

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



ALGUNAS RELACIONES ENTRE LA APROXIMACIÓN CLÁSICA  
Y BAYESIANA EN EL CONTRASTE DE HIPÓTESIS

Luis Sanz San Miguel

21.135

Memoria para optar al grado de Doctor en  
Ciencias Matemáticas, realizada bajo la  
dirección del Dr. D. Miguel Angel Gómez  
Villegas.

Madrid, Abril 1996



Archivo

D. MIGUEL A. GÓMEZ VILLEGAS, CATEDRÁTICO DEL DEPARTAMENTO DE ESTADÍSTICA E INVESTIGACIÓN OPERATIVA DE LA UNIVERSIDAD COMPLUTENSE DE MADRID.

CERTIFICO:

Que la presente memoria, titulada:

ALGUNAS RELACIONES ENTRE LA APROXIMACIÓN CLÁSICA  
Y BAYESIANA EN EL CONTRASTE DE HIPÓTESIS

ha sido realizada bajo mi dirección por el Licenciado en Ciencias Matemáticas D. Luis Sanz San Miguel y constituye su Tesis para optar al grado de Doctor en Ciencias Matemáticas.

Y para que conste, en cumplimiento de la legislación vigente y a los efectos que haya lugar, firmo la presente en Madrid a 23 de Abril de 1996.

A handwritten signature in black ink, appearing to read 'Miguel Gómez', with a long horizontal line extending to the left from the start of the signature.

A Carlos, Jorge y Clara.

# Prefacio

La relación entre las respuestas clásica y bayesiana a los problemas de la Inferencia Estadística viene siendo, desde hace años, ampliamente tratada por numerosos autores.

En gran número de ocasiones se ha conseguido, bajo determinadas condiciones, lograr acuerdo entre ambas aproximaciones. No obstante, persisten algunos contextos en los que la discrepancia de las dos posturas se mantiene.

En el terreno de los contrastes de hipótesis paramétricos, tenemos las dos situaciones. Si el contraste es de tipo unilateral, se ha logrado reconciliar ambas posturas. Mientras que si el contraste es de hipótesis nula puntual, se ha puesto de manifiesto, hasta ahora y por parte de numerosos autores, que hay discrepancia entre la aproximación clásica y la bayesiana.

En nuestro trabajo, presentamos una metodología basada en la relación entre el contraste de hipótesis nula puntual y el contraste de hipótesis nula de tipo intervalo que permite, en determinadas circunstancias, aproximar ambas posturas, clásica y bayesiana, cuando el contraste es de hipótesis nula puntual.

Quiero agradecer, sinceramente, la ayuda que con sus consejos, sugerencias y aportaciones me ha proporcionado el Profesor Miguel A. Gómez Villegas, bajo cuya dirección ha sido realizado este trabajo.

Mi agradecimiento, también a los profesores del Departamento de Estadística e Investigación Operativa de la Facultad de Matemáticas de la U.C.M. por su apoyo y estímulo a lo largo de estos años y, especialmente, quisiera agradecer a M.D. Esteban, E. Molina y A. Alonso su inestimable ayuda con el  $\text{\LaTeX}$ .

# Índice

|          |   |           |
|----------|---|-----------|
| <b>1</b> | <b>Introducción</b>   | <b>2</b>  |
| 1.1      | Antecedentes . . . . .  | 2         |
| 1.2      | Desarrollo de la memoria . . . . .                                    | 6         |
| <b>2</b> | <b>Contraste de Hipótesis. Resultados anteriores.</b>                 | <b>9</b>  |
| 2.1      | Introducción . . . . .  | 9         |
| 2.2      | Contraste de hipótesis unilateral . . . . .                           | 14        |
| 2.3      | Contraste de hipótesis nula puntual . . . . .                         | 18        |
| 2.4      | Contraste de hipótesis de tipo intervalo . . . . .                    | 23        |
| <b>3</b> | <b>Distribuciones iniciales unimodales y simétricas</b>               | <b>30</b> |
| 3.1      | Introducción . . . . .  | 30        |
| 3.2      | Ínfimo de la probabilidad final. . . . .                              | 33        |
| 3.3      | Comparación entre las evidencias clásica y bayesiana . . . . .        | 38        |
| <b>4</b> | <b>Distribuciones iniciales <math>\varepsilon</math>-contaminadas</b> | <b>51</b> |
| 4.1      | Introducción . . . . .  | 51        |
| 4.2      | Contaminaciones arbitrarias . . . . .                                 | 56        |
| 4.3      | Contaminaciones unimodales . . . . .                                  | 62        |
| 4.4      | Contaminaciones unimodales y simétricas . . . . .                     | 68        |
|          | <b>Bibliografía</b>   | <b>73</b> |

# Capítulo 1

## Introducción

### 1.1 Antecedentes

En el problema del contraste de hipótesis se puede pensar que la evidencia es una evaluación acerca de si los datos obtenidos de la observación experimental apoyan o no la hipótesis nula. Ésta, en general, consistirá en una afirmación que involucre un parámetro o alguna forma funcional no conocida de la distribución de interés, distribución de la que se han obtenido las observaciones.

Para un frecuentista la evidencia se concretará en el  $p$ -valor, que es la máxima probabilidad, supuesto que la hipótesis nula es cierta, de obtener valores más alejados de la hipótesis nula que los datos observados. Para un bayesiano, en cambio, la evidencia toma la forma de la probabilidad final de la hipótesis nula. Debe observarse que, aunque las aproximaciones clásica y bayesiana a los problemas del contraste de hipótesis y a los de la inferencia estadística en general son diferentes, se pueden, en numerosas situaciones, obtener relaciones que permiten adoptar la misma decisión desde ambos puntos de vista, bien sea para aceptar la hipótesis nula o bien para rechazarla.

La cuestión de comparar, en los contrastes de hipótesis, las medidas de evidencia frecuentistas, expresadas fundamentalmente en términos de los  $p$ -valores, y las bayesianas, expresadas mediante el factor Bayes o a través de la probabilidad final de la hipótesis nula, ha sido ampliamente tratada en la literatura.

Si la hipótesis nula es puntual, es decir consta de un único punto, ya 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 contra la hipótesis nula. En otras palabras, que el  $p$ -valor es (mucho) más pequeño que la probabilidad final de la hipótesis nula. Referencias a este trabajo de Lindley se pueden encontrar en Shafer(1982). Muchos autores han analizado dicha relación eligiendo distribuciones iniciales concretas. Este es el caso, además del ya citado de Lindley, de Jeffreys (1957, 1961), Good(1965, 1967, 1985, 1986), Lindley(1961, 1965, 1977), Edwards, Lindmann y Savage(1963), Smith(1965), Zellner(1971, 1984), Dickey (1971, 1973, 1974), Smith y Spiegelhalter(1980), Zellner y Siow(1980) y Gómez Villegas y De la Horra(1984).

La elección de la distribución inicial es uno de los aspectos problemáticos dentro del análisis bayesiano. Determinar una distribución de probabilidad que recoja nuestras opiniones a priori acerca del comportamiento del parámetro no es, en la práctica, sencillo ya que no resulta fácil discernir entre distribuciones que presenten características similares como unimodalidad, simetría, etc.. Es por ello que resulta interesante la consideración de clases de distribuciones iniciales. En estas clases se encontrarán todas aquellas distribuciones de probabilidad que, razonablemente y con cierta amplitud, recojan las opiniones iniciales acerca del parámetro o parámetros objeto del análisis. Edwards, Lindmann y Savage(1963) fueron los primeros que se aproximaron al problema de la relación entre el  $p$ -valor y la evidencia bayesiana utilizando una clase de distribuciones iniciales y calculando cotas inferiores para el factor Bayes y la probabilidad final de la hipótesis nula puntual sobre dicha clase. Su conclusión es que hay fuerte discrepancia entre los  $p$ -valores y las probabilidades finales.

Dickey(1977) consideró diferentes clases de distribuciones iniciales y examinó el ínfimo del factor Bayes sobre cada una de las clases. También concluyó que el p-valor exagera la evidencia contra la hipótesis nula incluso comparado con el ínfimo de las medidas de evidencia bayesianas. Análogamente Berger y Sellke(1987) utilizan diferentes clases de distribuciones iniciales para calcular, en este caso, el ínfimo de las probabilidades finales de la hipótesis nula sobre cada una de las clases. Ellos encuentran que ese ínfimo es mucho mayor que el correspondiente p-valor, lo que les lleva a concluir que los p-valores pueden ser medidas muy engañosas de la evidencia contra la hipótesis nula.

En cambio De Groot(1973), en un marco muy general, construye distribuciones alternativas y encuentra distribuciones iniciales impropias para las cuales el p-valor y la probabilidad final de la hipótesis nula son iguales. El supone que las distribuciones alternativas están estocásticamente ordenadas y eso, aunque no sea de una manera explícita, le coloca esencialmente en el problema del contraste de hipótesis unilateral. Otros autores que también realizan comparaciones entre la probabilidad final de la hipótesis nula y el p-valor son Pratt(1965), Dempster(1973), Bernardo(1980), Shafer(1982) y Good(1984).

En particular, Pratt(1965) afirma que en el contraste unilateral el p-valor puede ser aproximadamente igual a la probabilidad final de la hipótesis nula. En esta misma línea, Casella y Berger(1987) prueban que, utilizando diferentes clases de distribuciones iniciales, el ínfimo de la probabilidad final bayesiana es igual al p-valor e incluso, en determinados casos, ese ínfimo es menor que el correspondiente p-valor con lo que concluyen que en el problema del contraste unilateral aquel puede ser considerado como una aproximación a las medidas de evidencia bayesianas. Bernardo(1980) lleva a cabo un estudio alternativo del problema del contraste de hipótesis puntual más próximo a la teoría de la decisión al imponer la maximización de la información faltante. Su conclusión es que es posible establecer una relación uno a uno entre los p-valores y las probabilidades finales de la hipótesis nula.

Algunos trabajos que consideran cotas inferiores para el factor Bayes o para la

probabilidad final de la hipótesis nula son Berger(1986), Berger y Delampady(1987), Moreno y Cano(1989), Delampady(1989a, 1989b) y Delampady y Berger(1990).

Moreno y Cano(1989) tratan el caso de parámetro multidimensional y demuestran que, para una amplia clase de modelos muestrales y clases razonables de distribuciones iniciales, los p-valores no son necesariamente menores que la probabilidad final de la hipótesis puntual. En la misma situación, Delampady y Berger(1990) prueban que si hay simetría en las verosimilitudes, la discrepancia se mantiene para diferentes clases de distribuciones iniciales.

Delampady(1989b), Berger y Delampady(1987) además de Berger(1985) estudian la relación entre los contrastes de hipótesis nula puntual y de hipótesis nula de intervalo. Una vez más concluyen que hay discrepancia entre las respuestas clásica y bayesiana. En este sentido, el trabajo de Gómez Villegas y Gómez Sánchez-Manzano(1992) pone de manifiesto que cuando el intervalo es suficientemente pequeño ambos problemas, el puntual y el de intervalo, conducirían a tomar la misma decisión.

Como ya se ha comentado, parece adecuado considerar clases de distribuciones iniciales para efectuar un análisis bayesiano del contraste de hipótesis. A la hora de elegir una clase deben tenerse en cuenta ciertas consideraciones que permitan un desarrollo correcto del análisis correspondiente. Berger(1994) indica las propiedades que, idealmente, debería poseer una clase de distribuciones iniciales: sencillez en la especificación y en la interpretación, computacionalmente sencilla, que represente nuestras opiniones iniciales y que sea flexible.

Una clase de distribuciones iniciales, que responde a esas propiedades y que está siendo muy empleada, sobre todo en estudios de robustez bayesiana, es la clase de las distribuciones  $\epsilon$ -contaminadas. Estudios de robustez en los que se utilizan distribuciones de este tipo se pueden encontrar, entre otros, en Berger(1985), Berger y Berliner(1986), Sivaganesan(1988), Sivaganesan y Berger(1989) y Moreno y Cano(1991). Berger(1994) propone varias clases de contaminaciones para efectuar

un análisis bayesiano robusto. Por otra parte, Berger y Mortera(1994) estudian la robustez bayesiana en el contraste de hipótesis utilizando diferentes técnicas de aproximación para problemas en los que hay involucrados parámetros perturbadores.

## 1.2 Desarrollo de la memoria

En la presente memoria se aborda el estudio de la relación entre las evidencias clásica y bayesiana en el contraste de hipótesis. Más concretamente, el objetivo fundamental es mostrar que puede haber una reconciliación entre las dos posturas, clásica y bayesiana, en el contraste de hipótesis nula puntual en el caso de parámetro de localización unidimensional.

En el capítulo 2 se plantea, en forma resumida, el problema general del contraste de hipótesis tanto desde el punto de vista clásico como desde el punto de vista bayesiano. Asimismo se analizan con más detalle los contrastes unilaterales, de hipótesis nula puntual y de hipótesis nula de tipo intervalo. En el contraste de hipótesis unilateral se presentan los resultados obtenidos, fundamentalmente los de Casella y Berger(1987), en orden a reconciliar los puntos de vista clásico y bayesiano en este tipo de contraste. En concreto, ellos prueban que el ínfimo de la probabilidad final de la hipótesis nula es igual o menor que el p-valor. En cuanto al contraste de hipótesis nula puntual se ofrece un resumen de los resultados obtenidos hasta ahora, resultados que muestran una fuerte discrepancia entre la respuesta clásica, medida siempre en términos del p-valor, y la bayesiana, concretada bien en el factor Bayes bien en la probabilidad final de la hipótesis nula puntual. En este capítulo se estudia, también, la relación entre el contraste de hipótesis nula puntual y el contraste de hipótesis nula de tipo intervalo y, en particular, cuándo un contraste se puede aproximar por el otro de manera razonable. Aquí, la idea básica es que el contraste de hipótesis nula puntual raramente se plantea en situaciones prácticas y

que, por tanto, sería más razonable efectuar un contraste del tipo

$$H_0 : \theta \in I_b \quad \text{frente a} \quad H_1 : \theta \in I_b$$

donde  $I_b = (\theta_0 - b, \theta_0 + b)$  y  $b > 0$  debe ser elegido de tal manera que los  $\theta \in I_b$  sean "indistinguibles" de  $\theta_0$ .

