la siguiente forma:

$$\left[1 + \frac{1 - \pi_0}{\pi_0} \eta \right]^{-1}, \tag{2.7}$$

$$\text{siendo } \eta = \prod_{j=1}^s p_{0j}^{-m_j} \left\{ \frac{\prod_{i=1}^r \Gamma\left(\sum_{j=1}^s \alpha_{ij}\right)}{\prod_{i=1}^r \prod_{j=1}^s \Gamma(\alpha_{ij})} \right\} \left\{ \frac{\prod_{i=1}^r \prod_{j=1}^s \Gamma(n_{ij} + \alpha_{ij})}{\prod_{i=1}^r \Gamma\left(n_i + \sum_{j=1}^s \alpha_{ij}\right)} \right\}.$$

En concreto, si se asignan a los parámetros \mathbf{p}_i , $i = 1, \dots, r$, distribuciones a priori uniformes sobre Θ e independientes, entonces, la probabilidad a posteriori de la hipótesis nula se obtiene evaluando la expresión (2.7) en

$$\eta = \prod_{j=1}^s p_{0j}^{-m_j} \Gamma(s)^r \left\{ \frac{\prod_{i=1}^r \prod_{j=1}^s \Gamma(n_{ij} + 1)}{\prod_{i=1}^r \Gamma(n_i + s)} \right\}.$$

La probabilidad a posteriori calculada en la expresión (2.7) depende de π_0 , la probabilidad a priori inicial asignada a la hipótesis nula $H_0 : \mathbf{p}_1 = \dots = \mathbf{p}_r = \mathbf{p}_0$.

Denotando por $\mathbf{P}_0 = (\mathbf{p}_0, \dots, \mathbf{p}_0) \in \Theta^r \subset \mathbb{R}^{r(s-1)}$ y $\mathbf{P} = (\mathbf{p}_1, \dots, \mathbf{p}_r) \in \Theta^r \subset \mathbb{R}^{r(s-1)}$, la hipótesis nula del contraste (2.7) es $H_0 : \mathbf{P} = \mathbf{P}_0$.

Ahora, se puede pensar que en lugar de (2.6) es más realista

$$H_{0\delta} : d(\mathbf{P}_0, \mathbf{P}) \leq \delta, \text{ versus } H_{1\delta} : d(\mathbf{P}_0, \mathbf{P}) > \delta,$$

para una métrica d apropiada y un valor de $\delta > 0$ suficientemente pequeño.

Se propone utilizar $B(\mathbf{P}_0, \delta) = \left\{ \mathbf{P} \in \Theta^r, \sum_{i=1}^r \sum_{j=1}^s (p_{ij} - p_{0j})^2 \leq \delta^2 \right\}$. Entonces, aplicando el método de Gómez-Villegas y Sanz (2000) y Gómez-Villegas, Maín, y Sanz (2002), se podría usar $\pi(\mathbf{p}_1, \dots, \mathbf{p}_r) = \pi(\mathbf{P})$, la opinión inicial sobre \mathbf{P} , para calcular π_0 promediando,

$$\pi_0 = \int_{B(\mathbf{P}_0, \delta)} \pi(\mathbf{P}) d\mathbf{P}.$$

Con este método, si se asignan distribuciones a priori uniformes sobre Θ e independientes a cada \mathbf{p}_i , $i = 1, \dots, r$, se obtiene

$$\pi_0 = \frac{\pi^{\frac{r(s-1)}{2}} \delta^{r(s-1)}}{\Gamma\left(\frac{r(s-1)}{2} + 1\right)},$$

el volumen de la esfera de radio δ en $\mathbb{R}^{r(s-1)}$, para δ suficientemente pequeño.

Por otra parte, desde un punto de vista clásico, si se utiliza el estadístico χ^2 de Pearson como *estadístico de contraste*, es decir

$$\Lambda = \sum_{i=1}^r \sum_{j=1}^s \frac{n_{ij}^2}{n_i p_{0j}} - N,$$

y λ_0 denota el valor de Λ en el punto formado por los datos observados de la Tabla 1.5, es decir, $\Lambda(n_{ij0}, i = 1, \dots, r, j = 1, \dots, s) = \lambda_0$, entonces $\{\Lambda \geq \lambda_0\}$ es una posible *región crítica* y

$$p = P(\Lambda \geq \lambda_0 | \mathbf{p}_0) = P(\chi_{r(s-1)}^2 \geq \lambda_0).$$

es el p-valor.

Por consiguiente, es obvio que, para buscar una reconciliación entre ambos resultados, el clásico y el bayesiano, se puede seguir un razonamiento análogo al desarrollado en la sección 2.3, dado que la expresión (2.7) es similar a la expresión (2.1).

En conclusión, fijados n_i , $i = 1, \dots, r$ y p^* , si se denotan mediante

$$\ell_1 = \ell_1(p^*, n_1, \dots, n_r) = \max_{(n_{ij}), p > p^*} \eta(\eta + 1)^{-1}, \quad (2.8)$$

$$\ell_2 = \ell_2(p^*, n_1, \dots, n_r) = \min_{(n_{ij}), p \leq p^*} \eta(\eta + 1)^{-1}, \quad (2.9)$$

y p^* verifica que $\ell_1 < \ell_2$, entonces existe un intervalo, $I = I(p^*, n_1, \dots, n_r) = (\ell_1, \ell_2)$, de valores de π_0 , tal que, si $\pi_0 \in I$, el resultado que se obtiene con el método bayesiano propuesto para el contraste (2.6), utilizando un valor de δ tal que $\pi_0(\delta) \in I$, es el mismo que se obtiene mediante el método clásico χ^2 de Pearson. Por lo tanto, el acuerdo es posible en los términos de dicha condición suficiente.

2.5. Conclusiones

La probabilidad a posteriori de la hipótesis nula de homogeneidad de poblaciones multinomiales independientes en tablas $r \times s$, cuando \mathbf{p}_0 es conocido, para una distribución a priori de tipo mixto, que asigna una probabilidad inicial π_0 a la hipótesis nula mediante (1.5) y reparte de forma continua la probabilidad restante en los puntos de la alternativa mediante una densidad a priori Dirichlet, admite la siguiente representación:

$$P(H_0 | n_{11}, \dots, n_{rs}) = \left[1 + \frac{1 - \pi_0}{\pi_0} \eta \right]^{-1},$$

donde η es un estadístico que cuantifica la fuerza de la evidencia a favor de la hipótesis más verosímil, $\Lambda = h(\eta)$ es el *estadístico de contraste* para el método clásico χ^2 de Pearson, y $h : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ es una función no monótona de *tendencia creciente*.

Fijados $n_i \in \mathbb{N}$, $i = 1, \dots, r$ y $p^* \in (0, 1)$, $\ell_1 < \ell_2$, definidos en las expresiones (2.8) y (2.9), proporciona una condición suficiente para que la reconciliación entre ambos métodos

sea posible. Es decir, si p^* verifica que $\ell_1 < \ell_2$, entonces para δ tal que $\pi_0 = \pi_0(\delta, p^*) \in (\ell_1, \ell_2)$ se satisface uno y sólo uno de los dos postulados siguientes:

$$\text{“}p > p^* \text{ y además } \left[1 + \frac{1 - \pi_0}{\pi_0} \eta\right]^{-1} > \frac{1}{2}\text{”},$$

$$\text{“}p \leq p^* \text{ y además } \left[1 + \frac{1 - \pi_0}{\pi_0} \eta\right]^{-1} \leq \frac{1}{2}\text{”},$$

cualquiera que sea $(n_{ij0}, i = 1, \dots, r, j = 1, \dots, s)$, el punto formado por los datos observados de la Tabla 2.4. Por lo tanto, para contrastar (2.6), si $\ell_1 < \ell_2$, un clásico que utilizara p^* para cuantificar el p-valor a la hora de tomar una decisión, llegaría a la misma conclusión, en los términos de aceptar o rechazar, que un bayesiano que utilizara como masa a priori de la hipótesis nula un valor $\pi_0 \in (\ell_1, \ell_2)$.

La existencia de valores p^* que satisfagan dicha condición suficiente depende de la relación funcional, en términos de h , existente entre los estadísticos Λ y η . Por lo tanto, la reconciliación entre ambos métodos es posible en ese sentido.

Por ejemplo, en tablas 2×2 con $n_1 = 18$ y $n_2 = 12$, para contrastar (1.2) con $p_0 = \frac{1}{2}$ y $\pi(p_1, p_2) = I_{(0, 1)}(p_1) I_{(0, 1)}(p_2)$, cuando $p^* = 0,1$ el acuerdo se consigue para $\delta \in (0,353, 0,4)$ (véase la Tabla 2.3).

En las secciones 2.6 y 2.7, siguiendo un razonamiento similar al desarrollado en el caso \mathbf{p}_0 conocido, se generalizan los resultados anteriores para el problema del contraste de homogeneidad de poblaciones multinomiales independientes cuando \mathbf{p}_0 es desconocido o de forma funcional conocida, $\mathbf{p}_0 = \mathbf{p}(\omega)$, respectivamente.

2.6. Tablas $r \times s$ con \mathbf{p}_0 Desconocido

En los mismos supuestos de la sección 2.4, se pretende contrastar (2.6) con \mathbf{p}_0 desconocido. Por lo tanto, las hipótesis de interés son

$$H_0 : \mathbf{p}_1 = \dots = \mathbf{p}_r, \text{ versus } H_1 : \exists i \neq j, \mathbf{p}_i \neq \mathbf{p}_j. \quad (2.10)$$

En esta sección se desarrollan tres métodos bayesianos que permiten resolver este problema de contraste, mediante los cuales, se puede llegar a una reconciliación con el método clásico usual, en la línea que se ha explicado previamente.

2.6.1. Primer Método

Sea $\mathbf{p}_0 = (p_{01}, \dots, p_{0s}) \in \Theta$ el valor común desconocido de la hipótesis nula. Si la opinión a priori sobre \mathbf{p}_0 viene dada por la densidad $\pi(\mathbf{p}_0)$, para contrastar (2.10) se necesita una distribución a priori de tipo mixto. Siguiendo un razonamiento similar al desarrollado en las secciones anteriores, se propone

$$\pi^*(\mathbf{p}_1, \dots, \mathbf{p}_r, \mathbf{p}_0) = \pi_0 \pi(\mathbf{p}_0) I_{H_0}(\mathbf{p}_1, \dots, \mathbf{p}_r, \mathbf{p}_0) + (1 - \pi_0) \pi(\mathbf{p}_1, \dots, \mathbf{p}_r) I_{H_1}(\mathbf{p}_1, \dots, \mathbf{p}_r, \mathbf{p}_0),$$

siendo π_0 la masa a priori asignada a $H_0 : \mathbf{p}_1 = \dots = \mathbf{p}_r$, y $\pi(\mathbf{p}_1, \dots, \mathbf{p}_r) = \prod_{i=1}^r \pi(\mathbf{p}_i)$ la opinión inicial sobre $(\mathbf{p}_1, \dots, \mathbf{p}_r)$.

En este caso, la probabilidad a posteriori de la hipótesis nula de (2.10), observados los datos de la Tabla 2.4, es

$$\frac{\pi_0 \int_{\Theta} \prod_{j=1}^s p_{0j}^{\sum_{i=1}^r n_{ij}} \pi(\mathbf{p}_0) d\mathbf{p}_0}{\pi_0 \int_{\Theta} \prod_{j=1}^s p_{0j}^{\sum_{i=1}^r n_{ij}} \pi(\mathbf{p}_0) d\mathbf{p}_0 + (1 - \pi_0) \prod_{i=1}^r \int_{\Theta} \prod_{j=1}^s p_{ij}^{n_{ij}} \pi(\mathbf{p}_i) d\mathbf{p}_i}.$$

Sean, $\alpha_i = (\alpha_{i1}, \dots, \alpha_{is})$, con $\alpha_{ij} > 0$, para cada $j = 1, \dots, s$ e $i = 0, 1, \dots, r$, y supongamos que $\pi(\mathbf{p}_i)$ es la densidad de una distribución de Dirichlet de parámetro α_i , $D(\alpha_i)$, $i = 0, 1, \dots, r$. Entonces,

$$\begin{aligned} \int_{\Theta} \prod_{j=1}^s p_{0j}^{\sum_{i=1}^r n_{ij}} \pi(\mathbf{p}_0) d\mathbf{p}_0 &= \frac{\Gamma\left(\sum_{j=1}^s \alpha_{0j}\right)}{\prod_{j=1}^s \Gamma(\alpha_{0j})} \int_{\Theta} \prod_{j=1}^s p_{0j}^{m_j + \alpha_{0j} - 1} d\mathbf{p}_0 \\ &= \frac{\Gamma\left(\sum_{j=1}^s \alpha_{0j}\right) \prod_{j=1}^s \Gamma(m_j + \alpha_{0j})}{\prod_{j=1}^s \Gamma(\alpha_{0j}) \Gamma\left(N + \sum_{j=1}^s \alpha_{0j}\right)}. \end{aligned}$$

Por lo tanto, dicha probabilidad a posteriori se puede expresar de la siguiente forma:

$$B_1(\pi_0) = \left[1 + \frac{1 - \pi_0}{\pi_0} \eta_1 \right]^{-1} \quad (2.11)$$

$$\text{donde } \eta_1 = \frac{\prod_{j=1}^s \Gamma(\alpha_{0j}) \Gamma(N + \sum_{j=1}^s \alpha_{0j})}{\Gamma\left(\sum_{j=1}^s \alpha_{0j}\right) \prod_{j=1}^s \Gamma(m_j + \alpha_{0j})} \left\{ \frac{\prod_{i=1}^r \Gamma\left(\sum_{j=1}^s \alpha_{ij}\right)}{\prod_{i=1}^r \prod_{j=1}^s \Gamma(\alpha_{ij})} \right\} \left\{ \frac{\prod_{i=1}^r \prod_{j=1}^s \Gamma(n_{ij} + \alpha_{ij})}{\prod_{i=1}^r \Gamma\left(n_i + \sum_{j=1}^s \alpha_{ij}\right)} \right\}.$$

2.6.2. Segundo Método

Sea $\mathbf{p}_0 = (p_{01}, \dots, p_{0s}) \in \Theta$ el valor común desconocido de la hipótesis nula. Para cada $\mathbf{p}_0 \in \Theta$, se considera el siguiente contraste auxiliar:

$$H_0^{\mathbf{p}_0} : \mathbf{p}_1 = \dots = \mathbf{p}_r = \mathbf{p}_0, \text{ versus } H_1^{\mathbf{p}_0} : \exists i \neq j, \text{ con } \mathbf{p}_i \neq \mathbf{p}_j.$$

Observados los datos de la Tabla 2.4, con una distribución a priori adecuada, la probabilidad a posteriori de $H_0^{\mathbf{p}_0} : \mathbf{p}_1 = \dots = \mathbf{p}_r = \mathbf{p}_0$ se puede expresar de la siguiente forma:

$$\left[1 + \frac{1 - \pi_0}{\pi_0} \eta_2 \right]^{-1},$$

donde $\eta_2 = \eta_2(\mathbf{p}_0) = \prod_{j=1}^s p_{0j}^{-m_j} \left\{ \frac{\prod_{i=1}^r \Gamma\left(\sum_{j=1}^s \alpha_{ij}\right)}{\prod_{i=1}^r \prod_{j=1}^s \Gamma(\alpha_{ij})} \right\} \left\{ \frac{\prod_{i=1}^r \prod_{j=1}^s \Gamma(n_{ij} + \alpha_{ij})}{\prod_{i=1}^r \Gamma\left(n_i + \sum_{j=1}^s \alpha_{ij}\right)} \right\}$ y $\pi_0 = \pi_0^{\mathbf{p}_0}$ es la masa a priori asignada a $H_0^{\mathbf{p}_0} : \mathbf{p}_1 = \dots = \mathbf{p}_r = \mathbf{p}_0$.

Si la opinión a priori sobre \mathbf{p}_0 viene dada por la densidad $\pi(\mathbf{p}_0)$, se puede considerar la siguiente medida cuantitativa del grado en el que r poblaciones son homogéneas en el contraste (2.10):

$$B_2 = \int_{\Theta} \left[1 + \frac{1 - \pi_0^{\mathbf{p}_0}}{\pi_0^{\mathbf{p}_0}} \eta_2(\mathbf{p}_0) \right]^{-1} \pi(\mathbf{p}_0) d\mathbf{p}_0.$$

Cabe destacar que B_2 es la probabilidad a posteriori de $H_0 : \mathbf{p}_1 = \dots = \mathbf{p}_r$, obtenida a partir de la distribución a priori

$$\pi^*(\mathbf{p}_1, \dots, \mathbf{p}_r, \mathbf{p}_0) = \pi^*(\mathbf{p}_1, \dots, \mathbf{p}_r | \mathbf{p}_0) \pi(\mathbf{p}_0),$$

siendo $\pi^*(\mathbf{p}_1, \dots, \mathbf{p}_r | \mathbf{p}_0) = \pi_0^{\mathbf{p}_0} I_{H_0}(\mathbf{p}_1, \dots, \mathbf{p}_r, \mathbf{p}_0) + (1 - \pi_0^{\mathbf{p}_0}) \pi(\mathbf{p}_1, \dots, \mathbf{p}_r) I_{H_1}(\mathbf{p}_1, \dots, \mathbf{p}_r, \mathbf{p}_0)$.

Por lo tanto, desde este punto de vista bayesiano, se acepta $H_0 : \mathbf{p}_1 = \dots = \mathbf{p}_r$ cuando $B_2 > \frac{1}{2}$.

Si para todo $\mathbf{p}_0 \in \Theta$, la probabilidad a priori asignada a $H_0^{\mathbf{p}_0} : \mathbf{p}_1 = \dots = \mathbf{p}_r = \mathbf{p}_0$, $\pi_0^{\mathbf{p}_0}$, es constante e igual a un valor π_0 , entonces π_0 se puede interpretar como la probabilidad a priori de $H_0 : \mathbf{p}_1 = \dots = \mathbf{p}_r$, en cuyo caso,

$$B_2(\pi_0) = \int_{\Theta} \left[1 + \frac{1 - \pi_0}{\pi_0} \eta_2(\mathbf{p}_0) \right]^{-1} \pi(\mathbf{p}_0) d\mathbf{p}_0. \quad (2.12)$$

2.6.3. Tercer Método

En el mismo contexto en el que se ha desarrollado el segundo método, la idea es considerar, como medida cuantitativa del grado en el que r poblaciones son homogéneas en el contraste (2.10), el valor del supremo de $P(H_0^{\mathbf{p}_0} | n_{11}, \dots, n_{rs}, \mathbf{p}_0)$, cuando \mathbf{p}_0 varía en Θ . Como el valor donde se alcanza el ínfimo de $\prod_{j=1}^s p_{0j}^{-m_j}$ cuando \mathbf{p}_0 varía en Θ , $\hat{\mathbf{p}}_0$, se obtiene para $\hat{p}_{0j} = \frac{m_j}{N}$, $j = 1, \dots, s$, dicha medida es

$$B_3(\pi_0) = \left[1 + \frac{1 - \pi_0}{\pi_0} \eta_3 \right]^{-1}, \quad (2.13)$$

$$\text{donde } \eta_3 = N^N \prod_{j=1}^s m_j^{-m_j} \left\{ \frac{\prod_{i=1}^r \Gamma(\sum_{j=1}^s \alpha_{ij})}{\prod_{i=1}^r \prod_{j=1}^s \Gamma(\alpha_{ij})} \right\} \left\{ \frac{\prod_{i=1}^r \prod_{j=1}^s \Gamma(n_{ij} + \alpha_{ij})}{\prod_{i=1}^r \Gamma(n_i + \sum_{j=1}^s \alpha_{ij})} \right\}.$$

Por lo tanto, con este método, se rechaza $H_0 : \mathbf{p}_1 = \dots = \mathbf{p}_r$ cuando $B_3 < \frac{1}{2}$.

2.6.4. Comparación con el Método Clásico

Desde un punto de vista clásico, si se considera el estadístico χ^2 de Pearson como *estadístico de contraste*,

$$\Lambda = N \left(\sum_{i=1}^r \sum_{j=1}^s \frac{n_{ij}^2}{n_i m_j} - 1 \right),$$

y λ_0 denota el valor de Λ en el punto formado por los datos observados en la Tabla 2.4, es decir, $\Lambda(n_{ij0}, i = 1, \dots, r, j = 1, \dots, s) = \lambda_0$, entonces, $\{\Lambda \geq \lambda_0\}$ es una posible *región crítica*, y

$$p = P(\Lambda \geq \lambda_0 | \mathbf{p}_1 = \dots = \mathbf{p}_r) = P(\chi_{(r-1)(s-1)}^2 \geq \lambda_0)$$

es el p-valor.

Se puede observar que las medidas bayesianas dadas, respectivamente, en las expresiones (2.11) y (2.13) tienen la misma forma funcional que (2.7). Por lo tanto, siguiendo un razonamiento similar al desarrollado en las secciones 2.3 y 2.4, es posible una reconciliación entre los resultados clásicos y bayesianos, para cada uno de los tres métodos bayesianos expuestos anteriormente, en los términos de una condición suficiente.