En el capítulo 3 abordamos el problema del contraste de hipótesis nula puntual desde el punto de vista bayesiano y su relación con la respuesta clásica. Nuestro principal resultado es el teorema 3.1 que establece el ínfimo de la probabilidad final de la hipótesis nula puntual sobre la clase de distribuciones iniciales unimodales y simétricas,  $G_{US}$ . Esta clase es una de las utilizadas tanto por Casella y Berger(1987) como por Berger y Sellke(1987), los primeros para mostrar la reconciliación entre las aproximaciones clásica y bayesiana en el contraste unilateral y los segundos para mostrar la discrepancia en el contraste puntual.

La novedad radica en que podemos conseguir que este ínfimo sea aproximadamente igual al p-valor correspondiente a la observación realizada. Con lo que se puede concluir, en contraste con los resultados obtenidos hasta ahora, que el p-valor es una medida que puede estar dentro del rango de las medidas de evidencia bayesianas en el contraste de hipótesis nula puntual. Se comparan los p-valores con la cota inferior de la probabilidad final obtenida cuando las observaciones provienen de diferentes modelos muestrales. En particular, analizamos los casos Normal, Cauchy y doble exponencial que se corresponden, respectivamente, con modelos resistentes, propensos y neutros a datos atípicos (ver Maín, 1988).

Finalmente, en el capítulo 4, llevamos a cabo un amplio estudio con el objetivo de conseguir cotas inferiores de la probabilidad final de la hipótesis nula puntual utilizando la clase de distribuciones iniciales  $\varepsilon$ -contaminadas y comparamos estas cotas con el correspondiente p-valor en forma similar a como se hizo en el capítulo 3.

Con la clase de las distribuciones  $\varepsilon$ -contaminadas, como distribuciones iniciales,

introducimos tres diferentes clases de contaminación: (a) Todas las distribuciones; (b) Todas las distribuciones que son unimodales y (c) Todas las distribuciones unimodales y simétricas. Además de estas, en la literatura han sido tratadas otras clases de contaminaciones como describimos con más detalle en la introducción del capítulo 4. En cualquier caso, en los tres supuestos analizados, se llega a la conclusión de que se puede considerar que el  $p$ -valor queda dentro del rango de las medidas de evidencia bayesianas. En todos los casos se presentan ejemplos y se comentan los resultados.

Se puede, de esta forma, concluir que la discrepancia que otros autores han observado entre las aproximaciones clásica y bayesiana en el contraste de hipótesis nula puntual no se debe a la necesidad de utilizar, como se había argumentado hasta ahora, una distribución inicial mixta. El uso de otras clases de distribuciones iniciales diferentes a las utilizadas en esta memoria es un problema que queda abierto para posteriores investigaciones así como la utilización de clases de distribuciones para el modelo de muestreo.

## Capítulo 2

# Contraste de Hipótesis. Resultados anteriores.

### 2.1 Introducción

Una hipótesis estadística consiste en una afirmación acerca de un parámetro de la población objeto de estudio. Lo que se pretende al formular una hipótesis es conocer si es verdadera o falsa. Habitualmente, la hipótesis de la que se pretende averiguar si es verdadera o falsa, y que se denomina hipótesis nula, se enfrenta con la afirmación contraria a la que se ha recogida en ella. Esta última recibe el nombre de hipótesis alternativa y engloba, como su nombre indica, todas las posibles alternativas a la hipótesis nula.

De manera que un problema de contraste de hipótesis tiene, en la mayoría de las ocasiones, la forma siguiente:

Sea  $\theta$  un parámetro desconocido, posiblemente un vector, del que se sabe que pertenece a un espacio paramétrico,  $\Theta$ , y queremos determinar si  $\theta \in \Theta_0$  o  $\theta \in \Theta_1$ ,

donde  $\Theta_0 \cup \Theta_1 = \Theta$  y  $\Theta_0 \cap \Theta_1 = \emptyset$ .

Se definirá  $H_0$  como la hipótesis de que  $\theta \in \Theta_0$  y  $H_1$  como la hipótesis de que  $\theta \in \Theta_1$ . A  $H_0$  se le denomina *hipótesis nula* y a  $H_1$  *hipótesis alternativa*.

$$H_0 : \theta \in \Theta_0 \text{ frente a } H_1 : \theta \in \Theta_1. \quad (2.1)$$

Un test de hipótesis consiste en cualquier procedimiento que permita decidir si se acepta o se rechaza la hipótesis nula, entendiendo que rechazar la hipótesis nula equivale a aceptar la hipótesis alternativa. Antes de decidir qué hipótesis aceptar, se observa una muestra aleatoria simple,  $X_1, \dots, X_n$ , de la población que supondremos con función de masa o de densidad,  $f(x|\theta)$ , dependiente de  $\theta$ .

Consideraremos, por tanto, un problema de contraste de hipótesis como un problema en el que una de dos acciones será tomada: aceptar  $H_0$ , lo que equivale a rechazar  $H_1$ , o aceptar  $H_1$ , que equivaldrá a rechazar  $H_0$ .

Desde el punto de vista clásico, un test de hipótesis se evalúa a partir de las probabilidades de error de tipo 1 y de tipo 2. Se dice que se ha cometido un *error de tipo 1* si se rechaza  $H_0$  cuando, de hecho,  $H_0$  es cierta. Mientras que si se acepta  $H_0$  cuando, de hecho,  $H_1$  es cierta se dice que se ha cometido un *error de tipo 2*. Habitualmente, lo que se hace es construir la región crítica, conjunto de observaciones que conducen al rechazo de  $H_0$ , de tal manera que la probabilidad de error de tipo 2 sea lo más pequeña posible sujeta a la restricción de que la probabilidad de error de tipo 1 esté acotada por un valor previamente fijado,  $\alpha_0$ , que se conoce con el nombre de *nivel de significación* y que, normalmente, será un valor pequeño. Es corriente utilizar los valores 0.01, 0.05 y 0.1.

Debemos señalar que cuando se fija el nivel de significación de un contraste se está controlando únicamente la probabilidad de error de tipo 1 y que, por tanto, si se utiliza esta aproximación, las hipótesis deberían formularse teniendo en cuenta que lo que realmente se va a controlar es la probabilidad de *error de tipo 1* y no la de *tipo 2*.

Por otro lado, es interesante considerar el *tamaño*,  $\alpha$ , de un contraste concreto, que es la máxima probabilidad de tomar una decisión incorrecta entre todos los valores de  $\theta$  que satisfacen la hipótesis nula. Es decir,

$$\alpha = \sup_{\theta \in \Theta_0} Pr\{\text{rechazar } H_0 | \theta\}. \quad (2.2)$$

Nótese que  $\alpha \leq \alpha_0$ , siendo  $\alpha_0$  el nivel de significación.

El tamaño de un test aporta información importante, puesto que si es “pequeño” la decisión de rechazar  $H_0$  resulta convincente. En cambio, si el tamaño es “grande” la decisión de rechazar  $H_0$  puede no ser muy convincente ya que el contraste tiene una probabilidad alta de haber tomado la decisión errónea.

Otra medida clásica de evidencia contra la hipótesis nula,  $H_0$ , que informa de los resultados de un contraste y que depende de las observaciones realizadas es el p-valor. El p-valor mide el valor crítico correspondiente a la muestra observada, valor en el cual la muestra se encuentra en la frontera de rechazar o no rechazar.

Si se desea contrastar (2.1) y se ha observado  $X = x$ , el p-valor correspondiente a esta observación,  $p(x)$ , se define como

$$p(x) = \sup_{\theta \in \Theta_0} Pr\{T(X) \geq T(x) | \theta\}, \quad (2.3)$$

donde  $T(X)$  es un estadístico apropiado para el problema que se esté tratando. Observamos que el p-valor también depende de la hipótesis nula con la que se esté trabajando, así la región crítica que hemos utilizado en (2.3) se corresponderá, verosimilmente, con una hipótesis nula unilateral.

Si la observación muestral,  $x$ , está en la región crítica para un contraste de tamaño  $\alpha$  estará, asimismo, en la región crítica de cualquier contraste de tamaño  $\alpha' \geq \alpha$ . Entonces, el p-valor correspondiente a la observación  $x$  se puede interpretar como el valor más pequeño de  $\alpha$  para el que la observación de  $x$  conduciría a rechazar  $H_0$ . Ahora bien, conviene aclarar que el p-valor,  $p(x)$ , no es un nivel de significación y que no debe ser interpretado como tal.

Cuanto más pequeño sea el  $p$ -valor más fuerte será la evidencia muestral contra la hipótesis nula. En otras palabras, cuanto más pequeño sea  $p(x)$  las observaciones indicarán que hay evidencia de que la hipótesis cierta es  $H_1$ .

Desde el punto de vista bayesiano, para decidir entre  $H_0$  y  $H_1$ , basta con calcular las probabilidades finales

$$a_0 = Pr \{ \theta \in \Theta_0 | x \} \quad \text{y} \quad a_1 = Pr \{ \theta \in \Theta_1 | x \}, \quad (2.4)$$

y decidir entre  $H_0$  y  $H_1$  de acuerdo con esas probabilidades.

Si bien las probabilidades finales (2.4) son las medidas bayesianas utilizadas en primera instancia en el contraste de hipótesis, otros conceptos, que a continuación introducimos, también tienen interés.

Si llamamos  $\pi_0$  y  $\pi_1$  a las probabilidades iniciales de  $\Theta_0$  y  $\Theta_1$  respectivamente,

**Definición 2.1** *Llamaremos cociente de probabilidades iniciales de  $H_0$  contra  $H_1$  al cociente  $\pi_0/\pi_1$ . Análogamente, llamaremos cociente de probabilidades finales de  $H_0$  contra  $H_1$  al cociente  $a_0/a_1$ .*

Obsérvese que si nuestro cociente de probabilidades iniciales es próximo a la unidad significa que consideramos que  $H_0$  y  $H_1$  tienen inicialmente (casi) la misma probabilidad de ocurrir. Si este cociente de probabilidades es grande significa que consideramos  $H_0$  (mucho) más probable que  $H_1$  y si el cociente es pequeño ( $< 1$ ) es que consideramos que, inicialmente,  $H_0$  es menos probable que  $H_1$ .

**Definición 2.2** *El factor Bayes en favor de  $H_0$  contra  $H_1$  se define como*

$$B(x) = \frac{a_0/a_1}{\pi_0/\pi_1} = \frac{f(x|\Theta_0)}{f(x|\Theta_1)}.$$

El interés del factor Bayes está en que proporciona una medida de si los datos aumentan o disminuyen los cocientes de probabilidades de  $H_0$  respecto de  $H_1$ . Así,

si  $B(x) > 1$  quiere decir que, dado  $x$ ,  $H_0$  es ahora relativamente más plausible que  $H_1$ ; si  $B(x) < 1$  quiere decir que, dado  $x$ , ha aumentado la plausibilidad relativa de  $H_1$ .

Debe observarse que, puesto que  $\pi_1 = 1 - \pi_0$  y  $a_1 = 1 - a_0$ , la probabilidad final de  $H_0$  se puede obtener a partir de su probabilidad inicial y del factor Bayes. En efecto,

$$a_0 = \left( 1 + \frac{1 - \pi_0}{\pi_0} \frac{1}{B(x)} \right)^{-1}.$$

En ocasiones, el factor Bayes puede interpretarse como el cociente de verosimilitudes de  $H_0$  contra  $H_1$ . Esta interpretación es claramente válida en el caso de hipótesis simples,  $\Theta_0 = \{\theta_0\}$  y  $\Theta_1 = \{\theta_1\}$  puesto que entonces

$$B(x) = \frac{f(x|\theta_0)}{f(x|\theta_1)},$$

de manera que el factor Bayes, en este caso, no depende de las probabilidades iniciales.

En el caso general, en cambio,  $B(x)$  dependerá de la distribución inicial elegida, puesto que si se trata de contrastar  $H_0$  frente a  $H_1$  como en (2.1), obtenemos (Berger, 1985)

$$B(x) = \frac{\int_{\Theta_0} f(x|\theta) \frac{\pi(\theta)}{Pr(\Theta_0)} d\theta}{\int_{\Theta_1} f(x|\theta) \frac{\pi(\theta)}{Pr(\Theta_1)} d\theta},$$

donde  $\pi(\theta)/Pr(\Theta_i)$  es la densidad inicial truncada en  $\Theta_i$  ( $i = 0, 1$ ). De manera que el factor Bayes se puede interpretar como la razón ponderada de las verosimilitudes de  $H_0$  a  $H_1$  donde la ponderación es la densidad inicial truncada a  $\Theta_i$ .

Como ya hemos dicho, al estar involucrados los cocientes  $\pi(\theta)/Pr(\Theta_i)$ , el factor Bayes no puede ser considerado como una medida relativa de las verosimilitudes proporcionada únicamente por los datos, excepto si ambas hipótesis son simples.

En ocasiones, sin embargo,  $B(x)$  es relativamente poco sensible a determinadas (y razonables) elecciones de la distribución inicial. En estos casos la interpretación como cociente de verosimilitudes es razonablemente correcta. La principal ventaja operativa de tener un factor Bayes que dependa poco de la distribución inicial es que si un informe incluye este factor, entonces uno puede determinar su propio cociente de probabilidades finales simplemente multiplicando el factor Bayes por su propio cociente de probabilidades iniciales.

Una revisión del significado del factor Bayes, así como una calibración análoga a la del nivel de significación puede verse en Kass y Raftery(1995).

## 2.2 Contraste de hipótesis unilateral

Se dice que un contraste de hipótesis es de tipo unilateral si  $\Theta \subseteq \mathcal{R}$  y  $\Theta_0$  está enteramente a un lado de  $\Theta_1$ .

En este caso el uso de los p-valores tiene una justificación bayesiana como se muestra en el siguiente ejemplo

**Ejemplo 2.1.** Sea  $X$  una variable aleatoria con distribución  $N(\theta, \sigma^2)$ ,  $\sigma^2$  conocida y supongamos  $\pi(\theta) = 1$ . La distribución final,  $\pi(\theta|x)$ , en este caso es  $N(x, \sigma^2)$ .

En estas condiciones, consideremos el contraste

$$H_0 : \theta \leq \theta_0 \quad \text{frente a} \quad H_1 : \theta > \theta_0. \quad (2.5)$$

La probabilidad final de la hipótesis nula es

$$Pr(H_0|x) = Pr(\theta \leq \theta_0|x) = \Phi\left(\frac{\theta_0 - x}{\sigma}\right),$$

siendo  $\Phi$  la función de distribución de la variable aleatoria  $N(0, 1)$ .

En cuanto al p-valor contra  $H_0$  resulta

$$p(x) = Pr(X \geq x | \theta_0) = 1 - \Phi\left(\frac{x - \theta_0}{\sigma}\right) = \Phi\left(\frac{\theta_0 - x}{\sigma}\right)$$

que coincide con la probabilidad final de la hipótesis nula. No obstante, conviene observar que en la expresión bayesiana la variable aleatoria es  $\theta_0$  mientras que en la correspondiente al p-valor la variable aleatoria es  $x$ .  $\square$

El uso en el ejemplo anterior de una distribución inicial no informativa, que da masa infinita a cada una de las hipótesis, puede justificarse si se dispone de poca o nula información inicial. Además, cualquier distribución inicial que sea aproximadamente constante sobre el intervalo  $(\theta_0 - 2\sigma, \theta_0 + 2\sigma)$  (suponiendo  $x > \theta_0$ ) y que no sea significativamente grande fuera del mismo nos proporcionará una probabilidad final aproximadamente igual al p-valor.

Consideremos un contraste unilateral del tipo (2.5). Supongamos que  $X$  se distribuye mediante  $f(x|\theta)$ , con  $\theta$  parámetro de localización. Casella y Berger(1987) muestran que en este caso es posible reconciliar la medida de evidencia bayesiana, concretada en la probabilidad final de la hipótesis nula, con la medida clásica del p-valor. Si bien es preciso hacer notar que la probabilidad de que  $H_0$  ocurra no tiene sentido dentro de la teoría clásica.

En concreto, si pretendemos efectuar el contraste (2.5) y suponemos que se ha observado  $X = x$ , el p-valor correspondiente,  $p(x)$ , será

$$p(x) = Pr(X \geq x | \theta = \theta_0) = \int_x^{\infty} f(t - \theta_0) dt. \quad (2.6)$$

Si  $p(x)$  es “pequeño” la hipótesis nula sería rechazada, mientras que si  $p(x)$  es “grande” sería aceptada.

Sea  $\pi(\theta)$  una distribución inicial para el parámetro  $\theta$ . La probabilidad final de  $H_0$  será

$$Pr(H_0 | \dot{x}) = \int_{-\infty}^{\theta_0} \pi(\theta | \dot{x}) d\theta = \frac{\int_{-\infty}^{\theta_0} f(x - \theta) \pi(\theta) d\theta}{\int_{-\infty}^{\infty} f(x - \theta) \pi(\theta) d\theta}. \quad (2.7)$$

Se trata, entonces, de comparar estas dos medidas de evidencia, (2.6) y (2.7), para unas determinadas clases de distribuciones iniciales que dan el mismo peso a la hipótesis nula que a la hipótesis alternativa. Las clases utilizadas por Casella y Berger(1987) son:

$$G_A = \{\text{Todas las distribuciones que dan masa } 1/2 \text{ a } H_0 \text{ y } 1/2 \text{ a } H_1\}.$$

$$G_S = \{\text{Todas las distribuciones simétricas en torno a } \theta_0\}.$$

$$G_{US} = \{\text{Todas las distribuciones unimodales y simétricas en torno a } \theta_0\}.$$

$$G_{NOR} = \{N(\theta_0, \tau^2), 0 < \tau^2 < \infty\}.$$

Como medida de evidencia bayesiana toman el ínfimo de la probabilidad final de  $H_0$  sobre la clase de distribuciones iniciales elegida,  $G$ ,

$$\underline{Pr}(H_0|x, G) = \inf_{\pi \in G} Pr(H_0|x). \quad (2.8)$$

En todos los casos se considera  $x > \theta_0$  puesto que si  $x < \theta_0$  resulta  $p(x) > 1/2$  y ni desde el punto de vista clásico ni desde el punto de vista bayesiano se podría decir que los datos proporcionan evidencia contra  $H_0$ .

Si la clase de distribuciones iniciales elegida es  $G_A$  entonces, cualquiera que sea la función  $f(x - \theta)$ , resulta  $\underline{Pr}(H_0|x, G_A) = 0$  que es, por tanto, menor o igual que el p-valor. En los demás casos se considera que  $f$  es simétrica y que tiene cociente de verosimilitudes monótono (CVM).

**Definición 2.3** *Se dice que una variable aleatoria,  $X$ , con función de densidad  $f(x|\theta)$ ,  $\theta$  parámetro desconocido, tiene cociente de verosimilitudes monótono (CVM) si para cada  $\theta_1 > \theta_2$  el cociente  $f(x|\theta_2)/f(x|\theta_1)$  es una función no decreciente en el conjunto  $\{x, f(x|\theta_1) > 0 \text{ ó } f(x|\theta_2) > 0\}$ . ( $c/0$  se define  $\infty$  si  $c > 0$ ).*

Distribuciones con CVM son, entre otras, la distribución binomial, la distribución de Poisson y la normal de media desconocida y varianza conocida.

Además, se utiliza el hecho de que el ínfimo de las probabilidades de la hipótesis nula es el mismo si éste se toma sobre una clase de distribuciones iniciales,  $G$ , o sobre la clase  $G_M$  de todas las mixturas de  $G$ . Es decir,

$$\underline{Pr}(H_0|x, G_M) = \underline{Pr}(H_0|x, G) \quad (2.9)$$

**Definición 2.4** Sea  $G = \{\pi_\alpha, \alpha \in \mathcal{A}\}$  una clase de distribuciones iniciales sobre la recta real. Se dice que  $G_M$  es el conjunto de todas las mixturas de  $G$  si, dado  $B$  conjunto medible y  $P$  alguna medida de probabilidad sobre  $\mathcal{A}$ , todo elemento  $\pi$  de  $G_M$  se puede escribir

$$\pi(B) = \int_{\mathcal{A}} \pi_\alpha(B) dP(\alpha).$$

En estas condiciones, es decir tomando  $f$  simétrica y con CVM se obtiene que el p-valor,  $p(x)$ , es una cota superior para el ínfimo de la probabilidad final de  $H_0$ . Es más, si  $f$  no es simétrica como es el caso, por ejemplo, de la distribución exponencial o  $f$  no tiene CVM, como es el caso de la distribución de Cauchy, dicha acotación es estricta.

Si la distribución inicial es de la clase  $G_S$  o de la clase  $G_{US}$  el ínfimo de la probabilidad final se alcanza cuando la distribución inicial considerada es impropia, es decir para la medida de Lebesgue en  $(-\infty, +\infty)$ .

Ahora bien, si  $f$  no es simétrica, como es el caso de la distribución exponencial, o no tiene CVM, como la distribución de Cauchy, el ínfimo puede alcanzarse para una distribución inicial propia.

En cuanto a la clase  $G_{NOR}$ , al ser una subclase de  $G_{US}$  vale lo obtenido para ésta. No obstante, Casella y Berger(1987) encuentran que si  $f$  es simétrica y tiene CVM el p-valor es igual al ínfimo de la probabilidad final de  $H_0$  tomado sobre la clase de todas las transformaciones por escala de una densidad acotada simétrica y unimodal (nótese que  $G_{NOR}$  es de este tipo).

Para finalizar este apartado, cabe señalar que si el contraste unilateral es del tipo

$H_0 : \theta = \theta_0$  frente a  $H_1 : \theta > \theta_0$ , Berger y Sellke(1987) afirman que el p-valor y el ínfimo de la probabilidad final de  $H_0$  son completamente diferentes. Ellos obtienen, por ejemplo, que si  $X$  es una variable aleatoria que sigue una distribución  $N(\theta, 1)$  y se observa  $x = 1.645$  entonces el p-valor es 0.05 mientras que para la clase  $G_A$  resulta  $\underline{P}_T(H_0|x, G_A) = 0.21$ . Debemos hacer notar, sin embargo, que en este caso Berger y Sellke(1987) asignan masa 1/2 al punto  $\theta = \theta_0$ , lo que nos parece que es sesgar demasiado en favor de la hipótesis nula  $H_0$ .

### 2.3 Contraste de hipótesis nula puntual

En Estadística clásica es corriente encontrarse con el problema de contrastar si el parámetro,  $\theta$ , toma un valor concreto, esto es

$$H_0^* : \theta = \theta_0 \quad \text{frente a} \quad H_1^* : \theta \neq \theta_0 \quad (2.10)$$

Tal contraste, llamado de hipótesis nula puntual, es interesante por muchas razones. Para nosotros lo es, principalmente, porque las conclusiones hasta ahora obtenidas desde un enfoque bayesiano son completamente diferentes de aquellas que se obtienen desde el punto de vista clásico o frecuentista.

Lindley(1957), en lo que se ha conocido posteriormente como paradoja de Lindley, muestra la discrepancia entre las respuestas clásica y bayesiana al efectuar el contraste (2.10) suponiendo que las observaciones provienen de una distribución normal de varianza conocida. El asigna probabilidad inicial  $\pi_0$  a la hipótesis nula y distribuye el resto,  $1 - \pi_0$ , uniformemente sobre algún intervalo centrado en  $\theta_0$ . Su conclusión fundamental es que, sea cual sea el valor que se asigne a  $\pi_0$ , se puede encontrar un tamaño muestral,  $n$ , que depende de  $\pi_0$  y del nivel de significación del contraste, de tal manera que desde la óptica clásica se puede rechazar  $H_0^*$  mientras que desde el punto de vista bayesiano deberíamos aceptar  $H_0^*$ . Esa es la paradoja. Un estudio actualizado de la misma puede verse, como ya hemos dicho, en Shafer(1982).

Si suponemos que, desde el punto de vista clásico, el contraste se apoya en el valor de un determinado estadístico,  $T(X)$ , de tal manera que valores extremos de éste son considerados como evidencia contra la hipótesis nula,  $H_0^*$ , entonces el  $p$ -valor de los datos observados,  $p(x)$ , será

$$p(x) = Pr \{ |T(X)| \geq |T(x)| | \theta = \theta_0 \}, \quad (2.11)$$

siendo  $T(x)$  el valor del estadístico cuando  $X = x$ .

De manera que si  $p(x)$  es “pequeño” rechazaríamos  $H_0^*$ , mientras que si es “grande” aceptaríamos  $H_0^*$ .

Para contrastar (2.10) desde la perspectiva bayesiana, se debe utilizar una distribución inicial mixta que asigne probabilidad  $\pi_0$  al punto  $\theta_0$  y distribuya el resto,  $1 - \pi_0$ , en los valores  $\theta \neq \theta_0$  según una densidad (propia)  $\pi(\theta)$ . Es decir, si  $A$  es un elemento cualquiera de la  $\sigma$ -álgebra de Borel en  $\mathcal{R}$ , resulta

$$Pr(A) = \pi_0 I_A(\theta_0) + (1 - \pi_0) \int_A \pi(\theta) d\theta,$$

con  $I_A(\theta) = 1$  si  $\theta \in A$  e  $I_A(\theta) = 0$  si  $\theta \in A^c$ .

Si  $X_1, \dots, X_n$  es una muestra aleatoria simple de la variable aleatoria  $X$ , la densidad marginal de  $\mathbf{X} = (X_1, \dots, X_n)$  es

$$m(\mathbf{x}) = \pi_0 f(\mathbf{x} | \theta_0) + (1 - \pi_0) m_1(\mathbf{x}) \quad (2.12)$$

donde  $\mathbf{x} = (x_1, \dots, x_n)$  es el vector de observaciones y  $m_1(\mathbf{x}) = \int_{\theta \neq \theta_0} f(\mathbf{x} | \theta) \pi(\theta) d\theta$  es la densidad predictiva bajo  $H_1^*$ .

Entonces, las probabilidades finales de  $H_0^*$  y  $H_1^*$  son

$$Pr(H_0^* | \mathbf{x}) = \frac{\pi_0 f(\mathbf{x} | \theta_0)}{m(\mathbf{x})} \quad (2.13)$$

$$Pr(H_1^* | \mathbf{x}) = 1 - Pr(H_0^* | \mathbf{x}) = \frac{(1 - \pi_0) m_1(\mathbf{x})}{m(\mathbf{x})} \quad (2.14)$$

mientras que el factor Bayes

$$B(\mathbf{x}) = \frac{Pr(H_0^*|\mathbf{x})}{Pr(H_1^*|\mathbf{x})} \frac{1 - \pi_0}{\pi_0} = \frac{f(\mathbf{x}|\theta_0)}{m_1(\mathbf{x})}. \quad (2.15)$$

Si se dispone de un estadístico suficiente,  $T$ , es sencillo demostrar (Lee, 1989) que si  $t = T(\mathbf{x})$ , las expresiones (2.13), (2.14) y (2.15) siguen siendo válidas si sustituimos  $\mathbf{x}$  por  $t$ .

Berger(1985) y Lee(1989) prueban que en el caso de que las  $X_i$  sean variables aleatorias independientes e idénticamente distribuidas  $N(\theta, \sigma^2)$ , con  $\sigma^2$  conocida, si se elige  $\pi_0 = 1/2$  y una densidad inicial  $\pi(\theta)$  que sea  $N(\theta_0, \tau^2)$  sobre la hipótesis alternativa con  $\sigma = \tau$ , entonces los resultados clásico y bayesiano difieren cada vez más a medida que aumenta el tamaño muestral. De manera que para un p-valor fijo la probabilidad final de la hipótesis nula se aproxima a 1 cuando  $n$  se hace suficientemente grande. Situación similar a la obtenida por Lindley(1957) aunque allí la distribución inicial elegida fue la uniforme en un intervalo.

Debe, entonces, observarse que la discrepancia que obtienen estos autores no depende, actualmente, de la distribución inicial elegida. Es más, Jeffreys(1961) prefiere elegir una distribución inicial de Cauchy y llega a las mismas conclusiones que ya se han expuesto. Resultados similares obtienen, también, Smith y Spiegelhalter(1980) cuando utilizan una distribución inicial constante.