Para ilustrar el procedimiento, se va a considerar el caso más sencillo, en el que se pretende contrastar

$$H_0 : p_1 = p_2, \text{ versus } H_1 : p_1 \neq p_2. \quad (2.14)$$

Además, se supone que p_0 , p_1 y p_2 tienen, respectivamente, distribuciones uniformes sobre el intervalo unidad. En este caso, las medidas cuantitativas del grado en el que las dos poblaciones binomiales son homogéneas para los tres métodos bayesianos propuestos, dadas, respectivamente, en las expresiones (2.11), (2.12) y (2.13), se obtienen evaluando dichas expresiones en

$$\begin{aligned} \eta_1 &= \frac{\Gamma(N+2)}{\Gamma(m_1+1)\Gamma(m_2+1)} \frac{\Gamma(a+1)\Gamma(b+1)\Gamma(c+1)\Gamma(d+1)}{\Gamma(a+b+2)} \frac{\Gamma(c+1)\Gamma(d+1)}{\Gamma(c+d+2)}, \\ \eta_2 &= \eta_2(p_0) = p_0^{-m_1} (1-p_0)^{-m_2} \frac{\Gamma(a+1)\Gamma(b+1)\Gamma(c+1)\Gamma(d+1)}{\Gamma(a+b+2)} \frac{\Gamma(c+1)\Gamma(d+1)}{\Gamma(c+d+2)}, \\ \eta_3 &= N^N m_1^{-m_1} m_2^{-m_2} \frac{\Gamma(a+1)\Gamma(b+1)\Gamma(c+1)\Gamma(d+1)}{\Gamma(a+b+2)} \frac{\Gamma(c+1)\Gamma(d+1)}{\Gamma(c+d+2)}. \end{aligned}$$

La probabilidad a posteriori calculada mediante los tres métodos bayesianos propuestos depende de π_0 , la masa a priori inicial asignada a la hipótesis nula $H_0 : p_1 = p_2$.

Sea $\pi_0^k = \pi_0^k(\delta_k)$ el valor de π_0 que verifica que $B_k(\pi_0) > \frac{1}{2}$ si y sólo si $\pi_0 > \pi_0^k$, $k=1,2,3$.

Fijado un punto $(p_0, p_0) \in H_0$, sea

$$B((p_0, p_0), \delta) = \{(p_1, p_2) \in (0, 1) \times (0, 1), (p_1 - p_0)^2 + (p_2 - p_0)^2 \leq \delta^2\},$$

la bola de centro (p_0, p_0) y radio δ . Entonces, se podría utilizar $\pi(p_1, p_2)$, la opinión a priori sobre (p_1, p_2) , y calcular π_0 mediante

$$\pi_0 = \int_{C(\delta)} \pi(p_1, p_2) dp_2 dp_1, \quad (2.15)$$

siendo $C(\delta) = \bigcup_{p_0 \in (0, 1)} B((p_0, p_0), \delta) = \{(p_1, p_2) \in (0, 1) \times (0, 1), p_1 - 2\delta^2 < p_2 < p_1 + 2\delta^2\}$.

Se puede observar que si la opinión a priori sobre (p_1, p_2) es uniforme, $\pi(p_1, p_2) = I_{(0, 1)}(p_1) I_{(0, 1)}(p_2)$, entonces el valor de π_0 que se obtiene a partir de la expresión (2.15), para δ suficientemente pequeño, es $\pi_0 = 2\sqrt{2}\delta(1 - 2\delta^2) + 2\delta^2 = 2\sqrt{2}\delta + 2\delta^2 - 4\sqrt{2}\delta^3$.

Por otro lado, para contrastar (2.14) desde un punto de vista clásico, el *estadístico de contraste* usual obtenido por el método χ^2 de Pearson, evaluado en los datos de la Tabla 1.1, es el cuadrado del estadístico de Yule, es decir,

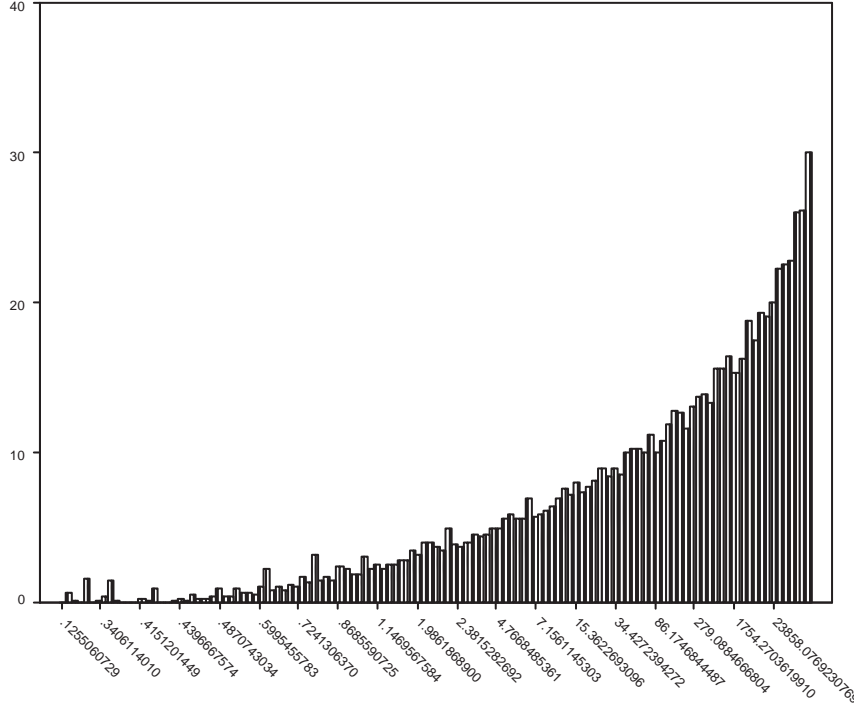
$$\Lambda = \{ad - bc\}^2 \frac{N}{n_1 n_2 m_1 m_2}.$$

En este caso, si observado un punto (a_0, c_0) , λ_0 denota el valor del estadístico de Yule en dicho punto, es decir $\Lambda(a_0, c_0) = \lambda_0$, entonces la evidencia utilizada es el p-valor,

$$p = P(\Lambda \geq \lambda_0 | p_1 = p_2) = P(\chi_1^2 \geq \lambda_0).$$

Cabe destacar que los estadísticos Λ , η_1 y η_3 no son suficientes. Además, se puede comprobar que los métodos bayesianos basados, respectivamente, en las medidas B_1 y B_3 no se pueden expresar en términos de Λ . No obstante, existen funciones $h_k : \mathbb{R}^+ \rightarrow \mathbb{R}^+$, $k=1, 3$, no monótonas, tales que $\Lambda = h_k(\eta_k)$, $k=1, 3$, (véanse las Figuras 2.2 y 2.3), con lo cual, el test clásico admite sendas representaciones en términos de η_1 y η_3 . También se puede observar que la función h_3 presenta irregularidades más pronunciadas que h_1 .

Figura 2.2: Diagrama de Barras $(\eta_1(a, c), \Lambda(a, c))$, para tablas 2×2 con $n_1 = 18$ y $n_2 = 12$. Existe una función $h_1 : \mathbb{R}^+ \rightarrow \mathbb{R}^+$, no monótona, tal que $\Lambda = h_1(\eta_1)$.



Sean n_1 , n_2 y p^* fijos y conocidos. Para comparar con el método clásico los resultados obtenidos con los tres métodos bayesianos desarrollados en esta sección, siguiendo el mismo razonamiento empleado en la sección 2.3, se puede calcular el valor de π_0 que verifica la ecuación

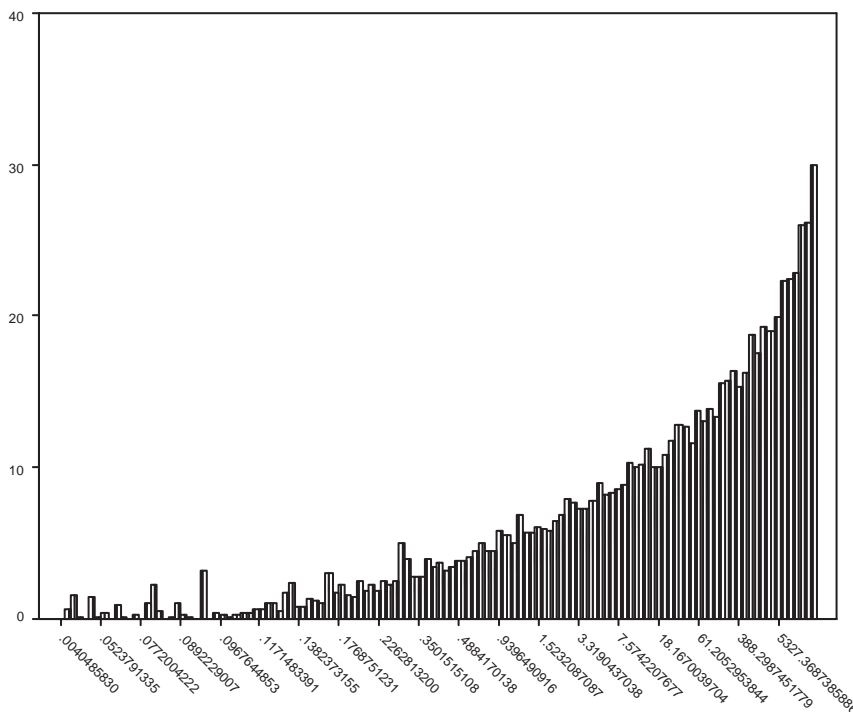
$$B_k(\pi_0) = \frac{p}{2p^*}, \quad k = 1, 2, 3. \quad (2.16)$$

En este caso, si dicho valor se denota por π_{k0}^* , $k = 1, 2, 3$, se obtiene

$$\pi_{k0}^* = \frac{\eta_k p}{\eta_k p + 2p^* - p}, \quad k = 1, 3,$$

mientras que para el segundo método habría que calcular su valor numéricamente, ya que no se puede obtener su expresión explícita. Entonces, utilizando π_{k0}^* se llegaría con el k -ésimo método Bayesiano, $k = 1, 2, 3$, a la misma conclusión que con el método clásico.

Figura 2.3: Diagrama de Barras $(\eta_3(a, c), \Lambda(a, c))$, para tablas 2×2 con $n_1 = 18$ y $n_2 = 12$. Existe una función $h_3 : \mathbb{R}^+ \rightarrow \mathbb{R}^+$, no monótona, tal que $\Lambda = h_3(\eta_3)$.



Para los datos del ejemplo de Pearson, los resultados obtenidos con el método clásico y los tres métodos bayesianos propuestos, se resumen en la Tabla 2.5. Con el método clásico se obtiene $\Lambda = 5,625$ y un p-valor $p = 0,017706$. Por ejemplo, para el primero de los métodos bayesianos propuestos, $\eta_1 = 5,8347$ y si $\pi_0 = \frac{1}{2}$, entonces $B_1 = 0,1463$. Además, para aceptar $H_0 : p_1 = p_2$, debe ser $\pi_0 > 0,8537$, o equivalentemente $\delta > 0,2913$. Si $p^* = 0,5$, el valor de π_0 que reconcilia dicho p-valor con la probabilidad a posteriori es $\pi_{10}^* = 0,09516$, o alternativamente $\delta_1^* = 0,03294$. Si $p^* = 0,1$ se obtiene $\pi_{10}^* = 0,3617$ con $\delta_1^* = 0,1211$ y se rechaza con una probabilidad a posteriori de 0,08853. Si $p^* = 0,05$ se obtiene $\pi_{10}^* = 0,5566$ con $\delta_1^* = 0,1852$ y se rechaza con 0,17706. Para $p^* = 0,01$ se obtiene $\pi_{10}^* = 0,9782$ con $\delta_1^* = 0,3435$ y se acepta con 0,8853. Además, para cada uno de los tres métodos bayesianos propuestos, al igual que ocurría con \mathbf{p}_0 conocido, se puede observar cómo el valor de π_0 , y por lo tanto el valor de δ , para el que se consigue acuerdo entre el

resultado clásico y bayesiano, disminuye cuando p^* aumenta.

También se aprecia que el tercer método bayesiano es el más conservador con respecto a la hipótesis nula del contraste (2.15), mientras que el segundo método es el que rechaza antes. Por ejemplo, si $\pi_0 = \frac{1}{2}$, aunque los tres métodos rechazan $H_0 : p_1 = p_2$, $B_1 = 0,1463$, $B_2 = 0,110919$ y $B_3 = 0,4484$. En consecuencia, como es posible observar en la Tabla 2.5, el valor de π_0 , o equivalentemente el valor de δ , para el que se alcanza el acuerdo con el tercer método es más pequeño que el correspondiente al de los métodos primero y segundo. De todas formas, como los valores de δ que se obtienen con el segundo método están muy próximos a los obtenidos con el primero y la dificultad de cálculo es considerablemente mayor, se propone utilizar el primer método bayesiano o, desde un punto de vista más conservador, el tercero.

Tabla 2.5: resumen de resultados para los datos del ejemplo de Pearson,

$$p_0 \text{ desconocido y } \pi(p_1, p_2) = I_{(0, 1)}(p_1) I_{(0, 1)}(p_2).$$

Tabla 2.5.1: método clásico

Λ	p
5,625	0,017706

Tabla 2.5.2: métodos Bayesianos

Tabla 2.5.2.1:

Tabla 2.1	η_k	$B_k\left(\frac{1}{2}\right)$	π_0^k	δ_k
Método 1	5,8347	0,1463	0,8537	0,2913
Método 2		0,110919	0,995416	0,3514
Método 3	1,2301	0,4484	0,5516	0,1836

Tabla 2.5.2.2:

π_{k0}^* δ_k^* $p(2p^*)^{-1}$	$p^* = 0,5$	$p^* = 0,1$	$p^* = 0,05$	$p^* = 0,01$
Método 1	0,09516 0,03294 0,017706	0,3617 0,1211 0,08853	0,5566 0,1852 0,17706	0,9782 0,3435 0,8853
Método 2	0,098975 0,034244 0,017706	0,41969 0,14001 0,08853	0,690015 0,23088 0,17706	0,9 ¹¹ 882 0,35355339 0,8853
Método 3	0,02169 0,00763 0,017706	0,1067 0,03686 0,08853	0,2093 0,07114 0,17706	0,9047 0,3117 0,8853

No obstante, el acuerdo obtenido en los términos de la expresión (2.16) es demasiado estricto, ya que π_{k0}^* , $k = 1, 2, 3$, depende de los datos, sin embargo da una idea de cómo debería ser el valor de δ con el que se obtenga una reconciliación entre el método clásico y cada uno de los tres métodos bayesianos. En la Tabla 2.5 se observa que $\delta_1^* \in (0,1211, 0,3435)$, $\delta_2^* \in (0,14, 0,3536)$ y $\delta_3^* \in (0,0369, 0,3117)$, cuando $p^* \in (0,01, 0,5)$.

Para intentar eliminar la dependencia que dicha reconciliación tiene de los datos se han generado todas las tablas 2×2 que se pueden obtener para n_1 y n_2 fijos y conocidos. En el caso que nos ocupa las entradas son $n_1 = 18$ y $n_2 = 12$ y se generan las 247 tablas posibles. Para cada una de estas tablas se efectúa el mismo estudio que se ha realizado previamente para la tabla de Pearson.

Mediante un análisis de datos sencillo se puede comprobar que para contrastar (2.14) con $\pi(p_1, p_2) = I_{(0, 1)}(p_1) I_{(0, 1)}(p_2)$, no existe acuerdo con ninguno de los métodos bayesianos primero y tercero, cuando $p^* = 0,5$, $p^* = 0,1$, $p^* = 0,05$ o $p^* = 0,01$. Sin

embargo se pueden determinar valores de p^* para los que existe un intervalo de valores de $\pi_0 = \pi_0(\delta)$, $I = I(p^*, n_1 = 18, n_2 = 12)$, tal que el resultado que se obtiene con estos dos métodos bayesianos, utilizando un valor de $\pi_0 \in I$, es el mismo que se obtiene mediante el método clásico χ^2 de Pearson. Por ejemplo, el mayor valor de p^* para el que existe acuerdo con el primer método es $p^* = 0,0635$, mientras que con el tercero es $p^* = 0,008$. Esto se debe a que el tercer método bayesiano es el más conservador. Además, si $p^* \in (0,0635, 0,0637)$ o $p^* \in (0,008, 0,0085)$, utilizando, respectivamente, el primer método bayesiano con $\delta \in (0,2222, 0,223)$, que se corresponde con $\pi_0 \in (0,6651, 0,6675)$, y $\delta \in (0,3218, 0,3252)$, correspondiente a $\pi_0 \in (0,9288, 0,9368)$, se obtiene el mismo resultado que con el método clásico. Con el tercer método bayesiano ocurre lo mismo si $p^* \in (0,08, 0,0085)$ utilizando $\delta \in (0,2478, 0,2503)$, o equivalentemente un valor $\pi_0 \in (0,73769, 0,74455)$.

En general, fijados n_i , $i = 1, \dots, r$ y p^* , si se denota mediante

$$\ell_1 = \ell_1(p^*, n_1, \dots, n_r) = \max_{(n_{ij}), p > p^*} \pi_0^k,$$

$$\ell_2 = \ell_2(p^*, n_1, \dots, n_r) = \min_{(n_{ij}), p \leq p^*} \pi_0^k,$$

y p^* verifica que $\ell_1 < \ell_2$, entonces existe un intervalo, $I = I(p^*, n_1, \dots, n_r) = (\ell_1, \ell_2)$, de valores de $\pi_0 = \pi_0(\delta, p^*)$, tal que si $\pi_0 \in I$, el resultado que se obtiene con el k -ésimo método Bayesiano propuesto para el contraste (2.10), $k = 1, 2, 3$, utilizando un valor de $\pi_0 \in I$, es el mismo que se obtiene mediante el método clásico χ^2 de Pearson. Es decir, si p^* verifica que $\ell_1 < \ell_2$, entonces para δ tal que $\pi_0 = \pi_0(\delta, p^*) \in (\ell_1, \ell_2)$ se satisface uno y sólo uno de los dos postulados siguientes,

$$“p > p^* \text{ y además } P(H_0 | n_{110}, \dots, n_{rs0}) > \frac{1}{2}”,$$

$$“p \leq p^* \text{ y además } P(H_0 | n_{110}, \dots, n_{rs0}) \leq \frac{1}{2}”,$$

cualquiera que sea $(n_{ij0}, i = 1, \dots, r, j = 1, \dots, s)$, el punto formado por los datos observados de la Tabla 2.4. Por lo tanto, el acuerdo es posible en este sentido.

2.7. Tablas $r \times s$ con $\mathbf{p}_0 = \mathbf{p}(\omega)$

Ahora, en los mismos supuestos de la sección 2.4, se pretende contrastar

$$H_0 : \mathbf{p}_1 = \cdots = \mathbf{p}_r = \mathbf{p}(\omega), \text{ versus } H_1 : \exists i \neq j, \text{ con } \mathbf{p}_i \neq \mathbf{p}_j \quad (2.17)$$

donde $\mathbf{p}: \Omega \rightarrow \Theta$, siendo $\Omega = \{ \omega = (\omega_1, \dots, \omega_q) , \mathbf{p}(\omega) = (p_1(\omega), \dots, p_s(\omega)) \in \Theta \} \subset \mathbb{R}^q$ y $q < s$ fijo. La hipótesis nula de (2.17) significa que la probabilidad de que un individuo seleccionado aleatoriamente pertenezca a la clase j es la misma para cada una de las i poblaciones, e igual a $p_{ij} = p_j(\omega_1, \dots, \omega_q)$, para cada $j = 1, \dots, s$, y para cada $i = 1, \dots, r$, es decir, las r poblaciones son homogéneas con una distribución de probabilidad especificada a falta de q parámetros desconocidos.

Tabla 2.6: datos en la tabla $r \times 3$.

	AA	Aa	Aa	Total
Muestra 1	a_1	b_1	c_1	n_1
Muestra 2	a_2	b_2	c_2	n_2
\vdots	\vdots	\vdots	\vdots	\vdots
Muestra r	a_r	b_r	c_r	n_r
Total	m_1	m_2	m_3	N

Por ejemplo, cuando se pretende estudiar si una o varias poblaciones se encuentran en equilibrio Hardy-Weinberg respecto a los tres genotipos posibles, AA, Aa y aa, la hipótesis nula es

$$H_0 : p_{i1} = p^2, p_{i2} = 2p(1-p), p_{i3} = (1-p)^2, i = 1, \dots, r,$$

para algún $p \in (0, 1)$. Este es un problema de contraste importante en genética. Para investigarlo, se seleccionan muestras aleatorias de individuos en cada una de las poblaciones

en estudio, contabilizándose el número de los que presentan, en cada una de ellas, cada uno los tres genotipos. Los datos se visualizan en la Tabla 2.6. Más adelante se estudia este problema con detalle.

A continuación se desarrollan tres métodos que permiten tratar el problema de contraste (2.17) desde una perspectiva bayesiana.

2.7.1. Primer Método

Si la densidad $\pi(\omega)$ describe la opinión que se tiene a priori sobre ω , para contrastar (2.17), se propone utilizar la siguiente distribución a priori de tipo mixto,

$$\pi^*(\mathbf{p}_1, \dots, \mathbf{p}_r, \omega) = \pi_0 \pi(\omega) I_{H_0}(\mathbf{p}_1, \dots, \mathbf{p}_r, \omega) + (1 - \pi_0) \pi(\mathbf{p}_1, \dots, \mathbf{p}_r) I_{H_1}(\mathbf{p}_1, \dots, \mathbf{p}_r, \omega),$$

siendo π_0 la probabilidad a priori asignada a $H_0 : \mathbf{p}_1 = \dots = \mathbf{p}_r = \mathbf{p}(\omega)$, y $\pi(\mathbf{p}_1, \dots, \mathbf{p}_r) = \prod_{i=1}^r \pi(\mathbf{p}_i)$ la opinión inicial sobre $(\mathbf{p}_1, \dots, \mathbf{p}_r)$.

Así, la probabilidad a posteriori de la hipótesis nula, observados los datos de la Tabla 2.4, es

$$\frac{\pi_0 \int_{\Omega} \prod_{j=1}^s p_j(\omega)^{\sum_{i=1}^r n_{ij}} \pi(\omega) d\omega}{\pi_0 \int_{\Omega} \prod_{j=1}^s p_j(\omega)^{\sum_{i=1}^r n_{ij}} \pi(\omega) d\omega + (1 - \pi_0) \prod_{i=1}^r \int_{\Theta} \prod_{j=1}^s p_{ij}^{n_{ij}} \pi(\mathbf{p}_i) d\mathbf{p}_i}.$$

Sean, $\alpha_i = (\alpha_{i1}, \dots, \alpha_{is})$, con $\alpha_{ij} > 0$, para cada $j = 1, \dots, s$ e $i = 1, \dots, r$, y $\pi(\mathbf{p}_i)$ la densidad de una distribución de Dirichlet de parámetro α_i , $D(\alpha_i)$, $i = 1, \dots, r$. En este caso, dicha probabilidad a posteriori se puede expresar de la siguiente forma:

$$B_1(\pi_0) = \left[1 + \frac{1 - \pi_0}{\pi_0} \eta_1 \right]^{-1}, \quad (2.18)$$

$$\text{donde } \eta_1 = \left[\int_{\Omega} \prod_{j=1}^s p_j(\omega)^{m_j} \pi(\omega) d\omega \right]^{-1} \left\{ \frac{\prod_{i=1}^r \Gamma(\sum_{j=1}^s \alpha_{ij})}{\prod_{i=1}^r \prod_{j=1}^s \Gamma(\alpha_{ij})} \right\} \left\{ \frac{\prod_{i=1}^r \prod_{j=1}^s \Gamma(n_{ij} + \alpha_{ij})}{\prod_{i=1}^r \Gamma(n_i + \sum_{j=1}^s \alpha_{ij})} \right\}.$$

En el ejemplo de los tres genotipos mencionado anteriormente, como el parámetro p representa la proporción de alelos de tipo A en la población, una posibilidad para

contrastar si una o varias poblaciones se encuentran en equilibrio Hardy-Weinberg es asignar a \mathbf{p} una distribución inicial beta, es decir

$$\pi(\mathbf{p}) = \frac{\Gamma(\alpha_{01} + \alpha_{02})}{\Gamma(\alpha_{01})\Gamma(\alpha_{02})} \mathbf{p}^{\alpha_{01}-1} (1 - \mathbf{p})^{\alpha_{02}-1}, \quad \mathbf{p} \in (0, 1), \quad (\alpha_{01}, \alpha_{02} > 0).$$

En este caso,

$$\int_{\Omega} \prod_{j=1}^s p_j(\omega)^{m_j} \pi(\omega) d\omega = 2^{m_2} \frac{\Gamma(\alpha_{01} + \alpha_{02})}{\Gamma(\alpha_{01})\Gamma(\alpha_{02})} \frac{\Gamma(2m_1 + m_2 + \alpha_{01}) \Gamma(m_2 + 2m_3 + \alpha_{02})}{\Gamma(2N + \alpha_{01} + \alpha_{02})}.$$

2.7.2. Segundo Método

Sea $\omega_0 = (\omega_{01}, \dots, \omega_{0q}) \in \Omega$ el valor común desconocido de la hipótesis nula. Para cada $\omega_0 \in \Omega$, se propone el siguiente contraste auxiliar:

$$H_0^{\omega_0} : \mathbf{p}_1 = \dots = \mathbf{p}_r = \mathbf{p}(\omega_0), \text{ versus } H_1^{\omega_0} : \exists i \neq j, \text{ con } \mathbf{p}_i \neq \mathbf{p}_j. \quad (2.19)$$

Observados los datos de la Tabla 2.4, con una distribución a priori adecuada, la probabilidad a posteriori de $H_0^{\omega_0} : \mathbf{p}_1 = \dots = \mathbf{p}_r = \mathbf{p}(\omega_0)$, se puede expresar de la siguiente forma:

$$\left[1 + \frac{1 - \pi_0}{\pi_0} \eta_2 \right]^{-1},$$

donde $\eta_2 = \eta_2(\omega_0) = \prod_{j=1}^s p_j(\omega_0)^{-m_j} \left\{ \frac{\prod_{i=1}^r \Gamma(\sum_{j=1}^s \alpha_{ij})}{\prod_{i=1}^r \prod_{j=1}^s \Gamma(\alpha_{ij})} \right\} \left\{ \frac{\prod_{i=1}^r \prod_{j=1}^s \Gamma(n_{ij} + \alpha_{ij})}{\prod_{i=1}^r \Gamma(n_i + \sum_{j=1}^s \alpha_{ij})} \right\}$ y $\pi_0 = \pi_0^{\omega_0}$

es la masa a priori asignada a $H_0^{\omega_0}$.

Si la opinión a priori sobre ω_0 viene dada por la densidad $\pi(\omega_0)$, una medida cuantitativa bayesiana del grado en el que r poblaciones son homogéneas en el contraste (2.19) es

$$B_2 = \int_{\Omega} \left[1 + \frac{1 - \pi_0^{\omega_0}}{\pi_0^{\omega_0}} \eta_2(\omega_0) \right]^{-1} \pi(\omega_0) d\omega_0.$$

Se puede observar que B_2 es la probabilidad a posteriori de la hipótesis nula de (2.17), obtenida a partir de la distribución a priori

$$\pi^*(\mathbf{p}_1, \dots, \mathbf{p}_r, \omega) = \pi^*(\mathbf{p}_1, \dots, \mathbf{p}_r | \omega) \pi(\omega),$$

donde $\pi^*(\mathbf{p}_1, \dots, \mathbf{p}_r | \omega) = \pi_0^\omega I_{H_0}(\mathbf{p}_1, \dots, \mathbf{p}_r, \omega) + (1 - \pi_0^\omega) \pi(\mathbf{p}_1, \dots, \mathbf{p}_r) I_{H_1}(\mathbf{p}_1, \dots, \mathbf{p}_r, \omega)$.

Por lo tanto, desde este punto de vista bayesiano, se aceptaría $H_0 : \mathbf{p}_1 = \dots = \mathbf{p}_r = \mathbf{p}(\omega)$ cuando $B_2 > \frac{1}{2}$.

Si para todo $\omega_0 \in \Omega$, la masa a priori asignada a $H_0^{\omega_0} : \mathbf{p}_1 = \dots = \mathbf{p}_r = \mathbf{p}(\omega_0)$, $\pi_0^{\omega_0}$, fuera constante e igual a un valor π_0 , entonces, π_0 podría interpretarse como la probabilidad a priori de $H_0 : \mathbf{p}_1 = \dots = \mathbf{p}_r = \mathbf{p}(\omega)$. En este caso,

$$B_2(\pi_0) = \int_{\Omega} \left[1 + \frac{1 - \pi_0}{\pi_0} \eta_2(\omega_0) \right]^{-1} \pi(\omega_0) d\omega_0. \quad (2.20)$$

2.7.3. Tercer Método

En el mismo contexto en el que ha sido desarrollado el segundo método, la idea es considerar como medida cuantitativa del grado en el que r poblaciones son homogéneas en el contraste (2.17) el valor donde se alcanza el supremo de $P(H_0^{\omega_0} | n_{11}, \dots, n_{rs}, \omega_0)$, cuando ω_0 varía en Ω .

Si se denota por $\hat{\omega}_0$ el valor donde se alcanza el ínfimo de $\prod_{j=1}^s p_j(\omega_0)^{-m_j}$, cuando ω_0 varía en Ω , dicha medida es

$$B_3(\pi_0) = \left[1 + \frac{1 - \pi_0}{\pi_0} \eta_3 \right]^{-1}, \quad (2.21)$$

$$\text{donde } \eta_3 = \prod_{j=1}^s p_j(\hat{\omega}_0)^{-m_j} \left\{ \frac{\prod_{i=1}^r \Gamma(\sum_{j=1}^s \alpha_{ij})}{\prod_{i=1}^r \prod_{j=1}^s \Gamma(\alpha_{ij})} \right\} \left\{ \frac{\prod_{i=1}^r \prod_{j=1}^s \Gamma(n_{ij} + \alpha_{ij})}{\prod_{i=1}^r \Gamma(n_i + \sum_{j=1}^s \alpha_{ij})} \right\}.$$

2.7.4. Comparación con el Método Clásico

Desde un punto de vista clásico, si se considera el estadístico χ^2 de Pearson como *estadístico de contraste*,

$$\Lambda = \sum_{i=1}^r \sum_{j=1}^s \frac{n_{ij}^2}{n_i p_j(\hat{\omega})} - N,$$

siendo $\hat{\omega}$ el *estimador de máxima verosimilitud* de ω , y λ_0 el valor de Λ en el punto formado por los datos observados en la Tabla 2.4, es decir, $\Lambda(n_{ij0}, i = 1, \dots, r, j = 1, \dots, s) = \lambda_0$, entonces $\{\Lambda \geq \lambda_0\}$ es una posible *región crítica* y la evidencia utilizada es el p-valor,

$$p = P(\Lambda \geq \lambda_0 | \mathbf{p}_1 = \dots = \mathbf{p}_r = \mathbf{p}(\omega)) = P(\chi_{rs-1-q}^2 \geq \lambda_0).$$

A continuación se va a comprobar que, con la metodología expuesta anteriormente y siguiendo un razonamiento similar al desarrollado en las secciones 2.3, 2.4 y 2.6, se puede concretar cómo y cuándo es posible una reconciliación entre los resultados clásicos y bayesianos.

Para ilustrar el procedimiento, con el método clásico usual y los tres métodos bayesianos desarrollados anteriormente, se va a considerar que se pretende contrastar si una o varias poblaciones se encuentran en equilibrio Hardy-Weinberg. En este caso, con el tercer método, la medida cuantitativa del grado en el que dichas poblaciones son homogéneas se obtiene evaluando la expresión (2.21) en $\hat{\omega}_0 = \hat{p} = \frac{2m_1+m_2}{2N}$, la proporción de alelos de tipo A en la muestra. Además, para los datos observados en la Tabla 2.6, si se asignan distribuciones a priori uniformes e independientes a cada $\mathbf{p}_{i1} = (p_{i1}, p_{i2}, p_{i3})$, $p_{i3} = 1 - p_{i1} - p_{i2}$, $i = 1, \dots, r$, se obtiene

$$\eta_3 = 2^{-m_2} \hat{p}^{-2m_1-m_2} (1 - \hat{p})^{-m_2-2m_3} \prod_{i=1}^r \frac{\Gamma(a_i + 1) \Gamma(b_i + 1) \Gamma(c_i + 1)}{\Gamma(a_i + b_i + c_i + 3)}.$$

Además, en los supuestos anteriores, si también se asigna a la proporción de alelos de tipo A una distribución a priori de Laplace, entonces las medidas cuantitativas del grado en el que r poblaciones son homogéneas para los métodos primero y segundo se obtienen, respectivamente, evaluando las expresiones (2.18) y (2.20) en

$$\eta_1 = 2^{-m_2} \frac{\Gamma(2N + 2)}{\Gamma(2m_1 + m_2 + 1) \Gamma(m_2 + 2m_3 + 1)} \prod_{i=1}^3 \frac{\Gamma(a_i + 1) \Gamma(b_i + 1) \Gamma(c_i + 1)}{\Gamma(n_i + 3)},$$

$$\eta_2 = \eta_2(\pi_0) = 2^{-m_2} p_0^{-2m_1-m_2} (1 - p_0)^{-m_2-2m_3} \prod_{i=1}^3 \frac{\Gamma(a_i + 1) \Gamma(b_i + 1) \Gamma(c_i + 1)}{\Gamma(a_i + b_i + c_i + 3)}.$$

Por lo tanto,

$$B_k(\pi_0) = \left[1 + \frac{1 - \pi_0}{\pi_0} \eta_i \right]^{-1}, k = 1, 3,$$

$$B_2(\pi_0) = \int_0^1 \left[1 + \frac{1 - \pi_0}{\pi_0} \eta_2(p_0) \right]^{-1} dp_0.$$

En particular, si $r = 1$ se obtiene

$$\eta_1 = 2^{-b_1} \frac{\Gamma(2n_1 + 2)}{\Gamma(2a_1 + b_1 + 1) \Gamma(b_1 + 2c_1 + 1)} \frac{\Gamma(a_1 + 1) \Gamma(b_1 + 1) \Gamma(c_1 + 1)}{\Gamma(n_1 + 3)},$$

$$\eta_2 = \eta_2(\pi_0) = 2^{-b_1} p_0^{-2a_1 - b_1} (1 - p_0)^{-b_1 - 2c_1} \frac{\Gamma(a_1 + 1) \Gamma(b_1 + 1) \Gamma(c_1 + 1)}{\Gamma(n_1 + 3)},$$

$$\eta_3 = 2^{-b_1} \hat{p}^{-2a_1 - b_1} (1 - \hat{p})^{-b_1 - 2c_1} \frac{\Gamma(a_1 + 1) \Gamma(b_1 + 1) \Gamma(c_1 + 1)}{\Gamma(n_1 + 3)},$$

donde $\hat{p} = \frac{2a_1 + b_1}{2n_1}$.

La probabilidad a posteriori calculada mediante los tres métodos bayesianos propuestos depende de π_0 , la masa a priori asignada a la hipótesis nula $H_0 : p^2, 2p(1 - p), (1 - p)^2$.

Sea π_0^k el valor de π_0 que verifica que $B_k(\pi_0) > \frac{1}{2}$ si y sólo si $\pi_0 > \pi_0^k$, $k = 1, 2, 3$, siendo B_k , $k = 1, 2, 3$, las medidas dadas respectivamente en las expresiones (2.18), (2.20) y (2.21).

Se puede observar que $\pi_0^k = \eta_k(\eta_k + 1)^{-1}$, $k = 1, 3$.

Tabla 2.7: ejemplo de Lindley.

	AA	Aa	Aa	Total
Muestra 1	31	38	31	100
Muestra 2	6	22	72	100
Muestra 3	2	6	92	100
Muestra 4	1	8	91	100
Total	40	74	286	400

Desde el punto de vista clásico, el estadístico χ^2 de Pearson para una tabla $r \times 3$ es

$$\Lambda = \sum_{i=1}^r \left\{ \frac{a_i^2}{n_i \hat{p}^2} + \frac{b_i^2}{n_i 2\hat{p}(1 - n_i \hat{p})} + \frac{c_i^2}{n_i (1 - \hat{p})^2} \right\} - N,$$

donde $\hat{p} = \frac{2m_1 + m_2}{2N}$.

Tabla 2.8: resumen de resultados para los datos del ejemplo de Lindley.

Tabla 2.8.1: método clásico.

Tabla 2.7	κ	p
Muestra 1	5,75999999	0,016395073
Muestra 2	4,85810336	0,027516572
Muestra 3	13,5734072	0,000229413
Muestra 4	2,49307479	0,114348168
Total	286,676	0

Tabla 2.8.2: métodos Bayesianos.

Tabla 2.7	\hat{p}	$p_{i1}(\hat{p})$	$p_{i2}(\hat{p})$	$p_{i3}(\hat{p})$
Muestra 1	0,5	0,25	0,5	0,25
Muestra 2	0,17	0,0289	0,2822	0,6889
Muestra 3	0,05	0,0025	0,095	0,9025
Muestra 4	0,05	0,0025	0,095	0,9025
Total	0,1925	0,03705625	0,3108875	0,65205625

Tabla 2.8.2.1: primer método

Tabla 2.8.2.1.1:

Tabla 2.7	η_1	$(\eta_1 + 1)^{-1}$	$\eta_1 (\eta_1 + 1)^{-1}$
Muestra 1	2,435434	0,29108404	0,70891596
Muestra 2	0,7328403	0,577087226	0,4229127742
Muestra 3	1,3135728	0,432231908	0,5677680921
Muestra 4	0,0999457	0,909135738	0,0908642615
Total	∞	0	1

Tabla 2.8.2.1.2:

$\eta_1 p (\eta_1 p + 2p^* - p)^{-1}$ $p (2p^*)^{-1}$	$p^* = 0,5$	$p^* = 0,1$	$p^* = 0,05$	$p^* = 0,01$
Muestra 1	0,039011 0,016395073	0,178626 0,081975364	0,323223 0,163950728	0,917193 0,81975364
Muestra 2	0,0203146 0,027516572	0,0104673 0,137582859	0,0217653 0,275165719	— —
Muestra 3	0,00030133 0,000229413	0,0015062 0,001147063	0,0030113 0,002294126	0,0150135 0,011470629
Muestra 4	0,01273978 0,114348168	0,0117723 0,571740838	— —	— —
Total	0 0	0 0	0 0	0 0

Tabla 2.8.2.2: tercer método

Tabla 2.8.2.2.1:

Tabla 2.7	η_3	$(\eta_3 + 1)^{-1}$	$\eta_3 (\eta_3 + 1)^{-1}$
Muestra 1	0,215029529	0,823025265	0,176974735
Muestra 2	0,04867245	0,953586604	0,046413396
Muestra 3	0,050914007	0,951552642	0,048447358
Muestra 4	0,003873892	0,996141057	0,003858943
Total	∞	0	1

Tabla 2.8.2.2.2:

$\eta_3 p (\eta_3 p + 2p^* - p)^{-1}$ $p (2p^*)^{-1}$	$p^* = 0,5$	$p^* = 0,1$	$p^* = 0,05$	$p^* = 0,01$
Muestra 1	0,0035714 0,016395073	0,0188394 0,081975364	0,0404615 0,163950728	0,494425 0,81975364
Muestra 2	0,0013753 0,027516572	0,0770497 0,137582859	0,0181421 0,275165719	— —
Muestra 3	0,0000117 0,000229413	0,00005846 0,001147063	0,000117 0,002294126	0,00059 0,011470629
Muestra 4	0,000499915 0,114348168	0,00514517 0,571740838	— —	— —
Total	0 0	0 0	0 0	0 0

Entonces, si el valor de Λ en el punto formado por los datos observados de la Tabla 2.6 es $\Lambda(a_{10}, b_{10}, \dots, a_{r0}, b_{r0}) = \lambda_0$, la evidencia utilizada es el p-valor,

$$\begin{aligned} p &= P\left(\Lambda \geq \lambda_0 \mid p_{i1} = p^2, p_{i2} = 2p(1-p), p_{i3} = (1-p)^2, i = 1, \dots, r\right) \\ &= P\left(\chi_{3r-2}^2 \geq \lambda_0\right). \end{aligned}$$

En particular, si $r = 1$ se obtiene

$$p = P\left(\Lambda \geq \lambda_0 \mid p^2, 2p(1-p), (1-p)^2\right) = P\left(\chi_1^2 \geq \lambda_0\right).$$

Por lo tanto, fijados $n_i, i = 1, \dots, r$, y p^* , siguiendo el mismo razonamiento empleado en las secciones 2.3, 2.4 y 2.6, para comparar con el método clásico los resultados obtenidos con los tres métodos bayesianos desarrollados en esta sección, si se considera la ecuación

$$B_k(\pi_0) = \frac{p}{2p^*}, k = 1, 2, 3, \quad (2.22)$$

y se denota por π_{k0}^* , $k = 1, 2, 3$, el valor que verifica (2.22), se obtiene

$$\pi_{k0}^* = \frac{\eta_k p}{\eta_k p + 2p^* - p}, k = 1, 3,$$

mientras que para el segundo método habría que calcular su valor numéricamente, ya que no se puede obtener su expresión explícita. Entonces, utilizando π_{k0}^* se llegaría con el k-ésimo método bayesiano, $k = 1, 2, 3$, respectivamente, a la misma conclusión que con el método clásico.

En la Tabla 2.8 se recogen los resultados que se obtienen para los datos del ejemplo de Lindley (véase la Tabla 2.7) con el método clásico y dos de los métodos bayesianos propuestos. En los tres casos la conclusión obtenida es la misma, se rechaza la homogeneidad de las cuatro muestras, $H_0 : p^2, 2p(1-p), (1-p)^2$, para algún $p \in (0, 1)$. Prácticamente el p-valor es cero y no existen valores de π_0 para los cuales se acepte la hipótesis nula de homogeneidad.

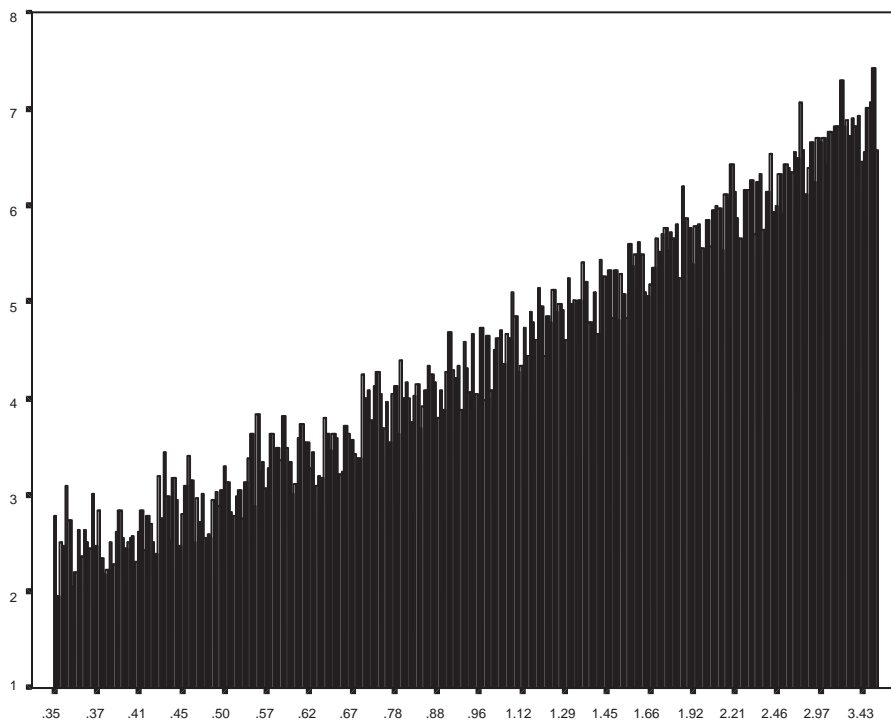
En la Tabla 2.8 también se presentan los resultados que se obtienen al contrastar individualmente si cada una de las cuatro muestras se encuentran en equilibrio Hardy-Weinberg. Se observa que para algunos valores de p^* no existe el valor exacto de π_0 , obtenido mediante (2.22), para el que se alcanza el acuerdo entre los resultados clásico y bayesiano, pero cuando existe, dicho valor disminuye cuando p^* aumenta, tanto con el tercer método como con el primero. Además, cuando el acuerdo es posible, el valor obtenido con el tercer método es más pequeño que el obtenido con el primero. Se observa que para la Muestra 1 de la Tabla 2.7 si $p^* = 0,01$ no hay suficiente evidencia estadística para rechazar H_0 desde el punto de vista clásico, lo cual se manifiesta en los métodos Bayesianos en un aumento significativo del valor de π_0 para el que se alcanza el acuerdo, $\pi_{10}^* = 0,9172$ y $\pi_{30}^* = 0,4944$. Sin embargo, para la muestra 3, si $p^* = 0,01$ se acepta H_0 y $\pi_{10}^* = 0,015$ y $\pi_{30}^* = 0,0115$.