Berger y Sellke(1987) calculan cotas inferiores para el factor Bayes y para la probabilidad final de la hipótesis nula puntual según sea la clase de distribuciones iniciales elegida. Esto es, si  $G$  es la clase de distribuciones iniciales, ellos calculan

$$\underline{Pr}(H_0^*|x, G) = \inf_{\pi \in G} Pr(H_0^*|x).$$

Las clases de distribuciones iniciales que consideran son  $G_A$ ,  $G_S$ ,  $G_{US}$  y  $G_{NOR}$  ya mencionadas en la Sección 2.2.

En principio puede pensarse que las distribuciones de las clases  $G$  son distribuciones de probabilidad sobre  $H_1^*$  aunque para evitar complicaciones en la notación

y puesto que los resultados son los mismos, se puede suponer que las  $G$  incluyen distribuciones que dan masa al punto  $\theta_0$ . Además, de esta forma, las cotas inferiores se alcanzan en todos los casos.

Si se considera ahora el caso de una variable aleatoria,  $X$ , cuya distribución sea  $N(\theta, \sigma^2)$ , con  $\sigma^2$  conocida,  $\pi_0 = 1/2$  y  $\pi(\theta)$  perteneciente a las clases de distribuciones iniciales mencionadas, Berger y Sellke(1987) obtienen cotas inferiores para la probabilidad final de  $H_0^*$  que son más grandes, en algunos casos mucho más grandes, que los correspondientes  $p$ -valores. Todo esto les lleva a afirmar que la interpretación de los  $p$ -valores como medida de evidencia contra  $H_0^*$  no es correcta, puesto que las probabilidades finales son bastante más grandes que el  $p$ -valor. En consecuencia no sería aceptable utilizar el  $p$ -valor de 0.05 como evidencia para rechazar la hipótesis nula.

En la gráfica de la figura 2.1, Berger y Sellke(1987), están representadas cotas inferiores para las probabilidades finales de la hipótesis nula junto con el  $p$ -valor y la probabilidad final,  $Pr(H_0^*|x)$ . Para representar esta última se han tomado  $n=10$ ,  $\pi_0 = 1/2$  y  $\sigma^2 = \tau^2$ .

En esta misma línea, Edwards Lindmann y Savage(1963) y Dickey(1973) proponen utilizar cotas superiores de la probabilidad final, si bien este análisis requeriría elecciones subjetivas de la distribución inicial,  $\pi(\theta)$ , sobre la hipótesis alternativa.

Dickey(1973) propone calcular la probabilidad final de  $H_0^*$ , y también el factor Bayes, para una amplia gama de distribuciones iniciales y estudiar cuál es el efecto que produce cada elección en los resultados finales. Se trataría, por ejemplo, si la verosimilitud es  $N(\theta, \sigma^2)$ , de considerar todas las distribuciones  $N(\mu, \tau^2)$  como iniciales y presentar gráficos de la probabilidad final,  $Pr(H_0^*|x)$ , y del factor Bayes,  $B(x)$ , con  $\mu$  y  $\tau^2$  variando. Esto último permitiría, a la vez, llevar a cabo estudios de robustez al recorrer los parámetros un determinado conjunto de valores.

Para concluir esta sección, cabe señalar que si la verosimilitud se supone de

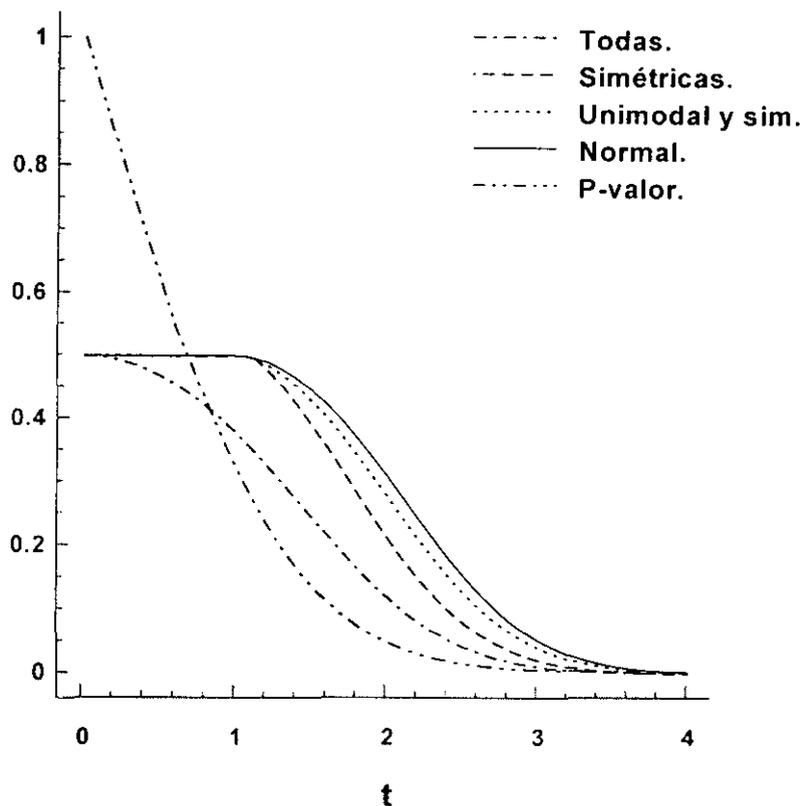


Figura 2.1: Probabilidades finales y p-valor, caso Normal.

otra forma funcional, por ejemplo Cauchy como hacen Berger y Sellke(1987), los resultados son similares a los obtenidos en el caso normal cuando se tienen valores moderados de las observaciones. En el caso concreto de la verosimilitud de Cauchy, si los valores de las observaciones son extremos las probabilidades finales y los p-valoros tienden a coincidir. Ahora bien, en este caso si  $|x|$  tiende hacia infinito la probabilidad final  $Pr(H_0^*|x)$  tiende a  $\pi_0$  cualquiera que sea la distribución inicial fija  $\pi(\theta)$ . De manera que las observaciones extremas no proporcionan información desde el punto de vista bayesiano.

## 2.4 Contraste de hipótesis de tipo intervalo

El contraste de hipótesis nula puntual, presentado en la sección precedente, raramente se presenta en la gran mayoría de las situaciones prácticas. Si exceptuamos, quizás, algún caso muy específico como puede ser el contraste de percepción extrasensorial(ESP) donde la hipótesis nula es que el individuo en cuestión no tiene ESP (Soal y Bateman, 1954) o el contraste de conexión entre dos factores genéticos (Lindley, 1957), prácticamente nunca se da el caso en que se estudie la hipótesis de que  $\theta = \theta_0$  “exáctamente” (Lehman, 1959). Nosotros añadiríamos el caso de comparar dos distribuciones normales, donde es capital suponer la hipótesis nula  $\sigma_1^2/\sigma_2^2 = 1$ , al menos desde el punto de vista frecuentista. Así, desde un punto de vista práctico sería más razonable contrastar

$$H_0 : \theta \in I_b \quad \text{frente a} \quad H_1 : \theta \in I_b^c, \quad (2.16)$$

siendo  $I_b = (\theta_0 - b, \theta_0 + b)$  con  $b > 0$  elegido de tal manera que los  $\theta \in I_b$  puedan considerarse “indistinguibles” de  $\theta_0$ . Se pueden encontrar ejemplos en Berger(1985), Berger y Delampady(1987) y Lee(1989) entre otros.

La cuestión que se plantea, entonces, es que dado que lo que realmente se debería contrastar es (2.16), ¿cuándo se puede aproximar (2.16) por (2.10)?

Desde el punto de vista clásico el problema de determinar el valor de  $b$  que hace que (2.16) pueda ser aproximado por (2.10) consiste en determinar cuándo los p-valores de (2.10),  $p(x)$ , dado por (2.11), y de (2.16),  $p_b(x)$ , sean aproximadamente iguales. Siendo

$$p_b(x) = \sup_{\theta \in I_b} Pr\{|T(X)| \geq |T(x)| | \theta\}. \quad (2.17)$$

Si  $X$  es una variable aleatoria que sigue una distribución  $N(\theta, \sigma^2)$ , con  $\sigma^2$  conocida, Berger y Delampady(1987) obtienen condiciones para aproximar  $p(x)$  y  $p_b(x)$  de manera que el error cometido no supere un determinado porcentaje. El tamaño muestral,  $n$ , influye en el valor de  $b$  que permite aproximar ambos p-valores de tal manera que si  $n$  es muy grande entonces  $b$  será muy pequeño.

Desde el punto de vista bayesiano, (2.16) se puede aproximar por (2.10) cuando las probabilidades finales de la hipótesis nula sean aproximadamente iguales en ambos contrastes. Un primer caso en el que esto ocurre es aquel en el que la verosimilitud es aproximadamente constante en  $I_b$  (Lee, 1989). Pero esta condición parece demasiado fuerte y, además presenta demasiadas dependencias ya que el factor que mide la variación de la verosimilitud en el intervalo  $I_b$  depende del tamaño muestral, de la magnitud de la observación y de la varianza de la verosimilitud.

Sea  $X$  una variable aleatoria continua con función de densidad  $f(x - \theta)$ ,  $\theta$  parámetro desconocido,  $\theta \in \Theta$ . Para efectuar el contraste (2.10) recordamos que es preciso elegir una distribución inicial mixta que asigne masa  $\pi_0$ ,  $0 < \pi_0 < 1$ , a la hipótesis nula,  $H_0$ , y reparta el resto,  $1 - \pi_0$ , sobre  $\theta \neq \theta_0$ , hipótesis alternativa, de acuerdo con una densidad que llamaremos  $\pi(\theta)$ .

En cuanto al contraste (2.16) supondremos, únicamente, que tenemos una densidad continua,  $\pi(\theta)$ , sobre  $\Theta$ . Esta  $\pi(\theta)$  será la misma que la utilizada en el contraste (2.10).

El factor Bayes del contraste (2.10) viene dado por (2.15), mientras que el del contraste (2.16) es

$$B_b(x) = \frac{\int_{I_b} f(x|\theta)\pi(\theta) d\theta}{\int_{I_b} \pi(\theta) d\theta} \frac{\int_{I_b^c} \pi(\theta) d\theta}{\int_{I_b} f(x|\theta)\pi(\theta) d\theta}. \quad (2.18)$$

Gómez Villegas y Gómez Sánchez-Manzano(1992) prueban que  $B_b(x)$  converge a  $B(x)$  cuando  $b$  tiende a cero. Esto es, desde el punto de vista bayesiano ambos problemas serían equivalentes para valores suficientemente pequeños de  $b$ . Estos autores sugieren elegir  $\pi_0$ , la probabilidad que se asigna a la hipótesis nula puntual, como

$$\pi_0 = \int_{I_b} \pi(\theta) d\theta \quad (2.19)$$

puesto que de esta manera el cociente de probabilidades finales del contraste (2.16)

$$\frac{Pr(I_b|x)}{Pr(I_b^c|x)} = B_b(x) \frac{\pi_0}{1 - \pi_0} \quad \text{converge a} \quad B(x) \frac{\pi_0}{1 - \pi_0} = \frac{Pr(H_0|x)}{Pr(H_1|x)}$$

**Tabla 2.1:** Probabilidades finales de  $H_0$  en los contrastes de hipótesis nula puntual y de intervalo, caso Normal.

|        | t      |        |        |        |
|--------|--------|--------|--------|--------|
|        | 1.645  | 1.960  | 2.576  | 3.291  |
| b= 0.1 | 0.0774 | 0.0477 | 0.0139 | 0.0021 |
|        | 0.0794 | 0.0482 | 0.0142 | 0.0022 |
| b= 0.2 | 0.1544 | 0.0983 | 0.0297 | 0.0045 |
|        | 0.1699 | 0.1083 | 0.0356 | 0.0066 |
| b= 0.3 | 0.2303 | 0.1515 | 0.0477 | 0.0074 |
|        | 0.2780 | 0.1890 | 0.0717 | 0.0161 |

que es el cociente de probabilidades finales del contraste (2.10). De manera que en los dos contrastes se llegaría a tomar la misma decisión, siempre que  $b$  fuese suficientemente pequeño. En efecto, si se toma  $\pi_0$  como indica (2.19) la probabilidad final de la hipótesis nula es muy parecida en ambos contrastes cuando  $b$  es pequeño, como se muestra en el siguiente ejemplo.

**Ejemplo 2.2.** Sea  $X_1, \dots, X_n$  una muestra aleatoria simple de una población  $N(\theta, \sigma^2)$ , con  $\sigma^2$  conocida. En este caso sabemos que la media muestral,  $\bar{X}$ , tiene una distribución  $N(\theta, \sigma^2/n)$ . Consideremos el estadístico  $T(\bar{X}) = \sqrt{n}(\bar{X} - \theta_0)/\sigma$ . Supongamos que la distribución inicial para  $\theta$ ,  $\pi(\theta)$ , es  $N(\theta_0, \tau^2)$ .

Entonces, si  $t = \sqrt{n}(\bar{x} - \theta_0)/\sigma$  y  $\Phi$  denota la función de distribución  $N(0, 1)$ , la tabla 2.1 muestra las probabilidades finales de la hipótesis nula en los contrastes (2.10) y (2.16) para distintos valores de  $b$  y de  $t$ . Para cada valor de  $b$ , la fila superior muestra la probabilidad final de la hipótesis nula puntual mientras que la fila inferior muestra la probabilidad final de la hipótesis nula de intervalo. Los cálculos se han efectuado tomando  $n = 10$  y  $\sigma^2 = \tau^2 = 1$ .  $\square$

En cuanto a la relación entre la probabilidad final de la hipótesis nula en (2.16) y el correspondiente p-valor,  $p_b(x)$ , Berger y Delampady(1987) muestran que fijada

**Tabla 2.2:** Probabilidades finales y p-valores de  $H_0$  en el contraste de hipótesis de tipo intervalo, caso Normal.

|         | t      |        |        |        |
|---------|--------|--------|--------|--------|
|         | 1.645  | 1.960  | 2.576  | 3.291  |
| b= 0.1  | 0.1169 | 0.0615 | 0.0138 | 0.0016 |
|         | 0.0732 | 0.0434 | 0.0120 | 0.0017 |
| b = 0.2 | 0.1670 | 0.0969 | 0.0266 | 0.0040 |
|         | 0.1571 | 0.0978 | 0.0302 | 0.0051 |
| b = 0.3 | 0.2479 | 0.1577 | 0.0520 | 0.0096 |
|         | 0.2583 | 0.1715 | 0.0613 | 0.0125 |

la densidad inicial,  $\pi(\theta)$ , la probabilidad final es aproximadamente igual al p-valor cuando el tamaño muestral,  $n$ , es grande. Más concretamente,  $Pr(H_0|\bar{x}_n)$  tiende a  $p_b(x)$  cuando  $n$  tiende a infinito. En el ejemplo siguiente comparamos estas dos medidas en un caso concreto.