La reconciliación obtenida en los términos de la expresión (2.22) es demasiado estricta ya que π_{k0}^* , $k = 1, 2, 3$, depende de los datos, aunque da una idea de cómo debería ser el valor de π_0 para el que se consiga un acuerdo entre el método clásico y cada uno de los tres métodos bayesianos. Por ejemplo, para intentar eliminar la dependencia que dicha reconciliación tiene de los datos en los contrastes individuales, se han generado todas las posibles tablas 3×1 , para n fijo y conocido. En el caso que nos ocupa, $n = 100$ y se generan las 5151 tablas posibles. Para cada una de estas tablas se ha efectuado el mismo estudio que se ha realizado previamente para cada una de las tablas del ejemplo de Lindley. Por ejemplo, las Muestras 1, 2, 3 y 4 se corresponden con las tablas que ocupan, respectivamente, las posiciones 1685, 1346, 1524 y 215, en la ordenación en sentido ascendente realizada según los valores del estadístico η_1 y 1719, 1286, 1306 y 87, según los valores de η_3 .

Mediante un análisis de datos sencillo se puede comprobar que $p^* = 0$ y $p^* = 1$ son prácticamente los únicos valores de p^* para los que existe un intervalo de valores de π_0 , $I = I(p^*, n = 100)$, tal que el resultado que se obtiene con los métodos bayesianos propuestos para contrastar si una población se encuentra en equilibrio Hardy-Weinberg,

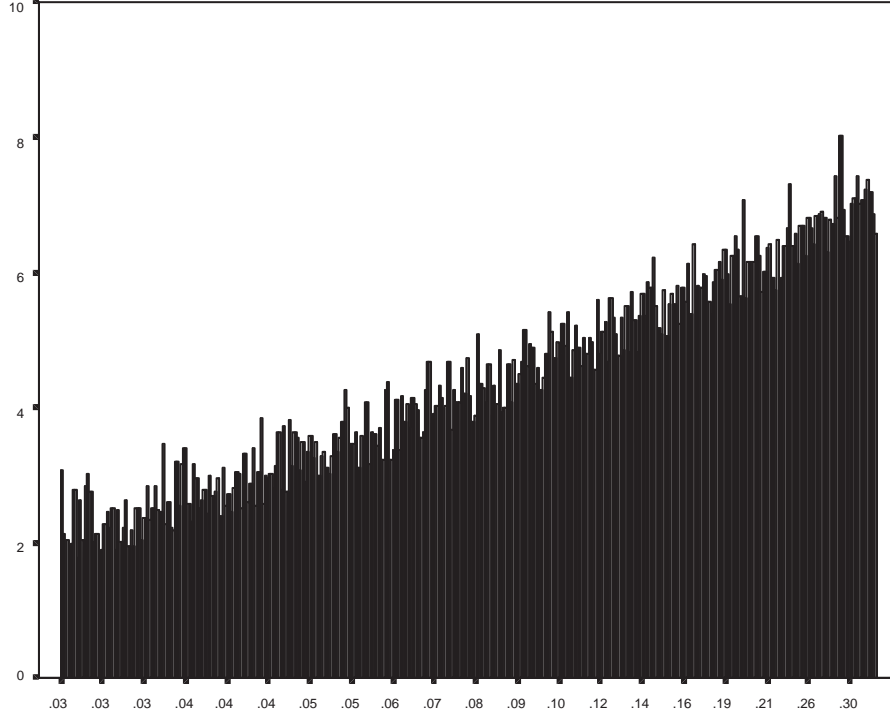
utilizando un valor de $\pi_0 \in I$, es el mismo que se obtiene mediante el método clásico χ^2 de Pearson. Por lo tanto, el acuerdo no es factible ni con el primer método bayesiano ni con el tercero.

Figura 2.4: Diagrama de Barras $(\eta_1(a, b, c), \Lambda(a, b, c))$, para tablas 3×1 con $n = 100$. Existe una función $h_1 : \mathbb{R}^+ \rightarrow \mathbb{R}^+$, no monótona, tal que $\Lambda = h_1(\eta_1)$.



En la Figura 2.4 se representa el diagrama de barras $(\eta_1(a, b, c), \Lambda(a, b, c))$, para las tablas 3×1 con $n = 100$ que ocupan las posiciones desde la 754, con $\Lambda = 2,79$, hasta la 1315, con $\Lambda = 6,57$, primera y última tabla con p-valor $\in (0,01, 0,1)$, en la ordenación en sentido ascendente según los valores del estadístico η_1 efectuada sobre el conjunto de las 4093 tablas 3×1 que verifican que $n\hat{p}^2 > 5$, $n2\hat{p}(1 - \hat{p}) > 5$ y $n(1 - \hat{p})^2 > 5$. Se puede observar que existe una función $h_1 : \mathbb{R}^+ \rightarrow \mathbb{R}^+$, no monótona, tal que $\Lambda = h_1(\eta_1)$.

Figura 2.5: Diagrama de Barras $(\eta_3(a, b, c), \Lambda(a, b, c))$, para tablas 3×1 con $n = 100$. Existe una función $h_3 : \mathbb{R}^+ \rightarrow \mathbb{R}^+$, no monótona, tal que $\Lambda = h_3(\eta_3)$.



En la Figura 2.5 se representa el diagrama de barras $(\eta_3(a, b, c), \Lambda(a, b, c))$, para las tablas para las tablas 3×1 que ocupan las posiciones desde la 722, con $\Lambda = 3,08$, hasta la 1331, con $\Lambda = 6,57$, primera y última tabla con p-valor $\in (0,01, 0,1)$, en la ordenación en sentido ascendente según los valores del estadístico η_3 efectuada sobre el conjunto de las 4093 tablas 3×1 que verifican que $n\hat{p}^2 > 5$, $n2\hat{p}(1 - \hat{p}) > 5$ y $n(1 - \hat{p})^2 > 5$. Se puede observar que existe una función $h_3 : \mathbb{R}^+ \rightarrow \mathbb{R}^+$, no monótona, tal que $\Lambda = h_3(\eta_3)$.

En general, la falta de acuerdo entre el procedimiento clásico usual y los procedimientos Bayesianos propuestos, para contrastar si una población se encuentra en equilibrio Hardy-Weinberg a partir de una muestra, se debe a que, fijado n , si se denotan mediante

$$\ell_1^k = \ell_1^k(p^*, n) = \max_{(a,b,c), a+b+c=n, p>p^*} \pi_0^k,$$

$$\ell_2^k = \ell_2^k(p^*, n) = \min_{(a,b,c), a+b+c=n, p\leq p^*} \pi_0^k,$$

entonces ningún $p^* \in (0, 1)$ verifica la condición suficiente $\ell_1^k < \ell_2^k$, $k = 1, 3$, para $n = 100$.

El motivo fundamental de que no se verifique dicha condición suficiente en el ejemplo estudiado, es que no es posible salvar la discrepancia existente entre el *estadístico de contraste* utilizado con el método clásico usual y el estadístico del que depende la probabilidad a posteriori para los métodos bayesianos propuestos, debido a que la relación funcional existente entre ambos es muy irregular (véanse las Figuras 2.4 y 2.5).

Por el mismo motivo, fijados n_1, \dots, n_r , en el caso de tablas $r \times 3$ (véase la Tabla 2.6) también es previsible esta falta de acuerdo cuando se pretende contrastar si r muestras han sido extraídas de una misma población que se encuentra en equilibrio Hardy-Weinberg para algún $p \in (0, 1)$ desconocido.

2.8. Observaciones

La descripción estadística del problema genético del ejemplo considerado en la sección 2.7 consiste en que se dispone de muestras aleatorias independientes de r poblaciones trinómicas (véase la Tabla 2.6). La distribución trinomial depende de dos parámetros desconocidos, (p_1, p_2, p_3) con $p_1 + p_2 + p_3 = 1$. Cuando $r = 1$, tal y como lo plantea Lindley (1988), se pretende contrastar la hipótesis de que realmente sólo depende de un parámetro p en el sentido Hardy-Weinberg, es decir $H_0 : p^2, 2p(1-p), (1-p)^2$. Alternativamente, se puede expresar $H_0 : 4p_1p_3 = p_2^2$, relación claramente cierta en el equilibrio, y que dibuja una curva en el plano (p_1, p_2) .

Fijado $p_0 \in (0, 1)$, sea

$$B((p_0^2, 2p_0(1-p_0)), \delta) = \left\{ (p_1, p_2) \in (0, 1) \times (0, 1), (p_1 - p_0^2)^2 + (p_2 - 2p_0(1-p_0))^2 \leq \delta^2 \right\},$$

la bola de centro $(p_0^2, 2p_0(1-p_0))$ y radio δ . Entonces, se podría utilizar $\pi(p_1, p_2)$, la opinión a priori sobre (p_1, p_2) , y calcular π_0 mediante

$$\pi_0 = \int_{C(\delta)} \pi(p_1, p_2) dp_2 dp_1,$$

siendo $C(\delta) = \cup_{p_0 \in (0, 1)} B(p_0^2, 2p_0(1 - p_0))$.

Por lo tanto, los resultados obtenidos en la sección 2.7.4 se pueden expresar en términos de δ .

2.9. Conclusiones y Comentarios

En algunas situaciones, utilizando nuestra metodología, es posible conseguir un acuerdo entre la probabilidad a posteriori de la hipótesis nula puntual, como medida de la evidencia bayesiana en el contraste de homogeneidad de poblaciones multinomiales independientes, y el p-valor clásico, como se muestra en los ejemplos que han sido estudiados.

Si se eligen valores adecuados de δ , la discrepancia entre la probabilidad a posteriori y el p-valor es menor que usando directamente $\pi_0 = \frac{1}{2}$ en la distribución a priori de tipo mixto correspondiente. El valor elegido para δ debe verificar que las probabilidades a posteriori de H_0 y $H_{0\delta}$ sean similares, para que la sustitución de una hipótesis por la otra sea coherente.

En los casos que han sido tratados en este capítulo, δ pertenece a un intervalo de valores, $I(p^*, n_1, \dots, n_r) = (\ell_1, \ell_2)$, véase, por ejemplo, la Tabla 2.3, donde se consigue acuerdo en los términos de una condición suficiente de la forma $\ell_1 < \ell_2$, por ejemplo, en el caso \mathbf{p}_0 conocido ℓ_1 y ℓ_2 vienen dados respectivamente en las expresiones (2.8) y (2.9). Es decir, si p^* verifica que $\ell_1 < \ell_2$, entonces tomando $\delta \in (\ell_1, \ell_2)$, para los datos observados de la Tabla 2.4 se satisface una y sólo una de las dos condiciones siguientes:

$$"p > p^* \text{ y además } P(H_0 | n_{110}, \dots, n_{rs0}) > \frac{1}{2},"$$

o

$$"p \leq p^* \text{ y además } P(H_0 | n_{110}, \dots, n_{rs0}) \leq \frac{1}{2},"$$

con lo cual, un clásico que utilizara p^* para ponderar el p-valor a la hora de decidir, llegaría a la misma conclusión que un bayesiano que usara $\pi_0 = \pi_0(\delta)$ con $\delta \in (\ell_1, \ell_2)$ en la distribución a priori de tipo mixto. Este resultado se formaliza el siguiente capítulo.

Cuando la condición suficiente con la que se ha formulado el acuerdo no se cumple, la reconciliación entre ambos métodos no es factible en estos términos. Esto es lo que ocurría en el ejemplo estudiado en la sección 2.7.4. El motivo fundamental de esta falta de acuerdo es que no es posible salvar la discrepancia existente entre el *estadístico de contraste*, que utiliza el método clásico para contruir una *región crítica* y calcular el p-valor, y el estadístico del que depende la probabilidad a posteriori, cuando se utiliza una distribución a priori de tipo mixto que asigna una masa a priori a la hipótesis nula y reparte la probabilidad restante en los puntos de la alternativa mediante una función de densidad.

Parte de estos resultados han sido aceptados para su publicación, véase Gómez-Villegas y González (2005).

Capítulo 3

Acuerdo entre la Aproximación Clásica y Bayesiana en Tablas de Contingencia

En el capítulo 2 se comprobó que, cuando se utiliza una distribución a priori de tipo mixto, no siempre es posible una reconciliación entre las medidas clásicas y bayesianas para el contraste de homogeneidad de distribuciones multinomiales independientes en tablas de contingencia. Para el análisis bayesiano introducido en el capítulo 1, se formalizan los resultados obtenidos en el capítulo 2 y se demuestra un teorema que pone de manifiesto los términos en los que es posible un acuerdo entre ambos métodos.

3.1. Introducción

En los capítulos anteriores, para el problema del contraste de homogeneidad de dos distribuciones binomiales independientes dado en (1.2), con una densidad a priori dada para el parámetro $\theta = (p_1, p_2)$, $\pi(\theta)$, se propuso la siguiente metodología: fijar una bola de radio δ alrededor del punto $\theta_0 = (p_0, p_0)$ de la hipótesis nula, y asignar una masa a priori, π_0 , a H_0 calculada integrando la densidad $\pi(\theta)$ sobre $B(\theta_0, \delta)$, repartiendo la probabilidad

restante, $1 - \pi_0$, sobre H_1 mediante $\pi(\theta)$. Esta metodología se podía generalizar al caso de tablas $r \times s$ con \mathbf{p}_0 conocido. Además, para el contraste de homogeneidad de r poblaciones multinomiales independientes con \mathbf{p}_0 desconocido o de forma funcional conocida, $\mathbf{p}_0 = \mathbf{p}(\omega)$, la masa a priori asignada respectivamente a la hipótesis nula de (2.10) y (2.17) se calculaba integrando la densidad $\pi(\theta)$ sobre $C(\delta) = \bigcup_{\theta_0 \in H_0} B(\theta_0, \delta)$.

Es claro que, en cada uno de los problemas de contraste mencionados anteriormente, la utilización de esta distribución a priori de tipo mixto permite calcular π_0 . Además, con este procedimiento, la probabilidad a posteriori de la hipótesis nula, observados los datos de la Tabla 2.4, se puede expresar de la siguiente forma

$$P(H_0 | n_{11}, \dots, n_{rs}) = \left[1 + \frac{1 - \pi_0}{\pi_0} \eta \right]^{-1}, \quad (3.1)$$

donde η es un estadístico que cuantifica la fuerza de la evidencia en contra de H_0 . Por ejemplo, cuando se asignan distribuciones a priori Dirichlet independientes a cada \mathbf{p}_i , $i = 1, \dots, r$, para contrastar (2.6), el valor de η es el que aparece en la expresión (2.7).

Por otra parte, los métodos clásicos usuales para contrastar (2.6), (2.10) y (2.17) con tablas de contingencia $r \times s$ utilizan como *estadístico de contraste* una medida de la discrepancia entre los valores observados en la tabla 2.4 y los valores esperados cuando H_0 es cierta, en los términos del estadístico χ^2 de Pearson. En este caso, observado un punto (n_{11}, \dots, n_{rs}) , la evidencia utilizada es el p-valor, $p = p(n_{11}, \dots, n_{rs})$. Por ejemplo, para contrastar (1.2) el *estadístico de contraste* es la variable aleatoria Λ dada en la expresión (1.13) y el p-valor correspondiente es (1.14).

Cuando se utiliza el método clásico, se rechaza H_0 cuando $p < p^*$, siendo $p^* \in (0, 1)$ un valor suficientemente pequeño. Por conveniencia, con el método bayesiano propuesto se va a rechazar H_0 cuando $P(H_0 | n_{11}, \dots, n_{rs}) < \frac{1}{2}$, aunque cabe destacar que se podría haber considerado un valor más restrictivo que $\frac{1}{2}$, o incluso otra medida de evidencia bayesiana, por ejemplo el factor Bayes.

En el capítulo 2 se comprobó que en algunas de las situaciones anteriores, utilizando dicha metodología, es posible conseguir un acuerdo cualitativo entre la probabilidad a

posteriori de H_0 y el p-valor. El objetivo ahora es formalizar los resultados obtenidos en el capítulo anterior y determinar las condiciones que tiene que cumplir un valor δ para que, fijado p^* , se satisfaga uno y sólo uno de los dos postulados siguientes:

$$\text{“}p(n_{11}, \dots, n_{rs}) > p^* \text{ y además } P(H_0 | n_{11}, \dots, n_{rs}) > \frac{1}{2}\text{”}, \quad (3.2)$$

$$\text{“}p(n_{11}, \dots, n_{rs}) \leq p^* \text{ y además } P(H_0 | n_{11}, \dots, n_{rs}) \leq \frac{1}{2}\text{”},$$

cualquiera que sea el punto observado (n_{11}, \dots, n_{rs}) .

En la sección 3.2 se establecen los desarrollos preliminares para la obtención de un acuerdo entre la aproximación clásica y bayesiana en tablas de contingencia. En la sección 3.3 se demuestra un teorema que reconcilia ambas aproximaciones en los términos de (3.2), cuando el estadístico η de la expresión (3.1) es el *estadístico de contraste*. En la sección 3.4 se formula un acuerdo entre ambos métodos en los términos de una condición suficiente cuando se utiliza el método clásico usual. En la sección 3.5 se realiza una reparametrización que transforma el espacio paramétrico en \mathbb{R}^2 . Finalmente, en la sección 3.6 se incluyen algunos comentarios.

3.2. Preliminares

Observados los datos de la Tabla 2.4, sea $\eta(n_{11}, \dots, n_{rs}) = \kappa$. Cabe destacar que basta con conocer el valor de η en el punto (n_{11}, \dots, n_{rs}) para calcular la probabilidad a posteriori dada en la expresión (3.1), aunque como tal, η no sea un *estadístico suficiente* para el modelo multinomial. Además, el valor de π_0 en (3.1) depende de la elección de δ . En concreto, cuando la densidad a priori $\pi(\theta)$ es uniforme para contrastar (1.2), $\pi_0 = \pi\delta^2$, para δ suficientemente pequeño.

Entonces, dicha probabilidad a posteriori es en realidad una función de (κ, δ) , es decir

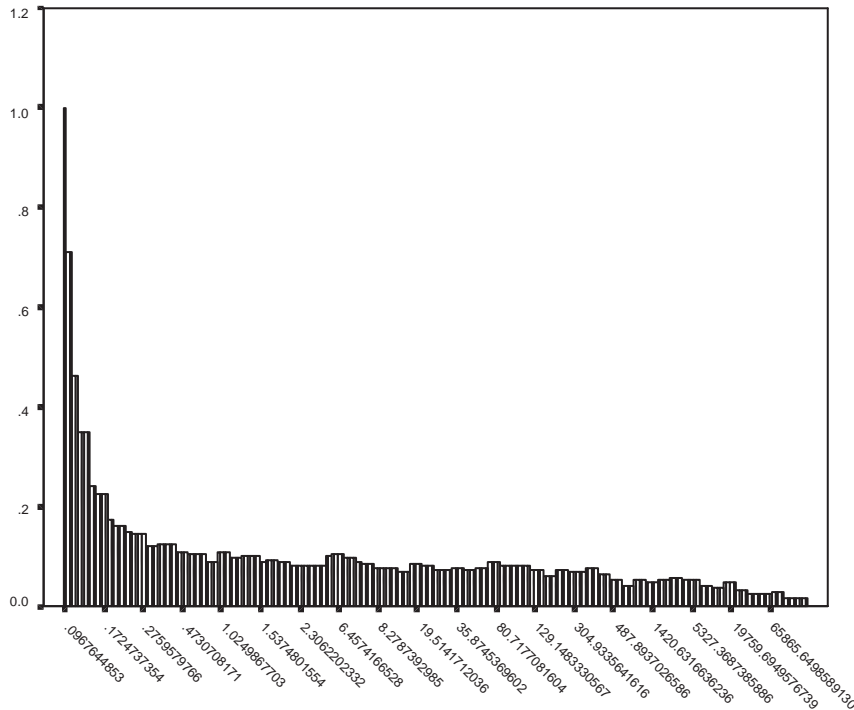
$$P(H_0 | \kappa, \delta) = \left[1 + \frac{1 - \pi_0(\delta)}{\pi_0(\delta)} \kappa \right]^{-1}, \quad (3.3)$$

y verifica las siguientes propiedades:

(i) Para κ fijo, $P(H_0|\kappa, \delta)$ es creciente como función de δ , con $P(H_0|\kappa, \delta = 0) = 0$ y $\lim_{\delta \rightarrow \infty} P(H_0|\kappa, \delta) = 1$.

(ii) Para δ fijo, $P(H_0|\kappa, \delta)$ es decreciente como función de κ , con $P(H_0|\kappa = 0, \delta) = 1$ y $\lim_{\kappa \rightarrow \infty} P(H_0|\kappa, \delta) = 0$.

Figura 3.1 Diagrama de Barras $(\eta(a, c), \pi_0(\delta(a, c)))$, para tablas 2×2 con $n_1 = 18$, $n_2 = 12$, $p_0 = \frac{1}{2}$ y $q_0(p_1, p_2) = I_{(0, 1)}(p_1) I_{(0, 1)}(p_2)$.



La probabilidad a posteriori dada en la expresión (3.1) depende del estadístico η , que se puede utilizar como *estadístico de contraste* para construir una *región crítica*. Observado un valor κ del estadístico η , $\{\eta \geq \kappa\}$ es una posible región crítica coherente

con la propiedad (ii), y el correspondiente p-valor en dicho punto es

$$p = p(\kappa) = \sup_{\theta \in H_0} P(\eta \geq \kappa | \theta) = \sup_{\theta \in H_0} \sum_{\eta(n_{11}, \dots, n_{rs}) \geq \kappa} f(n_{11}, \dots, n_{rs} | \theta), \quad (3.4)$$

donde $f(n_{11}, \dots, n_{rs} | \theta) = \prod_{i=1}^r \frac{n_i!}{\prod_{j=1}^s n_{ij}} \prod_{j=1}^s p_{ij}^{n_{ij}}$.

Con una elección adecuada de δ , la probabilidad a posteriori dada en la expresión (3.2) se puede hacer numéricamente igual al p-valor dado en la expresión (3.4). Igualando las expresiones (3.3) y (3.4), el valor $\delta = \delta(\kappa)$ obtenido de esta igualación es el correspondiente a

$$\pi_0(\delta(\kappa)) = \frac{kp(\kappa)}{kp(\kappa) + 1 - p(\kappa)} \quad (3.5)$$

En particular, si $n_1 = 18$ y $n_2 = 12$, el número de tablas 2×2 posibles es 247. En este caso, observados los datos de la Tabla 1.1, para contrastar (1.2) con $p_0 = \frac{1}{2}$, el p-valor es (2.4), y cuando la densidad a priori es $\pi(p_1, p_2) = I_{(0, 1)}(p_1) I_{(0, 1)}(p_2)$, la probabilidad a posteriori de H_0 se obtiene evaluando (2.1) en (2.2). Además, la función $\kappa = \eta(a, c)$ no es biyectiva y $\pi_0(\delta(\kappa))$, como función de κ , tiene su dominio en un conjunto de 70 valores. Como el p-valor es una función decreciente en κ , aunque $\pi_0(\delta(\kappa))$ no es decreciente (véase la Figura 3.1), si $\kappa_1 < \kappa < \kappa_2$ se puede determinar un conjunto de valores de δ , $D(\kappa_1, \kappa_2)$, tales que $\delta(\kappa) \in D(\kappa_1, \kappa_2)$ cuando $p(\kappa) \in (p(\kappa_2), p(\kappa_1))$. Por ejemplo, si $\kappa_1 = 1,025$ y $\kappa_2 = 8,28$ se obtiene $p(\kappa_1) = 0,1063$, $p(\kappa_2) = 0,0097$ y $D(\kappa_1, \kappa_2) = (0,1544, 0,186)$. Entonces, los cálculos numéricos efectuados en la Tabla 1.3 muestran que eligiendo un valor de δ en el intervalo $(0,17, 0,19)$ la probabilidad a posteriori es similar al p-valor. Por el contrario, si elegimos $\delta = \frac{1}{\sqrt{2\pi}} = 0,399$ entonces $\pi_0 = \frac{1}{2}$ y la probabilidad a posteriori es considerablemente mayor que el p-valor. De esta manera, en el capítulo 1 se concluía que era posible elegir δ de manera que la probabilidad a posteriori estuviera próxima al p-valor, al menos para los p-valores usuales. Además, como ya se indicó en el capítulo 2, a la hora de hacer efectivo el contraste, la probabilidad a posteriori se compara con p^* . Por ejemplo, para los datos de la Tabla 2.1 el p-valor es $p = 0,1529$, y para $\delta = 0,174$,

$\pi_0 = 0,095$ y la probabilidad a posteriori es $0,15366$, con lo que no hay suficiente evidencia estadística para rechazar H_0 , aunque la probabilidad a posteriori sea menor que 0.5 .

Aunque en el caso particular estudiado, con el razonamiento anterior es posible eliminar la dependencia que tiene de los datos el valor de δ obtenido a partir de la expresión (3.5), es previsible que la dificultad de cálculo sea considerablemente mayor cuando se estudien problemas de contraste más generales. Por otra parte, no se encuentra una caracterización que muestre cuáles son los elementos que influyen en la aproximación entre la probabilidad a posteriori y el p-valor.

Quizá sería más conveniente cambiar el punto de vista y considerar el valor

$$\pi_0(\delta(\kappa)) = \frac{\kappa p(\kappa)}{\kappa p(\kappa) + 2p^* - p(\kappa)}, \quad (3.6)$$

que verifica que $P(H_0|\kappa, \delta) = \frac{p(\kappa)}{2p^*}$, para $p^* \in (0, 1)$ fijo y suficientemente pequeño.

Tabla 3.1: resumen de resultados para los datos del ejemplo de Pearson, $p_0 = \frac{1}{2}$ y

$$\pi(p_1, p_2) = I_{(0, 1)}(p_1) I_{(0, 1)}(p_2).$$

κ	$p(\kappa)$	$(\kappa + 1)^{-1}$	$\kappa(\kappa + 1)^{-1}$	δ
6,7265	0,01529	0,1294	0,8706	0,53905

Tabla 2.1	$p^* = 0,5$	$p^* = 0,1$	$p^* = 0,05$	$p^* = 0,01$
$\kappa p(\kappa p + 2p^* - p)^{-1}$	0,09458	0,3577	0,5484	0,9562
δ	0,17351	0,33743	0,4178	0,6057
$p(2p^*)^{-1}$	0,01529	0,07645	0,1529	0,7645

En la Tabla 3.1 están calculados algunos valores de $\pi_0(\delta(\kappa))$ para los datos de la Tabla 2.1 cuando $p_0 = \frac{1}{2}$ en el contraste (1.2) y se utiliza $\pi(p_1, p_2) = I_{(0, 1)}(p_1) I_{(0, 1)}(p_2)$.

Cabe destacar que los resultados obtenidos son similares a los de la Tabla 2.2, donde el *estadístico de contraste* considerado era el estadístico χ^2 de Pearson dado en (2.3), en lugar del estadístico η dado en (2.2).

Aunque utilizando el valor de π_0 obtenido en la expresión (3.6) se llegaría al mismo resultado con los dos métodos en los términos de (3.2), el acuerdo obtenido es demasiado estricto, ya que dicho valor depende de los datos. Para eliminar esta dependencia es necesario considerar el conjunto de todas las posibles tablas $r \times s$ que se pueden formar con valores marginales por filas, n_1, \dots, n_r , fijos y conocidos. Fijado un valor posible κ del estadístico η , sea

$$A_\kappa = \{(n_{11}, \dots, n_{rs}) : \eta(n_{11}, \dots, n_{rs}) = \kappa\}.$$

Es claro que la probabilidad a posteriori dada en la expresión (3.1) es constante sobre A_κ .

En la siguiente sección se demuestra un teorema que caracteriza cómo y cuando es posible un acuerdo entre la aproximación clásica y bayesiana en tablas de contingencia.

3.3. Acuerdo entre la Aproximación Clásica y Bayesiana

En el siguiente Teorema se llega a una reconciliación entre la probabilidad a posteriori del método bayesiano dada en la expresión (3.3) y el p-valor del método clásico dado en la expresión (3.4) en los términos de (3.2).

Teorema 3.3.1 Sean

$$l_1 = l_1(p^*, n_1, \dots, n_r) = \max_{(n_{11}, \dots, n_{rs}), p(\kappa) > p^*} \eta(\eta + 1)^{-1},$$

$$l_2 = l_2(p^*, n_1, \dots, n_r) = \min_{(n_{11}, \dots, n_{rs}), p(\kappa) \leq p^*} \eta(\eta + 1)^{-1},$$

para n_1, \dots, n_r y p^* fijos y conocidos. Entonces $I = I(p^*, n_1, \dots, n_r) = (\ell_1, \ell_2)$ es un intervalo numérico de valores de $\pi_0 = \pi_0(\delta)$ donde se verifica (3.2).

Demostración Fijado un valor posible κ del estadístico η , para $(n_{11}, \dots, n_{rs}) \in A_\kappa$ se verifica que $P(H_0|\kappa, \delta) > \frac{1}{2}$, para $\pi_0 = \pi_0(\delta) > \frac{\kappa}{\kappa+1}$.

La función $g(\kappa) = \frac{\kappa}{\kappa+1}$ es estrictamente creciente. Además, si $\kappa_1 < \kappa_2$, entonces para cualesquiera $(n_{11}, \dots, n_{rs}) \in A_{\kappa_1}$ y $(n'_{11}, \dots, n'_{rs}) \in A_{\kappa_2}$ se verifica que $P(H_0|\kappa_1, \delta) > P(H_0|\kappa_2, \delta)$ y $p(\kappa_1) > p(\kappa_2)$.

Sean

$$\kappa^* = \min_{(n_{11}, \dots, n_{rs}), p(\kappa) \leq p^*} \eta,$$

$$\kappa_* = \max_{(n_{11}, \dots, n_{rs}), p(\kappa) > p^*} \eta,$$

entonces

$$\kappa^* > \kappa_*,$$

$$\min_{(n_{11}, \dots, n_{rs}), p(\kappa) \leq p^*} \eta(\eta+1)^{-1} = \kappa^*(\kappa^*+1)^{-1},$$

$$\max_{(n_{11}, \dots, n_{rs}), p(\kappa) > p^*} \eta(\eta+1)^{-1} = \kappa_*(\kappa_*+1)^{-1},$$

$$\ell_1 = \max_{(n_{11}, \dots, n_{rs}), p(\kappa) > p^*} \eta(\eta+1)^{-1} \leq \min_{(n_{11}, \dots, n_{rs}), p(\kappa) \leq p^*} \eta(\eta+1)^{-1} = \ell_2.$$

Sea $\pi_0 \in (\ell_1, \ell_2)$. Si $(n_{11}, \dots, n_{rs}) \in A_\kappa$, con $\kappa < \kappa_*$, entonces $p(\kappa) \geq p(\kappa_*) > p^*$ y $\pi_0 > \frac{\kappa}{\kappa+1}$. Recíprocamente, si $(n_{11}, \dots, n_{rs}) \in A_\kappa$, con $\kappa \geq \kappa^*$, entonces $p(\kappa) \leq p(\kappa^*) \leq p^*$ y $\pi_0 \leq \frac{\kappa}{\kappa+1}$.

Corolario 3.3.1 Fijados n_1, \dots, n_r y p^* , sean

$$\pi^* = \pi_0(\delta(\kappa^*)) = \frac{\kappa^* p(\kappa^*)}{\kappa^* p(\kappa^*) + 2p^* - p(\kappa^*)},$$

$$\pi_* = \pi_0(\delta(\kappa_*)) = \frac{\kappa_* p(\kappa_*)}{\kappa_* p(\kappa_*) + 2p^* - p(\kappa_*)},$$

donde κ^* y κ_* son los valores definidos en el Teorema 3.3.1. Si $2p^* \geq p(\kappa_*)$, entonces, $\pi^* \leq \pi_*$, $\pi^* \leq \ell_2$ y $\pi_* \geq \ell_1$.

Otra consecuencia inmediata del Teorema 3.3.1 es que el resultado que se obtiene con el método bayesiano basado en la probabilidad a posteriori dada en la expresión (3.1), utilizando un valor de δ tal que $\pi_0(\delta) \in (\ell_1, \ell_2)$, es el mismo que se obtiene mediante el método clásico basado en el p-valor dado en la expresión (3.4), utilizando p^* .

En particular, fijados n_1, n_2 y p^* , para contrastar (1.2) con $p_0 = \frac{1}{2}$, si $\pi_0 \in (\ell_1, \ell_2)$ y $\pi_0 \leq \frac{\pi}{4}$, entonces $\pi_0 = \pi\delta^2$, y para cualquier $\delta \in (\delta_1, \delta_2) = \left(\sqrt{\frac{\ell_1}{\pi}}, \sqrt{\frac{\ell_2}{\pi}}\right)$ se consigue acuerdo. Si $\pi_0 > \frac{\pi}{4}$, entonces $\pi_0 = \pi\delta^2 + 2\sqrt{\delta^2 - \frac{1}{4}} - 4\delta^2 \arcsen\sqrt{1 - \frac{1}{4\delta^2}}$, ya que la $B((p_0, p_0), \delta)$ se sale del cuadrado unidad, y el intervalo (δ_1, δ_2) se determina numéricamente. Cabe destacar que se puede evitar esta situación mediante una reparametrización sencilla que transforma el cuadrado unidad en \mathbb{R}^2 . Esta posibilidad se trata en la sección 3.5.

Tabla 3.2: resumen de resultados para tablas 2×2 con $n_1 = 18, n_2 = 12$,

$$p_0 = \frac{1}{2} \text{ y } \pi(p_1, p_2) = I_{(0, 1)}(p_1) I_{(0, 1)}(p_2).$$

$p^* \in$	(0,49, 0,53)	(0,091, 0,106)	(0,043, 0,053)	(0,0097, 0,0117)
$\delta \in$	(0,221, 0,23)	(0,40141, 0,405)	(0,4521, 0,4632)	(0,5476, 0,5523)
$\pi_0 \in$	(0,153, 0,167)	(0,5062, 0,5155)	(0,6421, 0,6742)	(0,8849, 0,8922)
π^*	0,16085	0,468	0,6276	0,8861
π_*	0,1711	0,5376	0,6717	0,917

Concretamente, en la Tabla 3.2 están calculados algunos de estos valores de δ cuando $n_1 = 18, n_2 = 12$ y $\pi(p_1, p_2) = I_{(0, 1)}(p_1) I_{(0, 1)}(p_2)$. Se puede observar que los resultados que se obtienen cuando el *estadístico de contraste* es Λ en lugar de η (véase la Tabla 2.3), son similares a los de la Tabla 3.2. En este caso, el acuerdo se puede formular en los

términos de una condición suficiente. Este problema se estudia con detalle en la sección 3.4.

3.4. Comparación con el Método Clásico Usual

En el Teorema 3.3.1 se prueba que es posible reconciliar los resultados clásicos y bayesianos en los problemas de contraste de homogeneidad de poblaciones multinomiales independientes estudiados en el capítulo 2. Esto ocurre cuando el *estadístico de contraste* es el correspondiente estadístico $\eta = \eta(n_{11}, \dots, n_{rs})$ del que depende la probabilidad a posteriori obtenida utilizando una distribución a priori de tipo mixto. No obstante, el *estadístico de contraste* del método clásico usual es el estadístico χ^2 de Pearson, $\Lambda = \Lambda(n_{11}, \dots, n_{rs})$. Ahora se va a tratar de comparar ambos métodos cuando el *estadístico de contraste* es Λ .

Fijado un valor posible λ del estadístico Λ , sea

$$B_\lambda = \{(n_{11}, \dots, n_{rs}) : \Lambda(n_{11}, \dots, n_{rs}) = \lambda\}.$$

Observado un punto $(n_{11}, \dots, n_{rs}) \in B_\lambda$, $\{\Lambda \geq \lambda\}$ es una posible región crítica y el p-valor en dicho punto es

$$p = p(\lambda) = \sup_{\theta \in H_0} P(\Lambda \geq \lambda | \theta) = \sup_{\theta \in H_0} \sum_{\Lambda(n_{11}, \dots, n_{rs}) \geq \lambda} f(n_{11}, \dots, n_{rs} | \theta). \quad (3.7)$$

En los casos estudiados en el capítulo 2 se puede comprobar que, para distribuciones a priori Dirichlet independientes, si $\eta(n_{11}, \dots, n_{rs}) = \eta(n'_{11}, \dots, n'_{rs}) = \kappa$, entonces $\Lambda(n_{11}, \dots, n_{rs}) = \Lambda(n'_{11}, \dots, n'_{rs})$. Por lo tanto, $\Lambda = h(\eta)$ para cierta función $h : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ (véanse las Figuras 2.1, 2.2, 2.3, 2.4 y 2.5), y $A_\kappa \subset B_{h(\kappa)}$. En consecuencia, el test clásico de la χ^2 puede expresarse en términos del estadístico η . Sin embargo, el recíproco no es cierto.

En el siguiente Teorema se llega a una reconciliación entre la probabilidad a posteriori del método bayesiano dada en la expresión (3.3) y el p-valor del método clásico usual

dado en (3.7), en los términos de (3.2).

Teorema 3.4.1 Sean

$$\ell_1 = \ell_1(p^*, n_1, \dots, n_r) = \max_{(n_{11}, \dots, n_{rs}), p(\lambda) > p^*} \eta(\eta + 1)^{-1},$$

$$\ell_2 = \ell_2(p^*, n_1, \dots, n_r) = \min_{(n_{11}, \dots, n_{rs}), p(\lambda) \leq p^*} \eta(\eta + 1)^{-1},$$

para n_1, \dots, n_r y p^* fijos y conocidos. Si $\ell_1 < \ell_2$ entonces $I = I(p^*, n_1, \dots, n_r) = (\ell_1, \ell_2)$ es un intervalo numérico de valores de $\pi_0 = \pi_0(\delta)$ donde se verifica (3.2).

Demostración Sean

$$\kappa^* = \min_{(n_{11}, \dots, n_{rs}), p(\lambda) \leq p^*} \eta,$$

$$\kappa_* = \max_{(n_{11}, \dots, n_{rs}), p(\lambda) > p^*} \eta,$$

entonces

$$\ell_1 = \max_{(n_{11}, \dots, n_{rs}), p(\lambda) > p^*} \eta(\eta + 1)^{-1} = \kappa_*(\kappa_* + 1)^{-1} = g(\kappa_*),$$

$$\ell_2 = \min_{(n_{11}, \dots, n_{rs}), p(\lambda) \leq p^*} \eta(\eta + 1)^{-1} = \kappa^*(\kappa^* + 1)^{-1} = g(\kappa^*).$$

Como la función g es estrictamente creciente, si $\ell_1 < \ell_2$, entonces $\kappa_* < \kappa^*$.

Sea $\pi_0 \in (\ell_1, \ell_2)$. Si $(n_{11}, \dots, n_{rs}) \in A_\kappa$, con $\kappa < \kappa_*$, entonces $p(h(\kappa)) > p(h(\kappa_*)) > p^*$ y $\pi_0 > \frac{\kappa}{\kappa+1}$. Recíprocamente, si $(n_{11}, \dots, n_{rs}) \in A_\kappa$, con $\kappa \geq \kappa^*$, entonces $p(h(\kappa)) < p(h(\kappa^*)) \leq p^*$ y $\pi_0 \leq \frac{\kappa}{\kappa+1}$.

Por lo tanto, como consecuencia inmediata del Teorema 3.4.1, fijados n_1, \dots, n_r y p^* , $\ell_1 < \ell_2$ es una condición suficiente para que el resultado que obtiene un bayesiano con la probabilidad a posteriori dada en la expresión (3.1), utilizando un valor de δ tal que $\pi_0(\delta) \in (\ell_1, \ell_2)$, sea el mismo que obtiene un clásico que utiliza p^* para ponderar el p-valor usual dado en (3.7).

Mientras que el intervalo de valores de $\pi_0 = \pi_0(\delta)$, $I = I(p^*, n_1, \dots, n_r) = (\ell_1, \ell_2)$, donde se consigue acuerdo entre ambas aproximaciones siempre existe cuando el *estadístico de contraste* es η (véase el Teorema 3.3.1), la condición $\ell_1 < \ell_2$ no siempre se cumple cuando el *estadístico de contraste* es Λ , de hecho, pueden existir valores de p^* para los que dicha condición no se verifique. Por ejemplo, en tablas de contingencia 2×2 con $n_1 = 18$ y $n_2 = 12$, para contrastar (1.2) con $p_0 = \frac{1}{2}$ y $\pi(p_1, p_2) = I_{(0,1)}(p_1) I_{(0,1)}(p_2)$, $p^* = 0,015$ es un valor que no verifica $\ell_1 < \ell_2$. En estos casos se puede decir que el acuerdo es posible cuando al menos exista un valor $p^* \in (0,01, 0,1)$ tal que $\ell_1 < \ell_2$, en los términos del Teorema 3.4.1 (para el ejemplo citado véase la Tabla 2.3). No obstante, en el ejemplo estudiado en la sección 2.7.4 el acuerdo no es factible para ningún $p^* \in (0,01, 0,1)$, debido fundamentalmente a que es imposible salvar la discrepancia existente entre Λ y η (véanse las Figuras 2.4 y 2.5).

3.5. Reparametrizaciones

Ya se ha comentado anteriormente que si el valor de δ no es suficientemente pequeño, la $B((p_0, p_0), \delta)$ se puede salir del espacio paramétrico $(0, 1) \times (0, 1)$. Esto puede llegar a ser un inconveniente, cuando, en la práctica, se desee calcular el valor de δ correspondiente a un determinado valor π_0 , siendo la dificultad de cálculo cada vez mayor a medida que aumenta la dimensión del espacio paramétrico. Este problema se puede resolver mediante un cambio de variable que transforme el cuadrado unidad en \mathbb{R}^2 , ya que, en este caso, $\pi_0 = \pi\delta^2$, el área de una esfera de radio δ .