**Ejemplo 2.3.** Supongamos, como en el ejemplo 2.2, que  $X_1, \dots, X_n$  es una muestra aleatoria simple de una variable aleatoria  $X$  con distribución  $N(\theta, \sigma^2)$ , con  $\sigma^2$  conocida, y tomemos como densidad inicial una  $N(\theta_0, \tau^2)$ . En estas condiciones, la tabla 2.2 muestra, para distintos valores de  $b$  y  $t$ , en la fila inferior los valores de la probabilidad final de la hipótesis nula en (2.16) y en la fila superior el p-valor correspondiente. Los cálculos se han efectuado tomando  $n = 10$ ,  $\sigma^2 = 1$  y  $\tau^2 = 2$ .  
□

Se observa que las diferencias entre el p-valor y la probabilidad final de la hipótesis nula en el caso del contraste de tipo intervalo no son grandes y que las conclusiones serían las mismas desde los puntos de vista clásico y bayesiano. No obstante, conviene observar que para valores muy pequeños de  $b$  el p-valor se estabiliza mientras que la probabilidad final de la hipótesis nula tiende a cero, como consecuencia de elegir su probabilidad inicial cada vez más pequeña.

**Tabla 2.3:** Probabilidad final de  $H_0$  en el contraste de hipótesis de tipo intervalo y p-valor en el contraste de hipótesis nula puntual, caso Normal.

|              |        |        |        |        |
|--------------|--------|--------|--------|--------|
| t            | 1.645  | 1.960  | 2.576  | 3.291  |
| p-valor      | 0.1169 | 0.0615 | 0.0138 | 0.0016 |
| b            | 0.0732 | 0.0434 | 0.0120 | 0.0017 |
| $\Pr(I_b t)$ | 0.1670 | 0.0969 | 0.0266 | 0.0040 |

Casella y Berger en la discusión del trabajo de Berger y Delampady(1987) muestran que se puede conseguir, tomando  $b$  adecuadamente, que los p-valores del contraste de hipótesis nula puntual (2.10) y las probabilidades de la hipótesis nula de intervalo en (2.16) sean, prácticamente, casi iguales. Reproducimos aquí, en la tabla 2.3, el ejemplo de Casella y Berger para algunos valores de  $t$ . La verosimilitud es  $N(\theta, 1)$  y la distribución inicial elegida para el parámetro es  $N(0, 4)$ .

Estos resultados les llevan a afirmar que los p-valores y las probabilidades finales de la hipótesis nula se parecen siempre que  $b$ , radio del intervalo, sea suficientemente pequeño.

Para finalizar esta sección, presentamos un ejemplo con verosimilitud binomial en el que se observa que, eligiendo  $b$  convenientemente, es posible aproximar el p-valor y la probabilidad final en el contraste de hipótesis nula puntual.

**Ejemplo 2.4.** Sea  $X$  una variable aleatoria con distribución binomial de parámetros  $n$  y  $\theta$ , este último desconocido perteneciente al espacio paramétrico  $\Theta = (0, 1)$ . La función de masa, por tanto, es

$$f(x|\theta) = \binom{n}{x} \theta^x (1 - \theta)^{n-x}, \text{ para } x = 0, 1, \dots, n.$$

Se desea contrastar  $H_0^* : \theta = \theta_0$  frente a  $H_1^* : \theta \neq \theta_0$ ,  $0 < \theta_0 < 1$ . Tomando para el p-valor el estadístico  $T(X) = 1/f(X|\theta_0)$ , resulta que si se ha observado  $X = x$  el

p-valor será:

$$p(x) = Pr(t; f(t|\theta_0) \leq f(x|\theta_0)).$$

Como distribución inicial,  $\pi(\theta)$ , para el parámetro elegimos una beta cuya media sea, precisamente  $\theta_0$ . Teniendo en cuenta que la densidad de la  $Be(r, s)$  es

$$\frac{\Gamma(r+s)}{\Gamma(r)\Gamma(s)} \theta^{r-1} (1-\theta)^{s-1}, \quad 0 < \theta < 1$$

y su media es  $r/(r+s)$ , para que esta media sea igual a  $\theta_0$  debemos elegir  $\pi(\theta) \sim Be(k\theta_0, k(1-\theta_0))$ , con  $k > 0$ . De esta manera tenemos que la densidad predictiva resulta

$$m_\pi(x) = \int f(x|\theta)\pi(\theta) d\theta = \binom{n}{x} \frac{\Gamma(k)\Gamma(x+k\theta_0)\Gamma(k(1-\theta_0)+n-x)}{\Gamma(k\theta_0)\Gamma(k(1-\theta_0))\Gamma(n+k-2)}$$

y, por tanto, el factor Bayes dado por (2.15),

$$B(x) = \frac{\theta_0^x (1-\theta_0)^{1-x} \Gamma(k\theta_0)\Gamma(k(1-\theta_0))\Gamma(n+k-2)}{\Gamma(k)\Gamma(x+k\theta_0)\Gamma(k(1-\theta_0)+n-x)}.$$

La probabilidad final de la hipótesis nula se obtiene a partir de  $B(x)$  ya que

$$Pr(H_0|x) = \left(1 - \frac{1-\pi_0}{\pi_0} \frac{1}{B(x)}\right)^{-1}.$$

En cuanto al contraste  $H_0 : \theta \in I_b$  frente a  $H_1 : \theta \in I_b^c$  tenemos

$$Pr(I_b) = \int_{I_b} \pi(\theta) d\theta, \tag{2.20}$$

mientras que la probabilidad final de la hipótesis nula es

$$Pr(I_b|x) = \int_{I_b} \pi(\theta|x) d\theta,$$

con  $\pi(\theta|x)$  densidad de la distribución  $Be(x+k\theta_0, n-x+k(1-\theta_0))$ .

La tabla 2.4 muestra los p-valor, la probabilidad final de la hipótesis nula puntual cuando se le asigna una masa inicial  $\pi_0 = 1/2$ , la probabilidad final de la hipótesis de intervalo con  $b = 0.05$  y la probabilidad final de la hipótesis nula puntual  $Pr(H_0^*|x)$  cuando asignamos la probabilidad inicial,  $\pi_0$ , calculada a partir de (2.20).

**Tabla 2.4:** Probabilidades finales y p-valores, caso Binomial.

| n  | x | $\theta_0$ | p-valor | $\Pr(H_0^* x, 1/2)$ | $\Pr(I_b)$ | $\Pr(I_b x)$ | $\Pr(H_0^* x, \pi_0)$ |
|----|---|------------|---------|---------------------|------------|--------------|-----------------------|
| 9  | 1 | 0.45       | 0.0476  | 0.2569              | 0.1917     | 0.1515       | 0.1161                |
| 10 | 7 | 0.40       | 0.1011  | 0.3470              | 0.1721     | 0.1055       | 0.0994                |
| 12 | 7 | 0.30       | 0.0524  | 0.3239              | 0.1824     | 0.0564       | 0.0633                |
| 14 | 8 | 0.25       | 0.0103  | 0.1685              | 0.1917     | 0.0554       | 0.0459                |
| 15 | 3 | 0.50       | 0.0352  | 0.1844              | 0.1689     | 0.0394       | 0.0439                |
| 17 | 8 | 0.25       | 0.0478  | 0.3565              | 0.1761     | 0.0593       | 0.0593                |
| 17 | 5 | 0.10       | 0.0827  | 0.3133              | 0.1824     | 0.1082       | 0.0966                |
| 19 | 2 | 0.40       | 0.0083  | 0.0740              | 0.1720     | 0.0140       | 0.0163                |
| 20 | 2 | 0.30       | 0.0525  | 0.2325              | 0.2694     | 0.1960       | 0.1440                |

Las probabilidades finales de las hipótesis nulas en los contrastes (2.10) y (2.16) son muy parecidas cuando en la distribución inicial mixta que utilizamos en el contraste (2.10) se elige  $\pi_0$ , probabilidad inicial asignada a  $H_0^*$ , como indica (2.20). Además, en algunos casos, dichas probabilidades alcanzan valores numéricos muy próximos al p-valor correspondiente. Así, por ejemplo, si  $n = 10$ ,  $x = 7$  y  $\theta_0 = 0.4$ , la probabilidad final de la hipótesis nula puntual es 0.0994 mientras que la probabilidad final de la hipótesis nula de intervalo es 0.1055 y el p-valor es 0.1011. En cambio, cuando se elige  $\pi_0 = 1/2$  se observa que la probabilidad final de  $H_0^*$  es mucho más grande que el p-valor.

A la vista de los resultados parciales obtenidos en los ejemplos anteriores parece que procede investigar si se puede obtener un resultado general que permita aproximar, en el contraste de hipótesis nula puntual, los valores de la evidencia bayesiana, expresada mediante una cota inferior para la probabilidad final de la hipótesis nula, y la evidencia clásica, concretada en el p-valor. A este objetivo se dedican los capítulos 3 y 4 de la presente memoria.

# Capítulo 3

## Distribuciones iniciales unimodales y simétricas

### 3.1 Introducción

En el contraste de hipótesis nula puntual, como ya mencionamos en el capítulo 2, Berger y Sellke(1987) y Berger y Delampady(1987) afirman que existe discrepancia entre las aproximaciones clásica y bayesiana, la primera expresada en términos del  $p$ -valor y la segunda a través del factor Bayes y de la probabilidad final de la hipótesis nula. Casella y Berger(1987) argumentan que tal discrepancia es debida a la necesidad de utilizar una distribución inicial de tipo mixto para efectuar el análisis bayesiano de dicho contraste.

En este capítulo veremos que la discrepancia observada por Berger y Sellke no se debe tanto a la necesidad de elegir una distribución inicial de tipo mixto para el parámetro como al reparto de la probabilidad que ellos llevan a cabo:  $1/2$  para el punto  $\theta = \theta_0$  y  $1/2$  para todos los  $\theta \neq \theta_0$ . Concluyendo en que ese reparto puede no ser representativo de una distribución inicial imparcial sino que mas bien refleja un

cierto sesgo en favor de la hipótesis nula,  $H_0^*$ . Es más, eligiendo  $\pi_0$  como proponen Gómez Villegas y Gómez Sánchez-Manzano (1992) probaremos que el ínfimo de la probabilidad final de la hipótesis nula puntual es aproximadamente igual que el correspondiente p-valor.

Sea  $\theta$  un parámetro de localización desconocido perteneciente a un espacio paramétrico  $\Theta \subset \mathbf{R}$ . Consideramos el contraste de hipótesis nula puntual

$$H_0^* : \theta = \theta_0 \quad \text{frente a} \quad H_1^* : \theta \neq \theta_0. \quad (3.1)$$

Supongamos que, para resolver el contraste, se observa una variable aleatoria,  $X$ , con función de densidad  $f(x - \theta)$ , continua en  $\theta_0$ . Llevar a cabo un análisis bayesiano de (3.1) implica, como ya se indicó en la Sección 2.3, utilizar una distribución inicial de tipo mixto en la que, entre otros elementos, hay que seleccionar una distribución sobre  $H_1^*$ .

Ahora bien, elegir una distribución inicial concreta,  $\pi(\theta)$ , no es sencillo en la mayoría de las ocasiones puesto que la información previa disponible no será, en general, lo suficientemente precisa como para poder determinar  $\pi(\theta)$  con total seguridad. Es habitual, en estos casos, dejar que la distribución inicial,  $\pi(\theta)$ , pertenezca a una clase amplia de distribuciones y, sobre esa clase, calcular una cota inferior para la probabilidad final. Nosotros calcularemos el ínfimo de la probabilidad final de la hipótesis nula cuando la distribución inicial varíe en la clase

$$G_{US} = \{ \text{Todas las distribuciones unimodales y simétricas en torno a } \theta_0 \}.$$

Pensamos que  $G_{US}$  es una clase razonable de distribuciones iniciales para el parámetro  $\theta$  en el problema del contraste de hipótesis nula puntual, en el sentido de que “no favorece” a la hipótesis alternativa,  $H_1^*$ . Esto es, el uso de  $G_{US}$  debería prevenir que haya excesivo sesgo hacia valores  $\theta \neq \theta_0$ . Además, esta clase ha sido utilizada por Casella y Berger(1987) y por Berger y Sellke(1987) para obtener cotas inferiores de la probabilidad final en los contrastes unilateral y bilateral respectivamente.

El contraste de hipótesis nula puntual en el problema de un parámetro de localización, que es el que nos ocupa en esta memoria, se elige más por conveniencia matemática que por metodología estadística (Casella y Berger, 1987). Para la mayoría de los autores, una hipótesis precisa puede ser representada, salvo raras excepciones, como

$$H_0 : |\theta - \theta_0| \leq b \quad \text{frente a} \quad H_1 : |\theta - \theta_0| \geq b, \quad (3.2)$$

donde  $b > 0$  es “pequeño”. Esta situación ya fue comentada en el capítulo 2.

De manera que lo que proponemos es reemplazar el contraste (3.1) por el contraste (3.2). Este cambio nos proporciona algunas ventajas:

(i) En (3.2) no es preciso utilizar una distribución inicial de tipo mixto.