Una posibilidad es utilizar distribuciones a priori $Be(0, 0)$ o de Haldane independientes, es decir

$$\pi(p_1, p_2) \propto p_1^{-1} (1 - p_1)^{-1} p_2^{-1} (1 - p_2)^{-1},$$

que es una densidad impropia y equivalente a la densidad uniforme impropia,

$$\pi(\Lambda_1, \Lambda_2) \propto 1,$$

en $\Lambda_1 = \log \frac{p_1}{1-p_1} \in \mathbb{R}$ y $\Lambda_2 = \log \frac{p_2}{1-p_2} \in \mathbb{R}$.

Entonces,

$$\begin{aligned} f(a, c | \Lambda_1, \Lambda_2) &\propto \left(\frac{e^{\Lambda_1}}{e^{\Lambda_1} + 1} \right)^a \left(\frac{1}{e^{\Lambda_1} + 1} \right)^{n_1 - a} \left(\frac{e^{\Lambda_2}}{e^{\Lambda_2} + 1} \right)^c \left(\frac{1}{e^{\Lambda_2} + 1} \right)^{n_2 - c} \\ &= \frac{e^{a\Lambda_1} e^{c\Lambda_2}}{(e^{\Lambda_1} + 1)^{n_1} (e^{\Lambda_2} + 1)^{n_2}}. \end{aligned}$$

Se pretende contrastar (1.2), es decir

$$H_0 : (\Lambda_1, \Lambda_2) = (\Lambda_0, \Lambda_0) \text{ versus } H_1 : (\Lambda_1, \Lambda_2) \neq (\Lambda_0, \Lambda_0),$$

siendo $\Lambda_0 = \log \frac{p_0}{1-p_0} \in \mathbb{R}$.

Sea $B((\Lambda_0, \Lambda_0), \delta) = \{(\Lambda_1, \Lambda_2) \in \mathbb{R}^2 : (\Lambda_1 - \Lambda_0)^2 + (\Lambda_2 - \Lambda_0)^2 < \delta^2\}$. Se propone calcular π_0 en

$$\pi^*(\Lambda_1, \Lambda_2) = \pi_0 I_{H_0}(\Lambda_1, \Lambda_2) + (1 - \pi_0) \pi(\Lambda_1, \Lambda_2) I_{H_1}(\Lambda_1, \Lambda_2),$$

mediante

$$\pi_0 = \int_{B((\Lambda_0, \Lambda_0), \delta)} \pi(\Lambda_1, \Lambda_2) d\Lambda_1 d\Lambda_2 = \pi \delta^2.$$

Entonces,

$$P(H_0 | a, c) = \left[1 + \frac{1 - \pi_0}{\pi_0} \eta \right]^{-1},$$

donde

$$\eta(a, c) = \frac{\int_{\mathbb{R}^2} f(a, c | \Lambda_1, \Lambda_2) \pi(\Lambda_1, \Lambda_2) d\Lambda_1 d\Lambda_2}{f(a, c | \Lambda_0, \Lambda_0)} = \frac{\left(\int_{\mathbb{R}} \frac{e^{a\Lambda_1}}{(e^{\Lambda_1} + 1)^{n_1}} d\Lambda_1 \right) \left(\int_{\mathbb{R}} \frac{e^{c\Lambda_2}}{(e^{\Lambda_2} + 1)^{n_2}} d\Lambda_2 \right)}{\frac{e^{a\Lambda_0 + c\Lambda_0}}{(e^{\Lambda_0} + 1)^{n_1 + n_2}}}.$$

Mediante un cambio de variable

$$\int_{\mathbb{R}^2} f(a, c | \Lambda_1, \Lambda_2) \pi(\Lambda_1, \Lambda_2) d\Lambda_1 d\Lambda_2 = \int_0^1 \int_0^1 f(a, c | p_1, p_2) \pi(p_1, p_2) dp_1 dp_2.$$

Por lo tanto

$$\begin{aligned}\eta(a, c) &= \frac{\int_0^1 \int_0^1 p_1^{a-1} (1-p_1)^{n_1-a-1} p_2^{c-1} (1-p_2)^{n_2-c-1}}{p_0^{a+c} (1-p_0)^{b+d}} \\ &= p_0^{-m_1} (1-p_0)^{-m_2} \frac{\Gamma(a) \Gamma(b) \Gamma(c) \Gamma(d)}{\Gamma(a+b) \Gamma(c+d)}.\end{aligned}$$

Siempre que se observe al menos un éxito y un fracaso en cada población, la probabilidad a posteriori que se obtiene es propia. Si $n_1 = 18$ y $n_2 = 12$, existen 187 tablas 2×2 con al menos un éxito en cada población. Si no se observa ningún éxito o ningún fracaso en al menos una de las dos poblaciones, se necesita pensar más cuidadosamente sobre la distribución a priori y seleccionar una distribución que refleje la opinión a priori de forma más precisa, porque en este caso los datos son menos útiles, y por lo tanto se tiene que confiar más en la distribución a priori.

Cabe destacar que, para distribuciones a priori uniformes o de Laplace independientes,

$$\begin{aligned}\eta_U &= p_0^{-m_1} (1-p_0)^{-m_2} \frac{\Gamma(a+1) \Gamma(b+1) \Gamma(c+1) \Gamma(d+1)}{\Gamma(a+b+2) \Gamma(c+d+2)} \\ &= \frac{abcd}{(n_1+1)n_1(n_2+1)n_2} \eta_H < \eta_H.\end{aligned}$$

Consecuentemente,

$$P_U(H_0|a, c) = \left[1 + \frac{1-\pi_0}{\pi_0} \eta_U\right]^{-1} > \left[1 + \frac{1-\pi_0}{\pi_0} \eta_H\right]^{-1} = P_H(H_0|a, c),$$

uniformemente sobre (a, c) , aunque η_H no conserva la misma ordenación en sentido ascendente respecto a los valores de η_U .

Se puede comprobar que si existe acuerdo en p^* con $P_U(H_0|a, c)$, no necesariamente también existe con $P_H(H_0|a, c)$ (por ejemplo, con distribuciones a priori de Haldane independientes no hay acuerdo para $p^* = 0,01$, mientras que con distribuciones a priori uniformes independientes sí, véase la Tabla 2.3), y en caso de que lo hubiere, es claro que $\delta_U = \delta_U(p^*, n_1, n_2)$ y $\delta_H = \delta_H(p^*, n_1, n_2)$ tienen distinto significado.

La Tabla 3.3 muestra algunos de estos valores de δ para los que se consigue acuerdo con el método χ^2 de Pearson, cuando $n_1 = 18$, $n_2 = 12$, $p_0 = \frac{1}{2}$ en (1.2) y se asignan distribuciones a priori de Haldane independientes. Se puede observar que los valores de π_0 para los que se consigue acuerdo son considerablemente mayores que cuando se utilizan distribuciones a priori uniformes independientes. Por lo tanto, este último es el caso más conservador con respecto a la hipótesis nula. También se puede comprobar que para $\pi_0 = \frac{1}{2}$, $P_H(H_0|a, c) < \frac{1}{2}$, cualquiera que sea el valor observado (a, c) , más aún, $P_H(H_0|a, c) < p(a, c)$. Por lo tanto, como $P_H(H_0|a, c)$ es creciente en π_0 , es necesario incrementar considerablemente el valor de π_0 para poder obtener acuerdo para p^* fijo.

Tabla 3.3: resumen de resultados para tablas 2×2 con $n_1 = 18$ y $n_2 = 12$,

$$p_0 = \frac{1}{2}, \text{ y } \pi(p_1, p_2) \propto p_1^{-1}(1-p_1)^{-1} p_2^{-1}(1-p_2)^{-1}.$$

$p^* \in$	(0,46, 0,51)	(0,087, 0,143)	(0,038, 0,062)	(0,0094, 0,0099)
$\delta \in$	(0,501, 0,507)	(0,5473, 0,5536)	(0,5585, 0,5602)	(0,5636, 0,5639)
$\pi_0 \in$	(0,788, 0,807)	(0,941, 0,963)	(0,98, 0,986)	(0,998, 0,999)

3.6. Comentarios

En los problemas de contraste de homogeneidad de poblaciones multinomiales independientes,

$$H_0 : \theta = \theta_0, \text{ versus } H_1 : \theta \neq \theta_0,$$

en los casos θ_0 conocido, desconocido y de forma funcional conocida, con una densidad a priori previamente dada para el parámetro $\theta = (p_{11}, \dots, p_{1s}, \dots, p_{r1}, \dots, p_{rs})$, se propone la siguiente metodología: fijar $C(\delta) = \bigcup_{\theta_0 \in H_0} B(\theta_0, \delta)$ y utilizar una distribución a priori de tipo mixto que asigna una masa a priori, π_0 , a H_0 calculada integrando la densidad

$\pi(\theta)$ sobre $C(\delta)$, y que reparte la probabilidad restante, $1 - \pi_0$ sobre H_1 mediante $\pi(\theta)$. Con esta aproximación, observados los datos de la Tabla 2.4, la probabilidad a posteriori de H_0 depende de cierto estadístico $\eta = \eta(n_{11}, \dots, n_{1s}, \dots, n_{r1}, \dots, n_{rs})$ que cuantifica la evidencia en contra de H_0 (véase la expresión (3.1)).

En el Teorema 3.3.1 se demuestra que, fijados los tamaños muestrales, n_1, \dots, n_r , y p^* el acuerdo entre la aproximación clásica y bayesiana en tablas de contingencia es siempre posible en los términos de (3.2) cuando se utiliza η como *estadístico de contraste*.

Cuando se utiliza el método clásico usual χ^2 de Pearson el *estadístico de contraste* es $\Lambda = \Lambda(n_{11}, \dots, n_{1s}, \dots, n_{r1}, \dots, n_{rs})$, medida de la discrepancia entre los valores observados y esperados cuando H_0 es cierta. En este caso, fijados n_1, \dots, n_r , y p^* , la reconciliación entre ambas aproximaciones depende de que se verifique la condición $\ell_1 < \ell_2$, donde ℓ_1 y ℓ_2 son los valores dados en el Teorema 3.4.1, siendo el acuerdo posible cuando al menos exista un valor $p^* \in (0,01, 0,1)$ tal que $\ell_1 < \ell_2$.

Por lo tanto, una posibilidad para poner de acuerdo los métodos clásicos y bayesianos en tablas de contingencia, con la metodología propuesta, es utilizar η como *estadístico de contraste*. Teniendo en cuenta que en los problemas estudiados en el capítulo 2, asignando distribuciones a priori Dirichlet independientes a cada $\mathbf{p}_i = (p_{i1}, \dots, p_{is})$, $i = 1, \dots, r$, existe una función $h : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ no invertible y no monótona tal que $\Lambda = h(\eta)$ y además, que tanto η como Λ no son *estadísticos suficientes*, se puede justificar la elección de η . De hecho, dicha relación funcional es una medida de la discrepancia entre ambos estadísticos, de la que depende directamente que se verifique la condición suficiente del Teorema 3.4.1. Cabe destacar que si h hubiera sido la identidad o al menos una función monótona la reconciliación siempre sería posible. Sin embargo, cuando el *estadístico del contraste* sea Λ y no exista un valor $p^* \in (0,01, 0,1)$ tal que $\ell_1 < \ell_2$ en los términos del Teorema 3.4.1 el acuerdo no es posible.

Capítulo 4

Acuerdo entre la Aproximación Clásica y Bayesiana en el Contraste de la Hipótesis Nula Puntual Multivariante

En los contrastes paramétricos de hipótesis nula puntual es sabido que los métodos clásicos y Bayesianos pueden dar lugar a diferentes decisiones. Para el problema del contraste de hipótesis nula puntual multivariante frente a alternativa bilateral se desarrolla un procedimiento que permite ver cómo y cuando es posible un acuerdo entre los procedimientos clásicos y bayesianos.

4.1. Introducción

Sea $(\mathcal{X}, \beta, P_\theta)_{\theta \in \Theta \subset \mathbb{R}^m}$ el modelo estadístico correspondiente a una muestra aleatoria (X_1, \dots, X_n) , $f(x_1, \dots, x_n | \theta)$ la función de verosimilitud de la muestra y $\pi(\theta)$ la densidad que modeliza la opinión inicial sobre $\theta = (\theta_1, \dots, \theta_m)$.

Se pretende contrastar

$$H_0 : \theta = \theta_0, \text{ versus } H_1 : \theta \neq \theta_0, \quad (4.1)$$

donde $\theta_0 = (\theta_{01}, \dots, \theta_{0m})$ es un vector conocido y $\theta \neq \theta_0$ significa que al menos uno de los elementos de θ es diferente del correspondiente elemento de θ_0 .

Entonces, para contrastar (4.1) se necesita una distribución a priori de tipo mixto. Se propone

$$\pi^*(\theta) = \pi_0 I_{H_0}(\theta) + (1 - \pi_0) \pi(\theta) I_{H_1}(\theta), \quad (4.2)$$

siendo π_0 la masa a priori asignada a H_0 .

Aunque existen muchas aproximaciones tanto clásicas como bayesianas para el problema del contraste de hipótesis nula puntual univariante, no es así en el caso multivariante. Algunas aportaciones son Oh (1988) que se ocupa de la distribución normal multivariante y Oh y DasGupta (1999) que investigan la relevancia de π_0 , la probabilidad a priori de H_0 , en la diferencia entre el ínfimo de la probabilidad a posteriori y el p-valor para algunas clases de distribuciones a priori sobre la hipótesis alternativa. Gómez-Villegas, Maín y Sanz (2004) desarrollan un test bayesiano para el problema de contraste dado en (4.1) y sugieren un procedimiento para llegar a la distribución de tipo mixto usando la densidad a priori. Para comparar las aproximaciones clásicas y bayesianas calculan cotas inferiores de la probabilidad a posteriori de H_0 , sobre algunas clases razonables de distribuciones a priori, que son comparadas con el p-valor del test clásico. El resultado es que se obtiene una mejor aproximación porque el p-valor se encuentra entre las medidas de evidencia bayesianas.

Se puede pensar que en lugar de (4.1) es más realista

$$H_{0\delta} : d(\theta_0, \theta) \leq \delta, \text{ versus } H_{1\delta} : d(\theta_0, \theta) > \delta, \quad (4.3)$$

con una métrica adecuada d y un valor δ suficientemente pequeño.

Se pueden considerar varias formas de especificar $d(\theta_0, \theta)$. Una posibilidad es la distancia euclídea, $d(\theta_0, \theta) = \sum_{i=1}^m (\theta_i - \theta_{0i})^2$.

Lo que se propone es utilizar la metodología de Gómez-Villegas, Maín, y Sanz (2004) que consiste en fijar una bola de radio δ centrada en θ_0 , $B(\theta_0, \delta) = \{\theta \in \mathbb{R}^m, d(\theta_0, \theta) \leq \delta\}$, y calcular $\pi_0 = \pi_0(\delta)$ mediante

$$\pi_0 = \int_{B(\theta_0, \delta)} \pi(\theta) d\theta. \quad (4.4)$$

Con este procedimiento la probabilidad a posteriori de la hipótesis nula puntual es

$$P_{\pi^*}(\theta_0 | x_1, \dots, x_n) = \left[1 + \frac{1 - \pi_0(\delta)}{\pi_0(\delta)} \eta(x_1, \dots, x_n) \right]^{-1}, \quad (4.5)$$

donde

$$\eta(x_1, \dots, x_n) = \frac{\int_{\Theta} f(x_1, \dots, x_n | \theta) \pi(\theta) d\theta}{f(x_1, \dots, x_n | \theta_0)} \quad (4.6)$$

es un estadístico que cuantifica la fuerza de la evidencia en contra de H_0 .

Lo que se propone no es sustituir (4.1) por (4.3) sino utilizar (4.3) para calcular π_0 en (4.2) mediante (4.4).

Gómez-Villegas, Maín, y Sanz (2004) dan tres razones que pueden justificar la elección de π_0 como en (4.4), a pesar de que el valor usual es $\pi_0 = \frac{1}{2}$. En primer lugar, en una dimensión, cuando se utilizan (4.2) y (4.4) con valores adecuados de δ , y $\pi(\theta)$ pertenece a la clase de las distribuciones unimodales y simétricas o a la clase de distribuciones ε -contaminadas, se obtiene una mejor aproximación entre el p-valor y la probabilidad a posteriori. En particular, cuando la verosimilitud es normal $\delta \in (0, 1, 0, 3)$. Estos resultados se pueden ver en Gómez-Villegas y Gómez (1992), Gómez-Villegas y Sanz (1998, 2000) y en Gómez-Villegas, Maín y Sanz (2002). Una segunda razón es que si la opinión a priori sobre θ viene dada por la densidad $\pi(\theta)$, entonces la probabilidad a priori del punto de la hipótesis nula de (4.1) es cero, pero utilizando (4.2), la masa a priori asignada a θ_0 es π_0 y esta probabilidad se obtiene mediante $\pi(\theta)$. La tercera razón es que debido a que H_0 es la hipótesis límite de $H_{0\delta}$ cuando $\delta \rightarrow 0$, si la opinión a priori para contrastar (4.3) es $\pi(\theta)$ y $\pi^*(\theta)$, dada en (4.2), es la opinión a priori para contrastar (4.1), parece natural que se satisfaga que $\mu(\pi^*, \pi) \rightarrow 0$ cuando $\delta \rightarrow 0$, para alguna medida adecuada de discrepancia,

μ . Una de las más populares es la medida de información de Kullback-Leibler, que para el problema considerado es $\mu(\pi^*, \pi) = -\ln(1 - \pi_0)$.

En cualquier caso, aquí los resultados se obtienen en función de π_0 y después se especifican para (4.4).

Por otra parte, una medida clásica de la evidencia en contra de la hipótesis nula, que depende de las observaciones, es el p-valor. Sea $\Lambda = \Lambda(X_1, \dots, X_n)$ un estadístico adecuado para contrastar (4.1), por ejemplo el estadístico del *test de la razón de verosimilitudes*,

$$\Lambda = \Lambda(x_1, \dots, x_n) = \frac{\sup_{\theta \in \Theta} f(x_1, \dots, x_n | \theta)}{f(x_1, \dots, x_n | \theta_0)},$$

entonces el p-valor correspondiente al punto (x_1, \dots, x_n) del espacio muestral es

$$\begin{aligned} p(x_1, \dots, x_n) &= P(\Lambda(X_1, \dots, X_n) > \Lambda(x_1, \dots, x_n) | \theta_0) \\ &\approx P(\chi_m^2 > 2 \ln \Lambda(x_1, \dots, x_n)). \end{aligned} \tag{4.7}$$

Con este procedimiento se rechaza H_0 si $p < p^*$, siendo $p^* \in (0, 1)$ un valor suficientemente pequeño.

Ahora, el objetivo es determinar las condiciones que tiene que cumplir un valor δ para que, fijado p^* , se satisfaga uno y sólo uno de los dos postulados siguientes:

$$\text{“}p(x_1, \dots, x_n) > p^*, \text{ y además } P_{\pi^*}(\theta_0 | x_1, \dots, x_n) > \frac{1}{2}\text{”}, \tag{4.8}$$

$$\text{“}p(x_1, \dots, x_n) \leq p^*, \text{ y además } P_{\pi^*}(\theta_0 | x_1, \dots, x_n) \leq \frac{1}{2}\text{”}.$$

cualquiera que sea el punto observado $(x_1, \dots, x_n) \in \chi$.

En la sección 4.2 se desarrolla el procedimiento que permite demostrar un teorema de caracterización del acuerdo entre la aproximación clásica y bayesiana en el problema multivariante del contraste de hipótesis nula puntual. En la sección 4.3 se consideran algunos ejemplos en los que se pueden aplicar los resultados obtenidos. En la sección 4.4 se incluyen algunas observaciones y en la sección 4.5 se finaliza con algunos comentarios relevantes.

4.2. Acuerdo entre la Aproximación Clásica y Bayesiana

En esta sección se definen los conceptos previos que se van a utilizar para la posterior obtención de un teorema de caracterización del acuerdo entre el p-valor dado en la expresión (4.7) y la probabilidad a posteriori dada en (4.5), en los términos de (4.8).

Definición 4.2.1 Sea $(\chi, \beta, P_\theta)_{\theta \in \Theta}$ el modelo estadístico asociado a una muestra aleatoria (X_1, \dots, X_n) y $\pi(\theta)$ la distribución a priori que modeliza la opinión inicial sobre θ . Un estadístico $T = T(X_1, \dots, X_n)$ es *suficiente para el contraste* $H_0 : \theta \in \Theta_0$, versus $H_1 : \theta \in \Theta_1$, con $\Theta_0 \cap \Theta_1 = \phi$ y $\Theta_0 \cup \Theta_1 = \Theta$, si $P_\pi(\Theta_0 | x_1, \dots, x_n) = P_\pi(\Theta_0 | t)$, cuando $T(x_1, \dots, x_n) = t$.

Proposición 4.2.1 Si $T = T(X_1, \dots, X_n)$ es un estadístico *suficiente* para estimar θ , entonces verifica la definición 4.2.1. El recíproco no es cierto.

Cabe destacar que el estadístico η dado en la expresión (4.6) es *suficiente para el contraste* (4.1) cuando la distribución a priori es (4.2). De hecho, cuando la distribución a priori es (4.2), si $T = T(X_1, \dots, X_n)$ es un estadístico *suficiente para el contraste* (4.1), existe una función $g : \mathbb{R} \rightarrow \mathbb{R}$ tal que $g(T) = \eta$.