(ii) Si queremos contrastar (3.1), dado  $\pi(\theta)$ , podemos fijar el valor de  $b$ , aquel que haga equivalentes (3.1) y (3.2) y calcular

$$\pi_0 = \int_{|\theta - \theta_0| \leq b} \pi(\theta) d\theta. \quad (3.3)$$

que es la masa de probabilidad inicial que, en la distribución mixta, asignaremos a la hipótesis nula,  $H_0^*$ , en el contraste (3.1). Por otra parte, elegir  $b$ , radio del intervalo, es más intuitivo que elegir  $\pi_0$  (habitualmente se ha tomado  $1/2$  como valor de  $\pi_0$ ).

Siguiendo con las notaciones introducidas en el capítulo 2, en la sección 3.2 presentamos el resultado principal que consiste en la obtención del ínfimo de la probabilidad final de la hipótesis nula en el contraste puntual cuando la distribución inicial varía en la clase  $G_{US}$ . Además, probamos que la distribución inicial de tipo mixto utilizada queda muy próxima a la distribución inicial utilizada en el contraste de tipo intervalo.

En la sección 3.3 llevamos a cabo la comparación entre el p-valor, como medida de evidencia clásica contra la hipótesis nula, y el ínfimo de la probabilidad final de  $H_0^*$  tomado sobre la clase  $G_{US}$ . Finalmente, analizamos algunos ejemplos y comentamos los resultados obtenidos.

Los resultados que se obtienen dependen de la distribución de la que provienen los datos. Siguiendo la clasificación de Main(1987) y Gómez Villegas y Maín(1992), si la verosimilitud es resistente a datos atípicos, como es el caso de la distribución normal, nuestro procedimiento proporciona un valor del ínfimo de la probabilidad final de  $H_0^*$  que es próximo tanto al p-valor como al ínfimo de la probabilidad final de  $H_0$ . En cambio si la verosimilitud es propensa a datos atípicos, como la distribución de Cauchy, o neutra, como la distribución doble exponencial, los ínfimos de  $H_0$  y  $H_0^*$  son aproximadamente iguales cuando  $b$  es pequeño, pero inferiores al p-valor. Los ejemplos 3.2 y 3.3 se dedican al estudio de estos problemas.

Asímismo, se obtienen algunos resultados asintóticos del ínfimo obtenido en relación con el p-valor. Estos resultados dependen, como era de esperar, de las colas de la distribución de muestreo.

### 3.2 Ínfimo de la probabilidad final.

En esta sección obtenemos el ínfimo de la probabilidad final de la hipótesis nula en el contraste (3.1) cuando la información inicial viene dada por una distribución de la clase de las distribuciones unimodales y simétricas en torno a  $\theta_0$ ,  $G_{US}$ .

Para ello, recogemos en primer lugar un Lema que facilita los cálculos posteriores y cuyo significado es que el ínfimo de la probabilidad final de la hipótesis nula es el mismo si se toma sobre una clase de distribuciones iniciales,  $G$ , o sobre la clase de todas las mixturas de elementos de  $G$ . La demostración del Lema puede encontrarse en Casella y Berger (1987).

**Lema 3.1.** Sea  $G = \{\pi_\alpha, \alpha \in \mathcal{A}\}$  una clase de distribuciones iniciales sobre la recta real ordenadas por un conjunto  $\mathcal{A}$ . Sea  $G_M$  la clase de todas las mixturas de elementos de  $G$ , es decir  $\pi$  pertenece a  $G_M$  si y sólo si

$$\pi(B) = \int_{\mathcal{A}} \pi_\alpha(B) dP(\alpha)$$

para alguna medida de probabilidad  $P$  sobre  $\mathcal{A}$  y todo conjunto medible  $B$ . Entonces,

$$\inf_{\pi \in G_M} Pr(H_0|x) = \inf_{\pi_\alpha \in G} Pr(H_0|x).$$

El teorema siguiente establece el ínfimo de la probabilidad final, para su demostración utilizamos el Lema 3.1 y el hecho conocido de que todo elemento de la clase  $G_{US}$  puede obtenerse mediante mixtura de elementos de la clase de todas las distribuciones uniformes y simétricas en torno a  $\theta_0$ , llamaremos  $G_U$  a esta clase.

**Teorema 3.1** *Si consideramos el contraste de hipótesis nula puntual (3.1) y en la distribución inicial de tipo mixto tomamos  $\pi_0$  como se indica en (3.3), supuesto que*

$$\int_{\mathcal{R}} f(x - \theta) d\theta < \infty \tag{3.4}$$

entonces

$$\inf_{\pi \in G_{US}} Pr(H_0^*|x) = \inf_{\pi \in G_U} Pr(H_0^*|x) = \left( 1 + \frac{1}{2b} \int_{-\infty}^{+\infty} \frac{f(x - \theta)}{f(x - \theta_0)} d\theta \right)^{-1}. \tag{3.5}$$

DEMOSTRACIÓN:

La primera igualdad de (3.5) es consecuencia del Lema 3.1. Para probar la segunda, consideremos una distribución  $\pi(\theta)$  uniforme en el intervalo  $(\theta_0 - k, \theta_0 + k)$ ,  $k > 0$  número real. En estas condiciones, si se ha observado  $x$ , la probabilidad final de la hipótesis nula,  $H_0^*$ , es

$$\begin{aligned} Pr(H_0^*|x) &= \frac{\frac{b}{k} f(x - \theta_0)}{\frac{b}{k} f(x - \theta_0) + \left(1 - \frac{b}{k}\right) \int_{|\theta - \theta_0| \leq k} \frac{1}{2k} f(x - \theta) d\theta} \\ &= \frac{2f(x - \theta_0)}{2f(x - \theta_0) + \left(\frac{1}{b} - \frac{1}{k}\right) \int_{|\theta - \theta_0| \leq k} f(x - \theta) d\theta}. \end{aligned}$$

La variación de  $Pr(H_0^*|x)$  con respecto a  $k$  viene dada por su derivada, es decir,

$$\frac{\partial}{\partial k} Pr(H_0^*|x) = \frac{-2f(x - \theta_0) \left[ \frac{1}{k^2} \int_{|\theta - \theta_0| \leq k} f(x - \theta) d\theta + \left( \frac{1}{b} - \frac{1}{k} \right) [f(x - k) + f(x + k)] \right]}{\left( 2f(x - \theta_0) + \left( \frac{1}{b} - \frac{1}{k} \right) \int_{|\theta - \theta_0| \leq k} f(x - \theta) d\theta \right)^2},$$

con lo que se observa que

$$\frac{\partial}{\partial k} Pr(H_0^*|x) < 0,$$

siempre que  $b < k$ . De manera que  $Pr(H_0^*|x)$  es decreciente en  $k$  y, por tanto, el mínimo se alcanzará cuando  $k$  tienda a infinito, es decir,

$$\begin{aligned} \inf_{\pi \in G_U} Pr(H_0^*|x) &= \lim_{k \rightarrow \infty} \frac{2f(x - \theta_0)}{2f(x - \theta_0) + \left( \frac{1}{b} - \frac{1}{k} \right) \int_{|\theta - \theta_0| \leq k} f(x - \theta) d\theta} \\ &= \frac{2f(x - \theta_0)}{2f(x - \theta_0) + \frac{1}{b} \int_{-\infty}^{+\infty} f(x - \theta) d\theta} \\ &= \left( 1 + \frac{1}{2b} \int_{-\infty}^{+\infty} \frac{f(x - \theta)}{f(x - \theta_0)} d\theta \right)^{-1}, \end{aligned}$$

puesto que de la condición 3.4

$$\frac{1}{k} \int_{|\theta - \theta_0| \leq k} f(x - \theta) d\theta = \frac{1}{k} \int_{\mathcal{R}} I_{(\theta_0 - k, \theta_0 + k)}(\theta) f(x - \theta) d\theta \leq \frac{1}{k} \int_{\mathcal{R}} f(x - \theta) d\theta < \infty,$$

lo que prueba el teorema.  $\square$

Obsérvese que, fijado  $b$  y después de obtener  $x$ , la expresión (3.5) nos proporciona el ínfimo de la probabilidad final de la hipótesis nula puntual.

**Ejemplo 3.1.** Sea  $X$  variable aleatoria con distribución  $N(\theta, \sigma^2)$ , con  $\sigma^2$  conocida y  $X_1, \dots, X_n$  una muestra aleatoria simple de la población  $X$ . En estas condiciones, es conocido que la distribución de  $\bar{X}$  es  $N(\theta, \sigma^2/n)$ , de manera que el ínfimo de la probabilidad final de la hipótesis nula puntual, tomado sobre la clase  $G_{US}$  dado por (3.5), es

$$\inf_{\pi \in G_{US}} Pr(H_0^*|\bar{x}) = \left[ 1 + \frac{1}{2b} \frac{\sqrt{2\pi}\sigma}{\sqrt{n}} \exp \left\{ \frac{n}{2\sigma^2} (\bar{x} - \theta_0)^2 \right\} \right]^{-1}. \quad (3.6)$$

**Tabla 3.1:** Ínfimo de la probabilidad final de la hipótesis nula sobre la clase  $G_{US}$ , caso Normal.

| $b$ | $t$    |        |        |        |
|-----|--------|--------|--------|--------|
|     | 1.645  | 1.960  | 2.576  | 3.291  |
| 0.1 | 0.0612 | 0.0356 | 0.0091 | 0.0011 |
| 0.3 | 0.1636 | 0.0998 | 0.0267 | 0.0034 |
| 0.5 | 0.2449 | 0.1560 | 0.0437 | 0.0056 |

La tabla 3.1 muestra, cuando  $\sigma^2 = 1$  y  $n = 10$ , el valor de (3.6) para algunos valores específicos de  $t$  y algunos valores de  $b$ , siendo  $t = \sqrt{n}|\bar{x} - \theta_0|/\sigma$ .  $\square$

La elección de la distribución inicial mixta como nosotros proponemos es la adecuada ya que, insistiendo en la relación entre el contraste de hipótesis nula puntual y el de tipo intervalo, si en (3.1) elegimos la distribución inicial de tipo mixto con  $\pi_0$  como indica (3.3) la distancia, medida en términos de la información de Shannon, entre esta distribución mixta y la densidad inicial que se toma en el contraste (3.2) es muy pequeña cuando  $b$ , radio del intervalo, es suficientemente pequeño. El resultado se recoge en el siguiente teorema

**Teorema 3.2** . Sea  $\pi_1(\theta)$  la distribución inicial mixta utilizada en el contraste (3.1)

$$\pi_1(\theta) = \pi_0 I_{\{\theta_0\}}(\theta) + (1 - \pi_0)\pi(\theta)I_{\{\theta \neq \theta_0\}}(\theta), \quad (3.7)$$

donde  $\pi_0$  viene dado por (3.3),  $\pi(\theta)$  es la densidad inicial utilizada en el contraste (3.2) y la función  $I_A(\theta)$  es la función indicatriz del conjunto  $A$ .

Si medimos la distancia entre  $\pi_1(\theta)$  y  $\pi(\theta)$  por

$$d(\pi_1, \pi) = \int \pi(\theta) \ln \frac{\pi(\theta)}{\pi_1(\theta)} d\theta \quad (3.8)$$

entonces

$$\lim_{b \rightarrow 0} d(\pi_1, \pi) = 0$$

DEMOSTRACIÓN:

La distancia entre  $\pi_1$  y  $\pi$  es

$$\begin{aligned} d(\pi_1, \pi) &= \int \pi(\theta) \ln \frac{\pi(\theta)}{\pi_1(\theta)} d\theta \\ &= \int \pi(\theta) \ln \frac{\pi(\theta)}{\pi_0 I_{\{\theta_0\}}(\theta) + (1 - \pi_0)\pi(\theta) I_{\{\theta \neq \theta_0\}}(\theta)} d\theta \\ &= \int \pi(\theta) \ln \pi(\theta) d\theta - \int \pi(\theta) \ln \left\{ \pi_0 I_{\{\theta_0\}}(\theta) + (1 - \pi_0)\pi(\theta) I_{\{\theta \neq \theta_0\}}(\theta) \right\} d\theta. \end{aligned}$$

Tomando límites cuando  $b$  tiende a cero, resulta

$$\begin{aligned} \lim_{b \rightarrow 0} d(\pi_1, \pi) &= \\ &= \int \pi(\theta) \ln \pi(\theta) d\theta - \int \pi(\theta) \lim_{b \rightarrow 0} \ln \left\{ \pi_0 I_{\{\theta_0\}}(\theta) + (1 - \pi_0)\pi(\theta) I_{\{\theta \neq \theta_0\}}(\theta) \right\} d\theta \\ &= \int \pi(\theta) \ln \pi(\theta) d\theta - \int \pi(\theta) \ln \left\{ \lim_{b \rightarrow 0} \pi_0 I_{\{\theta_0\}}(\theta) + \lim_{b \rightarrow 0} (1 - \pi_0)\pi(\theta) I_{\{\theta \neq \theta_0\}}(\theta) \right\} d\theta \\ &= \int \pi(\theta) \ln \pi(\theta) d\theta - \int \pi(\theta) \ln \pi(\theta) d\theta = 0, \end{aligned}$$

lo que prueba el teorema.  $\square$

De manera que si  $b$  es suficientemente pequeño, que es cuando tiene sentido hablar de la relación entre los contrastes (3.1) y (3.2), las distribuciones iniciales tomadas en ambos casos quedan muy próximas. En cambio si se elige  $\pi_0 = 1/2$ ; la distribución inicial mixta se escribe

$$\pi_2(\theta) = \frac{1}{2} I_{\{\theta_0\}}(\theta) + \frac{1}{2} \pi(\theta) I_{\{\theta \neq \theta_0\}}(\theta),$$

y la distancia de esta  $\pi_2(\theta)$  a la densidad  $\pi(\theta)$  resulta

$$d(\pi_2, \pi) = \int \pi(\theta) \ln \frac{\pi(\theta)}{\pi_2(\theta)} d\theta$$

$$\begin{aligned}
 &= \int \pi(\theta) \ln \frac{\pi(\theta)}{\frac{1}{2}I_{\{\theta_0\}}(\theta) + \frac{1}{2}\pi(\theta)I_{\{\theta \neq \theta_0\}}(\theta)} d\theta \\
 &= \int \pi(\theta) \ln \pi(\theta) d\theta - \int \pi(\theta) \ln \left( \frac{1}{2}\pi(\theta) \right) d\theta \\
 &= -\ln \left( \frac{1}{2} \right) = 0.693.
 \end{aligned}$$

De manera que  $\pi_2(\theta)$ , distribución inicial mixta con  $\pi_0 = 1/2$ , está más alejada de la distribución  $\pi(\theta)$  que  $\pi_1(\theta)$ , que es la distribución inicial mixta que nosotros proponemos para el contraste de hipótesis nula puntual.