Definición 4.2.2 Sea $(\chi, \beta, P_\theta)_{\theta \in \Theta}$ el modelo estadístico asociado una muestra (X_1, \dots, X_n) y sean $T_1 = T_1(X_1, \dots, X_n)$ y $T_2 = T_2(X_1, \dots, X_n)$ estadísticos univariantes. T_1 es de *tendencia creciente* respecto a T_2 en un valor $T_1 = t$ si

$$\sup_{T_1(x_1, \dots, x_n) < t} T_2(x_1, \dots, x_n) \leq \inf_{T_1(x_1, \dots, x_n) \geq t} T_2(x_1, \dots, x_n).$$

Proposición 4.2.3 Si $T_1 = h(T_2)$ y $h : \mathbb{R} \rightarrow \mathbb{R}$ es monótona no decreciente entonces T_1

es de *tendencia creciente* respecto a T_2 para todo valor $T_1 = t$. Además, si h es continua y estrictamente creciente entonces

$$\sup_{T_1(x_1, \dots, x_n) < t} T_2(x_1, \dots, x_n) = \inf_{T_1(x_1, \dots, x_n) \geq t} T_2(x_1, \dots, x_n) = h^{-1}(t).$$

Proposición 4.2.4 Si T_1 es de *tendencia creciente* respecto a T_2 en un valor $T_1 = t$ entonces la *región crítica* $\{T_1 \geq t\} = \{T_2 \geq \ell_2(t)\}$, siendo $\ell_2(t) = \inf_{T_1(x_1, \dots, x_n) \geq t} T_2(x_1, \dots, x_n)$.

Proposición 4.2.5 Sean $A_{T_i}(t_i) = \{(x_1, \dots, x_n), T_i(x_1, \dots, x_n) = t_i\}$, $i = 1, 2$. Si $T_1 = h(T_2)$ entonces $A_{T_2}(t_2) \subset A_{T_1}(h(t_2))$.

Teorema 4.2.1 Sea el contraste (4.1) con la distribución a priori (4.2) para π_0 calculado en (4.4). Si Λ es de *tendencia creciente* respecto a η en un valor $\Lambda = \lambda^*$, entonces para $p^* = P_{\theta_0} \{\Lambda \geq \lambda^*\}$ existe un intervalo numérico de valores de $\pi_0 = \pi_0(\delta)$, $I = I(p^*, n) = (\ell_1, \ell_2)$, donde se verifica (4.8).

Demostración Fijado un valor posible κ del estadístico η , para $(x_1, \dots, x_n) \in A_\eta(\kappa)$ se verifica que $P_{\pi^*}(\theta_0 | \kappa) = \left[1 + \frac{1 - \pi_0(\delta)}{\pi_0(\delta)} \kappa\right]^{-1} > \frac{1}{2}$, para $\pi_0 > \frac{\kappa}{\kappa + 1}$.

La función $\pi^0(\kappa) = \frac{\kappa}{\kappa + 1}$ es estrictamente creciente. Además, si $\lambda_1 < \lambda_2$, entonces

$$p(\lambda_1) = P_{\theta_0} \{\Lambda \geq \lambda_1\} \geq P_{\theta_0} \{\Lambda \geq \lambda_2\} = p(\lambda_2),$$

para cualesquiera $(x_1, \dots, x_n) \in A_\Lambda(\lambda_1)$ y $(x'_1, \dots, x'_n) \in A_\Lambda(\lambda_2)$.

Λ es de *tendencia creciente* respecto a η en $\Lambda = \lambda^*$, por lo tanto para $p^* = P_{\theta_0} \{\Lambda \geq \lambda^*\}$ se cumple que

$$\kappa_* = \sup_{(x_1, \dots, x_n), p(\lambda) > p^*} \eta = \sup_{\Lambda(x_1, \dots, x_n) < \lambda^*} \eta \leq \inf_{\Lambda(x_1, \dots, x_n) \geq \lambda^*} \eta = \inf_{(x_1, \dots, x_n), p(\lambda) \leq p^*} \eta = \kappa^*,$$

$$\ell_1 = \pi^0(\kappa_*) = \sup_{(x_1, \dots, x_n), p > p^*} \pi^0(\eta) \leq \inf_{(x_1, \dots, x_n), p \leq p^*} \pi^0(\eta) = \pi^0(\kappa^*) = \ell_2.$$

Sea $\pi_0 \in (\ell_1, \ell_2)$. Si $(x_1, \dots, x_n) \in A_\eta(\kappa)$, con $\kappa < \kappa_* \leq \kappa^*$, entonces $\pi_0 > \ell_1 > \frac{\kappa}{\kappa+1}$ y $P_{\theta_0} \{\Lambda \geq \Lambda(x_1, \dots, x_n)\} > p^*$.

Recíprocamente, si $(x_1, \dots, x_n) \in A_\eta(\kappa)$, con $\kappa \geq \kappa^* \geq \kappa_*$, entonces $\pi_0 < \ell_2 \leq \frac{\kappa}{\kappa+1}$ y $P_{\theta_0} \{\Lambda \geq \Lambda(x_1, \dots, x_n)\} \leq p^*$.

Corolario 4.2.1 Sea $\Lambda = h(\eta)$ y $h : \mathbb{R} \rightarrow \mathbb{R}$ es una función monótona no decreciente. Observada una muestra $(x_1, \dots, x_n) \in A_\eta(\kappa)$, el resultado que se obtiene para contrastar (4.1) con el método bayesiano basado en la probabilidad a posteriori

$$P_{\pi^*}(\theta_0|\kappa) = \left[1 + \frac{1 - \pi_0(\delta)}{\pi_0(\delta)} \kappa \right]^{-1},$$

utilizando un valor de δ tal que $\pi_0(\delta) \in (\ell_1, \ell_2)$, siendo

$$\ell_1 = \ell_1(p^*, n) = \sup_{(x_1, \dots, x_n), p > p^*} \eta(\eta + 1)^{-1}, \quad (4.9)$$

$$\ell_2 = \ell_2(p^*, n) = \inf_{(x_1, \dots, x_n), p \leq p^*} \eta(\eta + 1)^{-1}, \quad (4.10)$$

es el mismo que el que se obtiene mediante el método clásico utilizando $p^* = P(\Lambda \geq \lambda^*)$ para ponderar el p-valor, $p(h(\kappa)) = P_{\theta_0} \{\Lambda \geq h(\kappa)\}$. Además, si h es continua y estrictamente creciente, entonces

$$\ell_1 = \ell_2 = \pi_0(\delta, p^*, n) = \frac{h^{-1}(\lambda^*)}{h^{-1}(\lambda^*) + 1}. \quad (4.11)$$

Demostración La demostración es inmediata utilizando la Proposición 4.2.3.

Fijados el tamaño muestral n y p^* , otra consecuencia inmediata del Teorema 4.2.1 es que, para contrastar (4.1) (con la distribución a priori dada en (4.2), para π_0 calculado en (4.4)), una condición suficiente para que exista acuerdo en los términos de (4.8) (entre el p-valor dado en (4.7) y la probabilidad a posteriori dada en (4.5)), es que $\ell_1 \leq \ell_2$, siendo ℓ_1 y ℓ_2 los valores dados, respectivamente, en las expresiones (4.9) y (4.10). Además, si

$\Lambda = h(\eta)$ y $h : \mathbb{R} \rightarrow \mathbb{R}$ es monótona no decreciente, la condición $\ell_1 \leq \ell_2$ siempre se verifica para cualquier p^* , y cuando h es continua y estrictamente creciente, el acuerdo siempre es posible y se alcanza en (4.11).

No obstante, cuando el *estadístico de contraste* no es Λ , este resultado se puede enunciar en un contexto más general de la siguiente forma.

Teorema 4.2.2 Sea el contraste (4.1) con la distribución a priori (4.2) para π_0 calculado en (4.4). Sea $T_1 = T_1(X_1, \dots, X_n)$ el *estadístico de contraste*, $p(t_1) = P(T_1 \geq t_1)$ el p-valor correspondiente en cada $A_{T_1}(t_1)$ y $T_2 = T_2(X_1, \dots, X_n)$ un *estadístico suficiente para el contraste* tal que $\eta = g(T_2)$, con $g : \mathbb{R} \rightarrow \mathbb{R}$ y η dado en (4.6). Entonces, fijados n y p^* , si

$$\ell_1 = \sup_{(x_1, \dots, x_n), p > p^*} \frac{g(T_2)}{g(T_2) + 1} \leq \inf_{(x_1, \dots, x_n), p \leq p^*} \frac{g(T_2)}{g(T_2) + 1} = \ell_2, \quad (4.12)$$

el intervalo numérico de valores de $\pi_0 = \pi_0(\delta)$, $I = I(p^*, n) = (\ell_1, \ell_2)$, verifica (4.8). Además, si existe una función $h : \mathbb{R} \rightarrow \mathbb{R}$ monótona no decreciente tal que $T_1 = h(\eta)$, entonces, la condición (4.12) siempre se verifica para cualquier p^* , y cuando h es continua y estrictamente creciente, entonces

$$\ell_1 = \ell_2 = \frac{h^{-1}(t_1^*)}{h^{-1}(t_1^*) + 1},$$

siendo t_1^* tal que $p(t_1^*) = p^*$.

Demostración La demostración es similar a la del Teorema 4.2.1 definiendo, fijado un valor posible t_2 del estadístico T_2 , $\pi^0(t_2) = \frac{g(t_2)}{g(t_2)+1}$.

4.3. Aplicaciones

En esta sección se presentan algunos ejemplos en los que se aplican los resultados obtenidos en la sección anterior.

4.3.1. Paradoja de Lindley

Cuando la hipótesis nula es puntual es bien conocido que las medidas de evidencia clásicas y bayesianas pueden dar lugar a respuestas radicalmente distintas. En el célebre artículo de Lindley (1957) se ilustra la posible discrepancia en el caso normal. Lindley proporciona un ejemplo para mostrar que, en una dimensión, si H_0 es una hipótesis simple y x es el resultado de un experimento, los dos fenómenos siguientes pueden ocurrir simultáneamente: “un test para H_0 revela que x es significativo a un nivel del 5 %” y “la probabilidad a posteriori de H_0 , dado ese x , es, para probabilidades a priori bastante pequeñas de H_0 , mayor que el 95 %”. Se comenzará dando la formulación matemática de dicho ejemplo tal y como lo plantea Lindley, y posteriormente se analizará con detalle este fenómeno que, aunque pueda parecer paradójico, es bastante general.

Sea (x_1, \dots, x_m) el valor observado de una muestra aleatoria de una distribución normal de media θ y varianza conocida σ^2 . Sea c la probabilidad a priori de que $\theta = \theta_0$. Se supone que el resto de la probabilidad a priori está distribuída uniformemente sobre algún intervalo I que contenga a θ_0 . Se deberían considerar situaciones en las que \bar{x} , la media aritmética de las observaciones, se encuentre dentro del intervalo I para n suficientemente grande y por lo tanto $\bar{x} - \theta_0$ tienda a cero cuando n aumenta. Entonces, la probabilidad a posteriori de que $\theta = \theta_0$ es

$$\bar{c} = \frac{ce^{-\frac{n(\bar{x}-\theta_0)^2}{2\sigma^2}}}{K}, \quad (4.13)$$

donde $K = ce^{-\frac{n(\bar{x}-\theta_0)^2}{2\sigma^2}} + (1 - c) \int_I e^{-\frac{n(\bar{x}-\theta)^2}{2\sigma^2}} d\theta$. En virtud de las suposiciones sobre \bar{x} e I la integral puede ser evaluada como $\sigma\sqrt{\frac{2\pi}{n}}$.

Ahora, se va a suponer que el valor observado \bar{x} es significativo a nivel α con el test usual para la media θ_0 de una distribución normal con varianza conocida, es decir, $\bar{x} = \theta_0 + \lambda_\alpha \frac{\sigma}{\sqrt{n}}$, donde λ_α es un número que solamente depende de α tal que $\Phi(\lambda_\alpha) = 1 - \alpha$, siendo Φ la función de distribución de la normal estándar. Insertando este valor de \bar{x} en

(4.13) se obtiene la siguiente probabilidad a posteriori de que $\theta = \theta_0$:

$$\bar{c} = \frac{ce^{-\frac{1}{2}\lambda_\alpha^2}}{ce^{-\frac{1}{2}\lambda_\alpha^2} + (1-c)\sigma\sqrt{\frac{2\pi}{n}}}. \quad (4.14)$$

A partir de la expresión (4.14) se puede observar que $\bar{c} \rightarrow 1$ cuando $n \rightarrow \infty$. Por lo tanto, cualquiera que sea el valor de c , se puede encontrar un valor de n , dependiente de c y de α , tal que “ \bar{x} sea significativamente distinto de θ_0 al nivel α ” y “la probabilidad a posteriori de que $\theta = \theta_0$ sea del $100(1-\alpha)\%$ ”. Además, el fenómeno persistiría con casi cualquier distribución de probabilidad a priori que se concentrara únicamente sobre el valor de la hipótesis nula. Por ejemplo, si hay una cantidad c en $\theta = \theta_0$ y el resto se distribuye sobre I conforme a una densidad $p(\theta)$, donde $\int_I p(\theta)d\theta = 1-c$, entonces, si $p(\theta)$ está acotada, \bar{c} todavía converge a 1. Es suficiente con que $p(\theta)$ no converja a ∞ más rápidamente de lo que θ converge a θ_0 .

Esta es la *paradoja de Lindley*. La interpretación usual del primer resultado es que hay una buena razón para creer que $\theta \neq \theta_0$, mientras que la del segundo es que existen motivos para creer que $\theta = \theta_0$. Las dos interpretaciones están en conflicto directo. Lindley argumenta que una paradoja sólo habrá sido generada si es posible mostrar que existen situaciones donde sea razonable considerar una distribución a priori de esta forma y que a su vez, un test significativo de probabilidades cola de este tipo sea comunmente utilizado.

Lindley, utilizando el argumento de Savage (1954), considera la *paradoja* en aquellas situaciones donde la probabilidad a priori exista y tenga una concentración en el punto de la hipótesis nula. Según Lindley, como la probabilidad a posteriori varía enormemente con n para un nivel de significación fijo, en un caso extremo produciendo un resultado en conflicto directo con el nivel de significación, el grado de convicción no es aproximadamente el mismo en dos situaciones con niveles de significación iguales, es decir, “5% en una muestra pequeña hoy no significa lo mismo que 5% en una muestra más grande mañana”.

Cuando el nivel de significación está fijo, Jeffreys (véase, en particular, 1948, Apéndice) evalúa la variación en la probabilidad a posteriori, observando que \bar{c} dado en (4.14) converge a 1 muy lentamente, y que para valores moderados de n , \bar{c} puede ser menor que c

con un nivel de significación prescrito, estando los dos conceptos en un acuerdo razonable.

Sea $A = \frac{ce^{-\frac{1}{2}\lambda_\alpha^2}}{(1-c)\sqrt{2\pi}}$, entonces

$$\bar{c} = \frac{A}{A + \frac{\sigma}{\sqrt{n}}}$$

y $\bar{c} \rightarrow 0$ cuando $\frac{\sigma}{\sqrt{n}} \rightarrow \infty$. Por lo tanto, en un experimento de tamaño pequeño, una significación del 5% puede dar fuerte evidencia para dudar de la hipótesis nula.

Tabla 4.1: valores de \bar{c} calculados por Lindley (1987) para diferentes valores de $t = \frac{n}{\sigma^2}$, cuando $c = \frac{1}{2}$ y se utiliza un test significativo al 5%.

t	\bar{c}	t	\bar{c}	t	\bar{c}	t	\bar{c}
1	0,055	40	0,270	600	0,589	10000	0,854
2	0,076	60	0,312	800	0,623	20000	0,892
3	0,092	80	0,343	1000	0,649	40000	0,921
4	0,105	100	0,369	2000	0,723	60000	0,935
5	0,116	200	0,453	4000	0,787	80000	0,943
10	0,156	300	0,503	6000	0,819	100000	0,949
20	0,207	400	0,539	8000	0,839	∞	1

A continuación se reproduce el ejemplo numérico que aparece en el artículo citado de Lindley. Si se toma $c = \frac{1}{2}$ y se utiliza un test significativo al 5%, entonces $\lambda_\alpha^2 = \chi_{1,\alpha}^2 = 1,96^2 = 3,841$ y $A = 0,0584$. La tabla 4.1 proporciona los valores de \bar{c} para diferentes valores de $t = \frac{n}{\sigma^2}$. Si $\sigma = 1$, $t = n$ y se puede observar que para muestras pequeñas ($n \leq 10$) la probabilidad a posteriori de que $\theta = \theta_0$ ha disminuído apreciablemente con respecto a su valor inicial $\frac{1}{2}$, por lo tanto la evidencia permite rechazar la validez de la hipótesis nula. Para tamaños muestrales intermedios ($10 < n < 100$) la probabilidad sólo ha decrecido un poco, así que aunque no se esté tan seguro como inicialmente sobre la

hipótesis nula, las dudas no son tan grandes. Cuando n alcanza un valor en torno a 300, \bar{c} es igual a c y el experimento, a pesar de su 5% de significación, no ha alterado en absoluto la creencia en la hipótesis nula. Para alcanzar el fuerte contraste de la *paradoja* es necesario tomar n alrededor de 10000. Naturalmente, si σ es más pequeño, muestras más pequeñas serían suficientes.

Una vez expuestos los términos en los que se desarrolla la *paradoja de Lindley*, se va a comprobar que dicho fenómeno no resulta paradójico si se tiene en cuenta que, para n y α fijos, existe un valor de c , que no depende de las observaciones, para el que la conclusión que se obtiene con el método bayesiano basado en la probabilidad a posteriori dada en (4.13) es la misma que la que se obtiene con el método clásico. En primer lugar, se observa que dicha probabilidad a posteriori se puede expresar de la siguiente forma:

$$\bar{c} = \left(1 + \frac{1-c}{c} h^{-1}(T(\bar{x}, \theta_0)) \right)^{-1},$$

donde $h^{-1}(u) = \sigma \sqrt{\frac{2\pi}{n}} e^{\frac{1}{2}u}$ y $T(\bar{x}, \theta_0) = \frac{n(\bar{x}-\theta_0)^2}{\sigma^2}$. Entonces, como consecuencia inmediata del Teorema 4.2.2, el valor para el que se consigue acuerdo entre ambos métodos es

$$c = c(n, \alpha) = \frac{h^{-1}(\chi_{1,\alpha}^2)}{h^{-1}(\chi_{1,\alpha}^2) + 1} = \left(1 + \frac{1}{\sigma \sqrt{\frac{2\pi}{n}}} e^{-\frac{1}{2}\chi_{1,\alpha}^2} \right)^{-1}. \quad (4.15)$$

El argumento que utiliza Lindley para poner de manifiesto la existencia de una paradoja es que es posible encontrar un valor de $n = n(c, \alpha)$ para el que el resultado que se obtiene con la probabilidad a posteriori es radicalmente distinto al del método clásico. Para su ejemplo numérico utiliza $c = \frac{1}{2}$ y un nivel de significación del 5%, entonces si se evalúa la expresión (4.15) en $c = \frac{1}{2}$ se puede comprobar que dicho valor sólo se obtiene para $n = 2\pi e^{\chi_{1,\alpha}^2} = 292,8 \approx 300$. En general, si c y α están fijos, para que se alcance el acuerdo entre ambos métodos se debería tomar

$$n = n(c, \alpha) = \frac{1-c}{c} 2\pi e^{\chi_{1,\alpha}^2}.$$

Por lo tanto, desde nuestro punto de vista, dicho fenómeno no resulta paradójico.

4.3.2. Cotas Inferiores para Distribuciones Unimodales y Simétricas

Sea $\mathbf{X} = (X_1, \dots, X_m)'$ con distribución $N_m(\theta, \sigma^2 \mathbf{I})$, donde σ^2 es conocida, \mathbf{I} es la matriz identidad $m \times m$ y $\theta = (\theta_1, \dots, \theta_m)'$ desconocido. Se pretende contrastar (4.1) con una muestra de tamaño n . En esta situación el *estadístico de contraste* usual es

$$T(\bar{\mathbf{X}}, \theta_0) = \frac{n}{\sigma^2} (\bar{\mathbf{X}} - \theta_0)' (\bar{\mathbf{X}} - \theta_0),$$

donde $\bar{\mathbf{X}} = (\bar{X}_1, \dots, \bar{X}_m)'$. La distribución de $T(\bar{\mathbf{X}}, \theta_0)$ cuando H_0 es cierta es χ_m^2 . Por lo tanto, el p-valor correspondiente al valor observado $\bar{\mathbf{x}} = (\bar{x}_1, \dots, \bar{x}_m)'$ es

$$p(\bar{\mathbf{x}}) = P\{\chi_m^2 \geq T(\bar{\mathbf{x}}, \theta_0)\}.$$

Si $\pi^*(\theta)$ es la distribución a priori de tipo mixto dada en (4.2) con π_0 calculado mediante (4.4), aplicando los resultados de Gómez-Villegas, Maín y Sanz (2004), el ínfimo de la probabilidad a posteriori del punto de la hipótesis nula cuando $\pi(\theta) \in Q_{US}$, siendo Q_{US} la clase de distribuciones unimodales y simétricas en θ_0 , es

$$\inf_{\pi \in Q_{US}} P(H_0 | \bar{\mathbf{x}}) = \left[1 + \frac{2^{\frac{m}{2}} \Gamma\left(\frac{m}{2} + 1\right)}{\delta^{*m}} e^{\frac{1}{2}T(\bar{\mathbf{x}}, \theta_0)} \right]^{-1},$$

donde $\delta^* = \frac{\delta\sqrt{n}}{\sigma^2}$.