### 3.3 Comparación entre las evidencias clásica y bayesiana

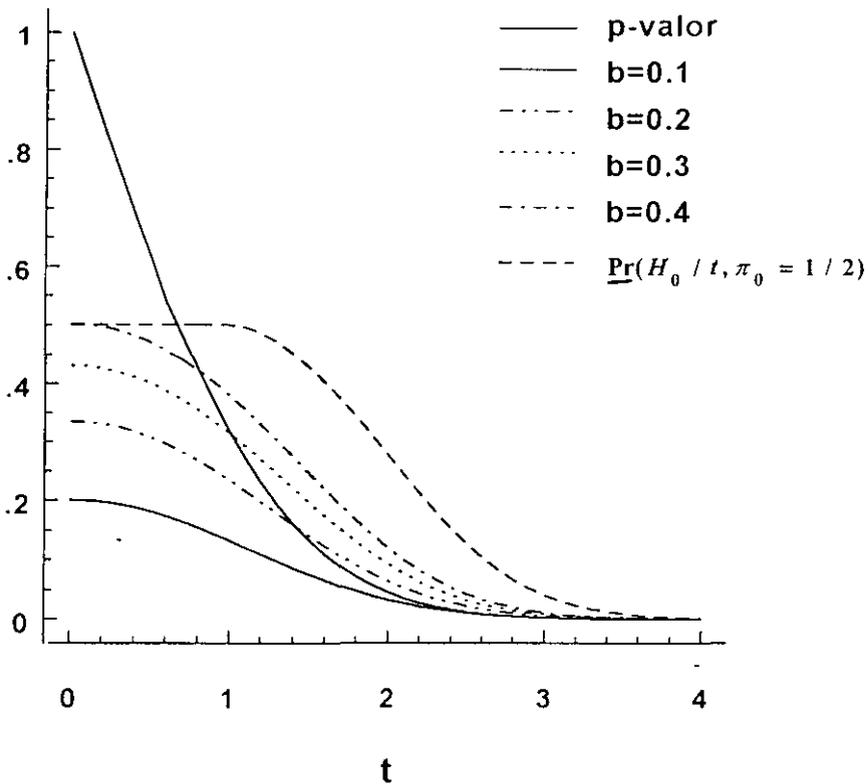
En la sección 3.2 hemos obtenido el ínfimo de la probabilidad final de la hipótesis nula puntual cuando la distribución inicial pertenece a la clase  $G_{US}$ . En esta sección probaremos que eligiendo  $b$  adecuadamente, el ínfimo que se obtiene es aproximadamente igual al p-valor, tomado este último como medida de la evidencia clásica contra  $H_0^*$ .

Recordemos que si  $T(X)$  es un estadístico adecuado para  $\theta$ , el p-valor correspondiente a los datos observados,  $x$ , viene dado por

$$p(x) = Pr(|T(X)| \geq |T(x)| | \theta = \theta_0). \tag{3.9}$$

Entonces, igualando (3.5) y (3.9) obtenemos un valor de  $b$ , digamos  $b^*$ , de tal manera que el ínfimo de la probabilidad final de  $H_0^*$  y el p-valor coinciden. El valor de  $b^*$  vendrá dado por

$$b^* = \frac{1}{2} \frac{p(x)}{1 - p(x)} \int_{-\infty}^{+\infty} \frac{f(x - \theta)}{f(x - \theta_0)} d\theta. \tag{3.10}$$



**Figura 3.1:** Ínfimo de la probabilidad final para  $b = 0.1, 0.2, 0.3$  y  $0.4$  y p-valor. Caso Normal

Además, si fijamos un valor de  $b$  próximo a  $b^*$ , como el  $\underline{\text{Pr}}(H_0^*|x)$  es función continua de  $b$ , obtendremos que el  $\underline{\text{Pr}}(H_0^*|x, b)$  queda próximo al  $\underline{\text{Pr}}(H_0^*|x, b^*)$  y, por tanto, próximo al p-valor. Todo ello se pone de manifiesto en el siguiente ejemplo

**Ejemplo 3.1** (Continuación). Los valores del ínfimo de la probabilidad final de la hipótesis nula puntual,  $H_0^*$ , quedan próximos a los correspondientes p-valores cuando  $b$  es moderadamente pequeño (ver tabla 3.1). Esta proximidad queda también reflejada en la figura 3.1 donde se ha representado gráficamente el p-valor junto con  $\inf_{\pi \in G_{US}} \text{Pr}(H_0^*|x, b)$  para  $b = 0.1, 0.2, 0.3$  y  $0.4$ . En la figura 3.1 se ha incluido también el ínfimo de la probabilidad final de  $H_0^*$  sobre  $G_{US}$  pero tomando  $\pi_0 = 1/2$  como la masa inicial asignada a la hipótesis nula. Podemos observar que esta última

**Tabla 3.2:** Valores de  $b$  que igualan el p-valor con el ínfimo de  $Pr(H_0^*|t)$  para una población Normal, ínfimo sobre  $G_{US}$ .

| $t$   | $p$ valor = $Pr(H_0^* t)$ | $b^*$ | $Pr(H_0 t, b)$ | $Pr(H_0^* t, \pi_0 = \frac{1}{2})$ |
|-------|---------------------------|-------|----------------|------------------------------------|
| 1.645 | 0.1                       | 0.170 | 0.1198         | 0.390                              |
| 1.960 | 0.05                      | 0.142 | 0.0575         | 0.290                              |
| 2.576 | 0.01                      | 0.111 | 0.0121         | 0.109                              |
| 3.291 | 0.001                     | 0.089 | 0.0011         | 0.018                              |

curva está lejos tanto del p-valor como de los  $\inf_{\pi \in G_{US}} Pr(H_0^*|x, b)$ .

Por otra parte, el valor de  $b^*$  dado por (3.10) es

$$b^* = \frac{1}{2} \frac{p(x)}{1 - p(x)} \frac{\sqrt{2\pi}\sigma}{\sqrt{n}} \exp \left\{ \frac{n}{2\sigma^2} (\bar{x} - \theta_0)^2 \right\},$$

que puesto en función de  $t = \sqrt{n}(|\bar{x} - \theta_0|)/\sigma$  resulta

$$b^* = \frac{1}{2} \frac{p(t)}{1 - p(t)} \frac{\sqrt{2\pi}\sigma}{\sqrt{n}} \exp \left\{ \frac{t^2}{2} \right\}. \tag{3.11}$$

La tabla 3.2 muestra, para algunos valores específicos de  $t$ , los valores de  $b^*$  que hacen que el ínfimo de  $Pr(H_0^*|x)$  sea igual al p-valor. De manera que si para efectuar el contraste de hipótesis nula puntual sustituimos éste por el contraste de tipo intervalo tomando un valor fijo de  $b$  próximo a  $b^*$ , entonces el ínfimo de la probabilidad final de  $H_0^*$  será aproximadamente igual al p-valor.

La equivalencia entre los contrastes de hipótesis nula puntual, (3.1), y de hipótesis nula de tipo intervalo, (3.2), queda también reflejada en la misma tabla donde, para los distintos valores de  $t$ , se ha calculado el ínfimo de  $Pr(H_0|t, b^*)$  dado por

**Tabla 3.3:** Razón entre  $\underline{Pr}(H_0^*|t, G_{US})$  y  $tp(t)$  en el caso Normal, con  $b = 0.125$ .

| t     | $\underline{Pr}(H_0^* t, G_{US})$ | $tp(t)$ | $\underline{Pr}(H_0^* t, G_{US})/(tp(t))$ |
|-------|-----------------------------------|---------|---|
| 1.645 | 0.0754                            | 0.1645  | 0.4584                                    |
| 1.960 | 0.0442                            | 0.0980  | 0.4510                                    |
| 2.576 | 0.0113                            | 0.0258  | 0.4387                                    |
| 3.291 | 0.0014                            | 0.0033  | 0.4254                                    |

$$\begin{aligned}
 \inf_{\pi \in G_{US}} Pr(H_0|t, b^*) &= \lim_{k \rightarrow \infty} Pr(H_0|t) \\
 &= \lim_{k \rightarrow \infty} \frac{\Phi\left(t + b^* \frac{\sqrt{n}}{\sigma}\right) - \Phi\left(t - b^* \frac{\sqrt{n}}{\sigma}\right)}{\Phi\left(t + k \frac{\sqrt{n}}{\sigma}\right) - \Phi\left(t - k \frac{\sqrt{n}}{\sigma}\right)} \\
 &= \Phi\left(t + b^* \frac{\sqrt{n}}{\sigma}\right) - \Phi\left(t - b^* \frac{\sqrt{n}}{\sigma}\right).
 \end{aligned}$$

Por último, también se ha recogido en la tabla 3.2 el ínfimo de la probabilidad final de  $H_0^*$  cuando  $\pi_0 = 1/2$ . Podemos, en este caso, observar la discrepancia que existe entre los p-valores y el ínfimo de la probabilidad final de  $H_0^*$  cuando se toma  $\pi_0 = 1/2$  en la distribución inicial mixta.  $\square$

Elegido un valor del radio del intervalo,  $b$ , resulta de interés estudiar la razón entre el  $\underline{Pr}(H_0^*|t, G_{US})$  y el producto  $tp(t)$ . En efecto, si tomamos  $b = 0.125$ , valor intermedio en la tabla 3.2, la última columna de la tabla 3.3 muestra que la razón entre  $\underline{Pr}(H_0^*|t, G_{US})$  y  $tp(t)$  es estable.

El comportamiento asintótico de la razón de estas dos cantidades queda reflejado de manera más precisa en el siguiente teorema.

**Teorema 3.3** Si  $b$  es el radio del intervalo en el contraste (3.2) y  $n$  el número de

observaciones, entonces

$$\frac{\inf_{\pi \in G_{US}} Pr(H_0^*|t)}{tp(t)} > b\sqrt{n}$$

Además,

$$\lim_{t \rightarrow +\infty} \frac{\inf_{\pi \in G_{US}} Pr(H_0^*|t)}{tp(t)} = b\sqrt{n}$$

DEMOSTRACIÓN:

De la expresión (3.6) obtenemos

$$\begin{aligned} \frac{\inf_{\pi \in G_{US}} Pr(H_0^*|t)}{tp(t)} &= \frac{b\sqrt{n}\varphi(t)}{(2b\sqrt{n}\varphi(t) + 1)t(1 - \Phi(t))} \\ &= \frac{b\sqrt{n}}{(2b\sqrt{n}\varphi(t) + 1)t \frac{1 - \Phi(t)}{\varphi(t)}}, \end{aligned} \quad (3.12)$$

donde  $\Phi$  y  $\varphi$  son, respectivamente, las funciones de distribución y densidad de la variable aleatoria  $N(0, 1)$ . Utilizando la desigualdad (Feller(1975), pg. 184),

$$1 - \frac{1}{t^2} < \frac{t(1 - \Phi(t))}{\varphi(t)}$$

que es válida para todo  $t > 0$ , se obtiene que

$$\frac{\inf_{\pi \in G_{US}} Pr(H_0^*|t)}{tp(t)} > \frac{b\sqrt{n}}{(2b\sqrt{n}\varphi(t) + 1)\left(1 - \frac{1}{t^2}\right)}. \quad (3.13)$$

Por último, el segundo miembro de (3.13) converge a  $b\sqrt{n}$  cuando  $t$  tiende a infinito, lo que concluye la demostración puesto que (3.12) es decreciente en  $t$ .  $\square$

Consecuencia del teorema anterior es que, cuando  $t$  es muy grande, el  $p$ -valor sobreestima la evidencia contra la hipótesis nula.

Además, observamos que cuando el tamaño muestral aumenta el valor de  $b$  debe disminuir para que las dos aproximaciones coincidan, resultado intuitivo y

que también ha sido puesto de manifiesto por Delampady y Berger(1987) respecto a los p-valores.

A la vista de los resultados obtenidos en esta sección y en la precedente, se pueden extraer las siguientes conclusiones:

En el problema del contraste de hipótesis nula puntual es posible reconciliar las medidas de evidencia clásicas y bayesianas ya que, como se ha puesto de manifiesto, es posible aproximar el ínfimo de la probabilidad final de la hipótesis nula puntual al correspondiente p-valor. Para ello basta elegir  $\pi_0$  en la distribución inicial como indica (3.3).

Eligiendo  $\pi_0$  de esta manera, obtenemos que el ínfimo de la probabilidad final de la hipótesis nula puntual,  $H_0^*$ , es aproximadamente igual al ínfimo de la probabilidad final de la hipótesis nula de tipo intervalo,  $H_0$ , si se elige  $b$ , radio del intervalo, suficientemente pequeño. Con lo que, desde el punto de vista bayesiano, tomaríamos la misma decisión en el contraste (3.1) que en el (3.2). Resultado que es coherente con la sustitución de un tipo de contraste por otro, como se ha comentado en la introducción de este capítulo.

Como en el problema de contraste de tipo intervalo (3.2) no se utiliza distribución inicial de tipo mixto, podemos concluir que la discrepancia observada por Berger y Sellke(1987) entre las respuestas clásica y bayesiana no se debe a la utilización de una distribución inicial de tipo mixto sino a la elección, en dicha distribución, de una masa  $\pi_0 = 1/2$  para el punto  $\theta = \theta_0$ .

A la hora de abordar un problema real de contraste de hipótesis nula puntual, creemos que si el decisor tiene claramente definido el valor de  $\pi_0$  entonces debe utilizar aquel sea cual sea. Ahora bien, si no tiene claro qué valor de  $\pi_0$  tomar, entonces debe, dentro de un contexto bayesiano, explicitar su función de densidad  $\pi(\theta)$ , elegir un radio de intervalo,  $b$ , adecuado para sustituir el problema puntual por el de intervalo y calcular  $Pr(H_0^*|x)$  utilizando  $\pi_0$  como en (3.3). Si, por último,

no puede determinar  $\pi(\theta)$  entonces debe tomar como distribución inicial la uniforme impropia sobre la recta real, ya que con esta es con la que se alcanza el ínfimo de la probabilidad final según se demuestra en el teorema 3.1.