Si $\bar{\mathbf{x}}$ es el valor observado en una muestra de tamaño n , para

$$\delta^* > \left[2^{\frac{m}{2}} \Gamma\left(\frac{m}{2} + 1\right) e^{\frac{1}{2}T(\bar{\mathbf{x}}, \theta_0)} \right]^{\frac{1}{m}}$$

se verifica que $\inf_{\pi \in Q_{US}} P(H_0 | \bar{\mathbf{x}}) > \frac{1}{2}$.

Además, fijado p^* , si $2p^* > p(\bar{\mathbf{x}})$, para

$$\delta^*(\bar{\mathbf{x}}) = \left[\frac{2^{\frac{m}{2}} \Gamma\left(\frac{m}{2} + 1\right) e^{\frac{1}{2}T(\bar{\mathbf{x}}, \theta_0)} p(\bar{\mathbf{x}})}{2p^* - p(\bar{\mathbf{x}})} \right]^{\frac{1}{m}}$$

se cumple que $\inf_{\pi \in Q_{US}} P(H_0 | \bar{\mathbf{x}}) = \frac{p(\bar{\mathbf{x}})}{2p^*}$.

Tabla 4.2: valores $\delta^*(p^*, n, m) = \frac{\delta\sqrt{n}}{\sigma^2}$ para los que se consigue acuerdo en la clase de distribuciones a priori unimodales y simétricas.

δ^*	$m = 2$	$m = 5$	$m = 10$	$m = 20$
$p^* = 0,5$	2	2,78	3,64	4,88
$p^* = 0,1$	4,47	4,53	5,08	6,12
$p^* = 0,05$	6,32	5,44	5,7	6,6
$p^* = 0,01$	14,14	8,13	7,28	7,7
$p^* = 0,001$	44,72	13,95	10,02	9,34

Tabla 4.3: acuerdo entre el p -valor y el ínfimo de la probabilidad a posteriori en $t = T(\bar{x}, \theta_0) = 7$.

$t = 7$	$m = 2$	$m = 5$	$m = 10$	$m = 20$
$p(t)$	0,0302	0,22064	0,72544	0,9967

$\inf_{\pi \in Q_{us}} P(H_0 t = 7, \delta^*)$	$m = 2$	$m = 5$	$m = 10$	$m = 20$
$p^* = 0,5$	0,0787	0,23133	0,76331	0,99791
$p^* = 0,1$	0,2992	0,77584	0,98894	0,99998
$p^* = 0,05$	0,46066	0,89648	0,99651	0,9999995
$p^* = 0,01$	0,81027	0,98473	0,9997	0,9999998
$p^* = 0,001$	0,97712	0,99897	0,999988	0,9999999

Sea t^* tal que $P\{\chi_m^2 \geq t^*\} = p^*$. Entonces, como consecuencia del Teorema 4.2.2, siempre existe acuerdo entre ambos métodos y se obtiene para

$$\delta^* = \left[2^{\frac{m}{2}} \Gamma\left(\frac{m}{2} + 1\right) e^{\frac{1}{2}t^*} \right]^{\frac{1}{m}} = \delta^*(t^*).$$

En la Tabla 4.2 están calculados los valores de δ^* que se obtienen para varios valores de m y de p^* . A modo de comprobación, en la Tabla 4.3 se comparan ambos métodos cuando el valor observado del *estadístico de contraste* es $t = T(\bar{\mathbf{x}}, \theta_0) = 7$, calculando, para los valores de m de la Tabla 4.2, los p-valores correspondientes y el ínfimo de la probabilidad a posteriori en los valores δ^* obtenidos en la Tabla 4.2. Por ejemplo, se puede observar que cuando $m = 2$ el correspondiente p-valor es $p(7) = 0,0302$. Por lo tanto, un clásico que utilizara $p^* = 0,05$ rechazaría H_0 , mientras que si utilizara $p^* = 0,01$ aceptaría H_0 , el mismo resultado que obtendría un bayesiano utilizando, respectivamente, los valores $\delta^* = 6,32$ y $\delta^* = 14,14$ de la Tabla 4.2, puesto que en este caso $\inf_{\pi \in Q_{us}} P(H_0 | t = 7, \delta^* = 6,32) = 0,46066$ y $\inf_{\pi \in Q_{us}} P(H_0 | t = 7, \delta^* = 14,14) = 0,81027$.

Se puede efectuar un estudio similar al anterior cuando las variables de la muestra tengan la misma varianza σ^2 , y un coeficiente de correlación común ρ , es decir cuando la distribución de $\mathbf{X} = (X_1, \dots, X_m)'$ sea normal multivariante, $N_m(\theta, \Sigma)$, con una estructura especial de correlación,

$$\Sigma = \sigma^2 \begin{pmatrix} 1 & \rho & \cdots & \delta \\ \delta & 1 & \cdots & \delta \\ \vdots & \vdots & \ddots & \vdots \\ \delta & \delta & \cdots & 1 \end{pmatrix},$$

con σ^2 y ρ conocidos, puesto que en este caso

$$\inf_{\pi \in Q_{US}} P(H_0 | \bar{\mathbf{x}}) = \left[1 + \frac{\Gamma\left(\frac{m}{2} + 1\right)}{\delta^{*m}} \left(2(1 - \rho)^{m-1} (1 + (m-1)\rho) \right)^{\frac{m}{2}} e^{\frac{t}{2}} \right]^{-1},$$

siendo $t = n\bar{\mathbf{x}}'\Sigma^{-1}\bar{\mathbf{x}}$ (véase Gómez-Villegas, Maín y Sanz (2004)).

4.3.3. Cotas Inferiores para Mixturas de Normales con Parámetro de Escala

En el mismo contexto de la sección 4.3.2, se pretende contrastar (4.1) con $\theta_0 = 0$. En este caso el *estadístico de contraste* usual es

$$T(\bar{\mathbf{X}}) = \frac{n}{\sigma^2} \bar{\mathbf{X}}' \bar{\mathbf{X}}.$$

Aplicando los resultados de Gómez-Villegas, Maín y Sanz (2004), si $m > 2$, el ínfimo de la probabilidad a posteriori del punto de la hipótesis nula cuando $\pi(\theta) \in Q_N$, siendo

$$Q_N = \left\{ \pi(\theta|v^2) \approx N_m(\mathbf{0}, v^2\mathbf{I}), \pi(v^2) \text{ función no decreciente en } (0, \infty) \right\},$$

es

$$\inf_{\pi \in Q_N} P(H_0|t) = \left[1 + \frac{1}{\delta^{*2}} \frac{F_{m-2}(t)}{f_m(t)} \right]^{-1},$$

donde $t = \frac{n}{\sigma^2} \bar{\mathbf{x}}' \bar{\mathbf{x}}$, $\delta^* = \frac{\delta\sqrt{n}}{\sigma^2}$ y F_{m-2} es la función de distribución de una variable aleatoria χ_{m-2}^2 con f_m la función de densidad de una variable aleatoria χ_m^2 .

Si $\bar{\mathbf{x}}$ es el valor observado en una muestra de tamaño n y $T(\bar{\mathbf{x}}) = t$, entonces el p-valor es $p(t) = P\{\chi_m^2 \geq t\}$, y para

$$\delta^* > \left[\frac{F_{m-2}(t)}{f_m(t)} \right]^{\frac{1}{2}}$$

se verifica que $\inf_{\pi \in Q_N} P(H_0|t) > \frac{1}{2}$.

Tabla 4.4: valores $\delta^*(p^*, n, m) = \frac{\delta\sqrt{n}}{\sigma^2}$ para los que se consigue acuerdo en la clase de *mixturas de normales con parámetro de escala*.

δ^*	$m = 5$	$m = 10$	$m = 15$	$m = 20$
$p^* = 0,5$	2,38	2,72	2,95	3,13
$p^* = 0,1$	5,14	5,77	6,21	6,55
$p^* = 0,05$	7,15	7,96	8,52	8,97
$p^* = 0,01$	15,51	17,01	18,05	18,89
$p^* = 0,001$	48,02	50,63	54,3	56,48

Además, fijado p^* , para

$$\delta^*(t) = \left[\frac{F_{m-2}(t)}{f_m(t)} \frac{p(t)}{2p^* - p(t)} \right]^{\frac{1}{2}}$$

se cumple que $\inf_{\pi \in Q_{US}} P(H_0|t) = \frac{p(t)}{2p^*}$.

Tabla 4.5: acuerdo entre el p-valor y el ínfimo de la probabilidad a posteriori para $t = T(\bar{\mathbf{x}}) = 20$.

$t = 20$	$m = 5$	$m = 10$	$m = 15$	$m = 20$
$p(t)$	0,00125	0,02925	0,1719	0,45793

$\inf_{\pi \in Q_{us}} P(H_0 t = 20, \delta^*)$	$m = 5$	$m = 10$	$m = 15$	$m = 20$
$p^* = 0,5$	0,00304	0,06592	0,2691	0,4781
$p^* = 0,1$	0,01407	0,24163	0,6205	0,8011
$p^* = 0,05$	0,02689	0,3772	0,7549	0,883
$p^* = 0,01$	0,1155	0,7343	0,9325	0,971
$p^* = 0,001$	0,5546	0,9622	0,99206	0,9966

Sea t^* tal que $P\{\chi_m^2 \geq t^*\} = p^*$. Entonces, como consecuencia del Teorema 4.2.2, siempre existe acuerdo entre ambos métodos y se obtiene para

$$\delta^*(p^*, n, m) = \left[\frac{F_{m-2}(t^*)}{f_m(t^*)} \right]^{\frac{1}{2}} = \delta^*(t^*).$$

En la Tabla 4.4 están calculados los valores de δ^* que se obtienen para varios valores de m y de p^* . Además, en la Tabla 4.5 se comparan numéricamente ambos métodos cuando el valor observado del *estadístico del contraste* es $t = T(\bar{\mathbf{x}}) = 20$, calculando para los valores de m utilizados en la Tabla 4.4, los p-valores correspondientes y el ínfimo de la probabilidad a posteriori en los valores δ^* obtenidos en la Tabla 4.4. Por ejemplo, se observa que cuando $m = 10$ el p-valor es $p(10) = 0,02925$. Así, un clásico que utilizara $p^* = 0,05$ rechazaría H_0 , mientras que si utilizara $p^* = 0,01$ aceptaría H_0 . Se puede comprobar que un bayesiano obtendría éste es el mismo resultado utilizando, respectivamente, los valores $\delta^* = 7,96$ y $\delta^* = 17,01$ de la Tabla 4.3. Esto se debe a que, en este caso, $\inf_{\pi \in Q_{us}} P(H_0 | t = 20, \delta^* = 7,96) = 0,3772$ y $\inf_{\pi \in Q_{us}} P(H_0 | t = 20, \delta^* = 17,01) = 0,7343$.

4.4. Observaciones

Se puede observar que los resultados obtenidos en el capítulo 3 para tablas de contingencia $r \times s$, particularizados en el problema de contraste dado en (2.6), se pueden considerar como una aplicación del Teorema 4.2.1 cuando el muestreo se realiza sobre r poblaciones multinomiales independientes y $m = r(s - 1)$.

Además, cuando el problema estadístico consiste en contrastar

$$H_0 : \theta_0 = \theta_0(\omega), \text{ versus } H_1 : \theta_0 \neq \theta_0(\omega),$$

donde $\theta_0 : \Omega \rightarrow \Theta$, siendo $\Omega = \{\omega = (\omega_1, \dots, \omega_q), \theta(\omega) = (\theta_1(\omega), \dots, \theta_m(\omega)) \in \Theta \subset \mathbb{R}^q\}$ y $q < m$ fijo, se pueden generalizar los teoremas de la sección 4.2. Una posibilidad es utilizar la distribución a priori

$$\pi^*(\theta) = \pi_0 \pi_1(\omega) I_{H_0}(\theta) + (1 - \pi_0) \pi_2(\theta) I_{H_1}(\theta),$$

donde $\pi_1(\omega)$ y $\pi_2(\theta)$ son, respectivamente, densidades sobre Ω y Θ , y π_0 es la masa a priori asignada a la hipótesis nula, ya que en este caso, al igual que en la sección 2.7, la probabilidad a posteriori de H_0 tiene la forma de (4.5) para

$$\eta(x_1, \dots, x_n) = \frac{\int_{\Theta} f(x_1, \dots, x_n | \theta) \pi_2(\theta) d\theta}{\int_{\Omega} f(x_1, \dots, x_n | \theta_0(\omega)) \pi_1(\omega) d\omega}.$$

Otras dos posibilidades serían, respectivamente, abordar este problema mediante razonamientos similares a los de las secciones 2.7.3 y 2.7.4.

4.5. Conclusiones y Comentarios

La conclusión más importante es que los p-valores y las probabilidades a posteriori se pueden reconciliar en el problema del contraste multivariante de hipótesis nula puntual dado en (4.1).

Se propone asignar una masa a priori a θ_0 igual a la probabilidad de una bola de radio δ centrada en θ_0 , calculada mediante $\pi(\theta)$, y utilizar la distribución a priori de

tipo mixto definida en (4.2). En este caso, la probabilidad a posteriori del punto de la hipótesis nula es (4.5). Esta metodología permite desarrollar un procedimiento con el que es posible demostrar un teorema que caracteriza el acuerdo entre ambas aproximaciones en los términos de (4.8).

Además, en los ejemplos estudiados en la sección 4.3 se muestra que el acuerdo es siempre posible para un valor $\delta^* = \delta^*(p^*, n, m) = \frac{\delta\sqrt{n}}{\sigma^2}$. Esto se debe fundamentalmente a que el ínfimo de la probabilidad a posteriori sobre las clases de distribuciones consideradas respectivamente en los ejemplos mencionados es una función del *estadístico de contraste* del método clásico usual.

Referencias

- [1] BERGER, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis*. Springer, New York.
- [2] BERGER, J. O. (1994). An overview over robust Bayesian analysis (with discussion). *Test* **3**(1), 5-124.
- [3] BERGER, J. O. and BERLINER, L. M. (1986). Robust Bayes and empirical Bayes analysis with ε -contaminated priors. *Ann. Statist.*, **14**(2), 461-486.
- [4] BERGER, J. O., BOUKAI, B. and WANG, Y. (1997). Unified Frequentist and Bayesian Testing of a Precise Hypothesis. *Statistical Science*, **12**(3), 133-160.
- [5] BERGER, J. O., BOUKAI, B. and WANG, Y. (1999). Simultaneous Bayesian-Frequentist Sequential Testing of Nested Hypothesis. *Biometrika*, **86**, 79-92.
- [6] BERGER, J. O. and DELAMPADY, M. (1987). Testing Precise Hypotheses, (with Discussion). *Statistical Science*, **2**(3), 317-352.
- [7] BERGER, J. O. and SELKE, T. (1987). Testing a Point Null Hypotheses: The Irreconcilability of p-Values and Evidence, (with Discussion). *J. Amer. Statist. Assoc.*, **82**, 112-139.
- [8] BERNARDO, J. M. (1980). A Bayesian Analysis of Classical Hypothesis Testing. In: Bernardo, J.M., DeGroot, M.H., Lindley, D.V., Smith, A.F.M., (Eds.) *Bayesian Statistics*. University Press, Valencia, pp. 605-647 (with discussion).

- [9] CASELLA, G. and BERGER, R. L. (1987). Reconciling Bayesian and Frequentist Evidence in the One-Sided Testing Problem, (with Discussion). *J. Amer. Statist. Assoc.*, **82**, 106-135.
- [10] COW, D. R. and HINCKLEY, D. V. (1974). *Theoretical Statistics*. Chapman & Hall. London.
- [11] DEGROOT, M. H. (1974). Reaching a consensus. *J. Amer. Statist. Assoc.*, **68**, 966-969.
- [12] DE LA HORRA, J. and RODRÍGUEZ-BERNAL, M. T. (2003). Bayesian robustness of the posterior predictive p-value. *Commun. Statist.-Theory Meth.*, **32**, 1493-1503.
- [13] DE LA HORRA, J. (2005). Reconciling classical and prior predictive p-values in the two sided location parameter testing problem. *Aceptado en Test*.
- [14] DICKEY, J. M. and LIENZ, B. P. (1970). The weighted likelihood ratio, sharp hypothesis about chances, the order of a Markov chain. *Ann. Math. Statist.*, **41**, 214-226.
- [15] EDWARDS, W. L., LINDMAN, H. and SAVAGE L. J. (1963). Bayesian Statistical Inference for Psychological Research. *Psychol. Rev.*, **70**, 193-248.
- [16] GÓMEZ-VILLEGAS, M. A. and GÓMEZ SÁNCHEZ-MANZANO, E. (1992). Bayes Factor in Testing Precise Hypotheses. *Commun. Statist-Theory Meth.*, **21**, 1707-1715.
- [17] GÓMEZ-VILLEGAS, M. A. and GONZÁLEZ, B. (2005). Bayesian Analysis of Contingency Tables. *Aceptado en Commun. Statist-Theory Meth.*
- [18] GÓMEZ VILLEGAS, M. A., MAÍN, P. and SANZ, L. (2002). A suitable bayesian approach in testing point null hypothesis: some examples revisited. *Commun. Statist.-Theory Meth.*, **31**(2), 201-217.
- [19] GÓMEZ VILLEGAS, M. A., MAÍN, P. and SANZ, L. (2004). A Bayesian Analysis for the Multivariate Point Null Testing Problem. *Technical Report Dpto. EIO-I. Universidad Complutense de Madrid. 04-01*.

- [20] GÓMEZ VILLEGAS, M. A., MAÍN, P., SANZ, L. and NAVARRO, H. (2004). Asymptotic relationships between posterior probabilities and p-values using the hazard rate. *Statistics & Probability Letters*, **66**, 59-66.
- [21] GÓMEZ-VILLEGAS, M. A. and SANZ, L. (1998). Reconciling Bayesian and Frequentist Evidence in the Point Null Testing Problem. *Test*, **7**(1), 207-216.
- [22] GÓMEZ-VILLEGAS, M. A. and SANZ, L. (2000). ε -contaminated priors in testing point null hypothesis: a procedure to determine the prior probability. *Statistics & Probability Letters*, **47**, 53-60.
- [23] GHOSH, J. K. and MUKERJEE, R. (1992). Non-informative priors. In: Bernardo, J.M., Berger, J.O., Dawid, A.P., Smith, A.F.M. (Eds), *Bayesian Statistics*, vol. **4**. University Press, Oxford, 195-210 (with discussion).
- [24] GHOSH, J. K. and RAMAMOORTHI R. V. (2003). *Bayesian Nonparametrics*. Springer.
- [25] HOWARD, J. V. (1998). The 2×2 Table: A Discussion from a Bayesian Viewpoint. *Statistical Science*, **13**(4), 351-367.
- [26] HUBER, P. J. (1973). The use of Choquet capacities in statistics. *Bull. Int. Statist. Inst.*, **45**, 181-191.
- [27] JEFFREYS, H. (1948) *Theory of Probability*, 2nd ed. Oxford: Clarendon Press.
- [28] LEE, P. M. (1997). *Bayesian Statistics: An Introduction*. London: Arnold.
- [29] LINDLEY, D. V. (1957). A Statistical Paradox. *Biometrika*, **44**, 187-192.
- [30] LINDLEY, D. V. (1998). Statistical Inference Concerning Hardy-Weinberg Equilibrium. *Bayesian Statistics*, **3**, 307-326.
- [31] MCCULLOCH, R. E. and ROSSI, P. E. (1992). Bayes factors for non-linear hypothesis and likelihood distributions. *Biometrika*, **79**, 663-676.

- [32] MUKHOPADHYAY, S. and DASGUPTA, A. (1997). A Uniform Approximation of Bayes Solutions and Posteriors: Frequentistly Valid Bayes Inference. *Statistics and Decisions*, **15**, 51-73.
- [33] OH, M. S. (1988). A Bayes test for simple versus one-sided hypothesis on the mean vector of a multivariate normal distribution. *Commun. Statist.-Theory Meth.*, **27**(10), 2371-2389.
- [34] OH, H. S. and DASGUPTA, A. (1999). Comparison of the p-Value and Posterior Probability. *J. of Statist. Planning and Inference*, **76**, 93-107.
- [35] PEARSON, E. S. (1947). The choice of statistical tests illustrated on the interpretation of data classed in a 2×2 table. *Biometrika*, **34**, 139-167.
- [36] PEARSON, K. (1900). On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Phil. Mag.*, **5**(50), 157-175.
- [37] PRATT, J. V. (1965). Bayesian Interpretation of Standard Inference Statements. *J. Roy. Statist. Soc. B*, **27**, 169-203.
- [38] QUINTANA F. A. (1998). Nonparametric Bayesian Analysis for Assessing Homogeneity in $k \times l$ Contingency Tables With Fixed Right Margin Totals. *Journal of the American Statistical Association, Theory and Methods*, **93**(443), 1140-1149.
- [39] ROBERT, Ch. P. (2001). *The Bayesian Choice*. Springer, New York.
- [40] RUBIN, D. B. (1984). Bayesianly Justifiable and Relevant Frequency Calculations for the Applied Statistician. *Ann. Statist.*, **12**, 1151-1172.
- [41] SAVAGE, L. J. *The Foundations of Mathematics*. London: Routledge and Kegan Paul.

- [42] SIVAGANESAN, S. (1988). Range of the posterior measures for priors with arbitrary contaminations. *Commun. Statist.-Theory Meth.*, **17**, 1591-1612.
- [43] SIVAGANESAN, S. and Berger, J. O. (1989). Ranges of posterior measures for priors with unimodal contaminations. *Ann. Statist.*, **17**(2), 869-889.
- [44] SPIEGELHALTER, D. J. and SMITH, A. F. M. (1982). Bayes factors for linear and log-linear models with vague prior information. *J. Roy. Statist. Soc., Ser. B*, **44**, 377-387.