Hasta aquí hemos expuesto los resultados obtenidos con el ínfimo de la probabilidad final resultante del teorema 3.1 cuando la verosimilitud es Normal. La distribución Normal está clasificada dentro de la clase de distribuciones resistentes a datos atípicos para el parámetro de localización, ver Maín (1987) y Gómez Villegas y Maín (1992). En este caso ya hemos visto que el ínfimo de la probabilidad final de la hipótesis nula puntual es aproximadamente igual que el correspondiente p-valor.

Cabe, entonces, preguntarse si estos resultados serán válidos caso de que la verosimilitud sea de una clase de distribuciones no resistente a datos atípicos. Los ejemplos siguientes tratan estas dos situaciones. En el primero se supone que las observaciones provienen de una variable aleatoria que sigue una distribución de Cauchy, que es propensa a observaciones atípicas, mientras que en el segundo suponemos que se han obtenido de una variable aleatoria doble exponencial, distribución que es neutra respecto a observaciones anómalas.

**Ejemplo 3.2.** Sea  $X$  variable aleatoria con distribución de  $Cauchy(\theta, \sigma^2)$ ,  $\theta$  parámetro desconocido que, en este caso, representa la mediana de la distribución y  $\sigma$  conocido. Por simplicidad, supondremos que  $\sigma = 1$ . En estas condiciones,

$$f(x - \theta) = \frac{1}{\pi} \frac{1}{1 + (x - \theta)^2}.$$

Sea  $X_1, \dots, X_n$  una muestra aleatoria simple de  $X$ . Entonces, la verosimilitud resulta

$$f(x_1, \dots, x_n) = \prod_{i=1}^n \frac{1}{\pi} \frac{1}{1 + (x_i - \theta)^2}.$$

Para contrastar  $H_0^* : \theta = 0$  frente a  $H_1^* : \theta \neq 0$  se tiene que el ínfimo de la

**Tabla 3.4:** p-valores e ínfimo de  $Pr(H_0^*|x)$ . Distribución de Cauchy, ínfimo sobre  $G_{US}$ .

|              | $M_1$   | $M_2$  | $M_3$   | $M_4$  | $M_5$   |
|--------------|---------|--------|---------|--------|---------|
| $\bar{x}$    | -0.9691 | 2.1472 | -2.9429 | 4.4188 | -7.8613 |
| $b=0.1$      | 0.1573  | 0.1090 | 0.0893  | 0.0780 | 0.1708  |
| $b=0.2$      | 0.2719  | 0.1966 | 0.1639  | 0.1447 | 0.2918  |
| $b=0.3$      | 0.3590  | 0.2685 | 0.2272  | 0.2023 | 0.3819  |
| $b=0.4$      | 0.4275  | 0.3286 | 0.2816  | 0.2528 | 0.4517  |
| $b=0.5$      | 0.4828  | 0.3796 | 0.3289  | 0.2972 | 0.5074  |
| $p(\bar{x})$ | 0.5100  | 0.2775 | 0.2084  | 0.1417 | 0.0805  |
| $b^*$        | 0.5575  | 0.3139 | 0.2687  | 0.1952 | 0.0425  |

probabilidad final de  $H_0^*$  dado por (3.5) es

$$\begin{aligned} \inf_{\pi \in G_{US}} Pr(H_0^*|x_1, \dots, x_n) &= \left( 1 + \frac{1}{2b} \int_{-\infty}^{+\infty} \frac{f(x_1, \dots, x_n|\theta)}{f(x_1, \dots, x_n|\theta_0)} d\theta \right)^{-1} \\ &= \left( 1 + \frac{1}{2b} \int_{-\infty}^{+\infty} \prod_{i=1}^n \frac{1 + (x_i - \theta_0)^2}{1 + (x_i - \theta)^2} d\theta \right)^{-1}. \end{aligned} \quad (3.14)$$

Si consideramos la región crítica  $\{|\bar{X}| \geq c\}$ , siendo  $\bar{X}$  la media muestral, el p-valor correspondiente a las observaciones  $x_1, \dots, x_n$  vendrá dado por

$$p(\bar{x}) = 1 - \frac{2}{\pi} \arctan(\bar{x} - \theta_0). \quad (3.15)$$

La tabla 3.4 recoge los valores del ínfimo de la probabilidad final dados por (3.14) para distintos valores de  $b$ , radio del intervalo, y un conjunto de cinco muestras de tamaño  $n = 10$  simuladas de una distribución de Cauchy(0,1). En la tabla se recoge, igualmente, el p-valor correspondiente al valor de  $\bar{x}$  en cada muestra.

Además, se obtienen los valores de  $b^*$  dados por (3.10) que hacen que el ínfimo de la probabilidad final de  $H_0^*$  se iguale al p-valor. Lo mismo que sucedía en el

**Tabla 3.5:** Razón entre  $xPr(H_0^*|x, G_{US})$  y  $p(x)$  en el caso Cauchy,  $b = 0.2$ .

| $x$ | $xPr(H_0^*, x, G_{US})$ | $p(x)$ | $xPr(H_0^* x, G_{US})/p(x)$ |
|-----|-------------------------|--------|-----------------------------|
| 1   | 0.0599                  | 0.5    | 0.1198                      |
| 1.5 | 0.0566                  | 0.3743 | 0.1511                      |
| 2   | 0.0496                  | 0.2952 | 0.1680                      |
| 2.5 | 0.0433                  | 0.2422 | 0.1786                      |
| 3   | 0.0378                  | 0.2048 | 0.1846                      |

caso de verosimilitud Normal, fijando un valor del radio del intervalo próximo a este valor de  $b^*$ , se obtiene que el p-valor es aproximadamente igual que el ínfimo de la probabilidad final de la hipótesis nula puntual. En cualquier caso, si  $b \leq b^*$  entonces dicho ínfimo será menor o igual que el p-valor correspondiente.

Por otro lado, tomando valores pequeños de  $b$  resulta que los ínfimos de la probabilidad final de  $H_0^*$  y el de  $H_0$  son muy parecidos. En efecto, si tomamos, por ejemplo,  $b = 0.1$  los valores del ínfimo de la probabilidad final de  $H_0$  para cada una de las muestras son 0.1848, 0.1217, 0.0980, 0.0844 y 0.2038 respectivamente. Esos valores son aproximadamente iguales a los del ínfimo de la probabilidad final de  $H_0^*$ . Ahora bien, sólo cuando  $|\bar{x}|$  toma valores “grandes” coinciden las cotas inferiores de la probabilidad final con el p-valor ya que si fijamos los valores de  $b^*$  obtenidos en la tabla 3.4 resulta que  $Pr(H_0|x_1, \dots, x_n, G_{US})$  toma los valores 0.7972 0.3258, 0.2633, 0.1640 y 0.0875 respectivamente para las muestras  $M_1, \dots, M_5$ , que se aproximan a los p-valores sólo en los últimos casos.

Para estudiar el comportamiento asintótico del  $Pr(H_0^*|x, G_{US})$  con respecto al p-valor, resultará de interés observar la razón entre  $xPr(H_0^*|x, G_{US})$  y el p-valor,  $p(x)$ , que puede verse en la tabla 3.5. Se ha considerado una única observación.

Si  $x$  es estrictamente positivo, el cociente  $xPr(H_0^*|x, G_{US})/p(x)$  es creciente y está acotado por  $b$ . el comportamiento asintótico queda reflejado en el siguiente teorema.

**Teorema 3.4** . Si  $x$  es estrictamente positivo y  $b$  es el radio del intervalo elegido para el contraste (3.2), entonces

$$\frac{xPr(H_0^*|x, G_{US})}{p(x)} < b$$

y, además

$$\lim_{x \rightarrow \infty} \frac{xPr(H_0^*|x, G_{US})}{p(x)} = b.$$

DEMOSTRACIÓN:

Aplicando la regla de L'Hopital para el cálculo de límites indeterminados, en este caso de la forma 0/0, se tiene

$$\begin{aligned} \lim_{x \rightarrow \infty} \frac{xPr(H_0^*|x, G_{US})}{p(x)} &= \lim_{x \rightarrow \infty} \frac{2bx}{2b + \pi(1 + x^2)} \\ &= \lim_{x \rightarrow \infty} \frac{2}{1 - \frac{2}{\pi} \arctan x} \\ &= \lim_{x \rightarrow \infty} \frac{-b\pi(1 + x^2)(4b + 2\pi - 2\pi x^2)}{2(2b + \pi(1 + x^2))^2} = b. \square \end{aligned}$$

Es decir, el p-valor es, en este caso, asintóticamente mucho más grande que el ínfimo de la probabilidad final de la hipótesis nula puntual.

Para concluir esta sección, consideramos el caso en que las observaciones provengan de una distribución doble exponencial, distribución que es neutra respecto a datos atípicos (Maín, 1988).

**Ejemplo 3.3.** Sea  $X$  variable aleatoria con distribución doble exponencial, es decir

$$f(x - \theta) = \frac{1}{2\sigma} \exp \left\{ -\frac{|x - \theta|}{\sigma} \right\}$$

con  $\sigma > 0$  conocido.

**Tabla 3.6:** p-valores e ínfimo de  $Pr(H_0^*|x)$ . Caso doble exponencial e ínfimo sobre  $G_{US}$ .

|         | x      |        |        |        |        |
|---------|--------|--------|--------|--------|--------|
|         | 1      | 1.5    | 2      | 2.5    | 3      |
| $b=0.1$ | 0.0355 | 0.0218 | 0.0134 | 0.0081 | 0.0049 |
| $b=0.2$ | 0.0685 | 0.0427 | 0.0264 | 0.0162 | 0.0099 |
| $b=0.3$ | 0.0995 | 0.0627 | 0.0390 | 0.0240 | 0.0147 |
| $b=0.4$ | 0.1283 | 0.0819 | 0.0514 | 0.0318 | 0.0195 |
| $p(x)$  | 0.3678 | 0.2231 | 0.1353 | 0.0821 | 0.0498 |

El ínfimo de la probabilidad final de  $H_0^*$  sobre la clase  $G_{US}$  dado por 3.5 es

$$\inf_{\pi \in G_{US}} Pr(H_0^*|x) = \left( 1 + \frac{\sigma}{b} \exp \left\{ \frac{|x - \theta_0|}{\sigma} \right\} \right)^{-1}. \tag{3.16}$$

La tabla 3.6 recoge, con  $\sigma = 1$ , los valores del ínfimo dado por 3.16 para algunos  $x$  y diferentes  $b$ . La última fila muestra los p-valores corespondientes a cada observación, dados por

$$p(x) = 2 \int_{|x|}^{\infty} \frac{1}{2\sigma} \exp \left\{ -\frac{|u - \theta_0|}{\sigma} \right\} du.$$

Al igual que sucedió en el ejemplo 3.2, observamos que el p-valor es mucho más grande que el ínfimo de la probabilidad final de la hipótesis nula puntual.

Igualmente los valores de  $b^*$  dados por 3.10, que hacen que el ínfimo de la probabilidad final de  $H_0^*$  sea igual al p-valor, son elevados de forma que no tendrán interés desde un punto de vista estadístico. Así, por ejemplo, si  $x = 1$  resulta  $b^* = 1.582$ .

No obstante, tomando valores moderados de  $b$ , radio del intervalo, el ínfimo de la probabilidad final de  $H_0^*$  es aproximadamente igual al ínfimo de la probabilidad final de  $H_0$ . En efecto, si elegimos  $b = 0.2$  y  $x$  toma los valores 1, 1.5, 2, 2.5 y 3 resulta que  $\underline{Pr}(H_0|x, G_{US})$  es 0.0741, 0.0449, 0.02725, 0.01653 y 0.0100 respectivamente.

**Tabla 3.7:** Razón entre  $\underline{Pr}(H_0^*|x, G_{US})$  y  $p(x)$  en el caso doble exponencial,  $b = 0.2$ .

| $x$ | $\underline{Pr}(H_0^* x, G_{US})$ | $p(x)$ | $\underline{Pr}(H_0^* x, G_{US})/p(x)$ |
|-----|-----------------------------------|--------|--|
| 1   | 0.0685                            | 0.3678 | 0.1863                                 |
| 1.5 | 0.0427                            | 0.2231 | 0.1915                                 |
| 2   | 0.0264                            | 0.1353 | 0.1951                                 |
| 2.5 | 0.0162                            | 0.0821 | 0.1973                                 |
| 3   | 0.0099                            | 0.0498 | 0.1987                                 |

Valores que son muy parecidos a los del ínfimo de la probabilidad final de  $H_0^*$  como se puede observar en la tabla 3.6.

En cuanto al comportamiento asintótico del ínfimo de la probabilidad final de la hipótesis nula puntual con respecto al  $p$ -valor tenemos que fijado un valor del radio del intervalo,  $b$ , la tabla 3.7 nos muestra los valores del cociente  $\underline{Pr}(H_0^*|x, G_{US})/p(x)$  para distintos  $x$ . Se puede observar que a medida que  $x$  crece el cociente se estabiliza cerca de  $b = 0.2$ . Lo mismo sucede fijando cualquier otro valor de  $b$ . Este comportamiento queda recogido en el siguiente teorema

**Teorema 3.5** Dado  $b > 0$  y para todo  $x$ , se tiene que

$$\frac{\underline{Pr}(H_0^*|x, G_{US})}{p(x)} < \frac{b}{\sigma}$$

y, además,

$$\lim_{x \rightarrow \infty} \frac{\underline{Pr}(H_0^*|x, G_{US})}{p(x)} = \frac{b}{\sigma}.$$

DEMOSTRACIÓN: El cociente

$$\begin{aligned} \frac{\underline{Pr}(H_0^*|x, G_{US})}{p(x)} &= \frac{\left(1 + \frac{\sigma}{b} \exp\left\{\frac{|x - \theta_0|}{\sigma}\right\}\right)^{-1}}{\exp\left\{\frac{-|x - \theta_0|}{\sigma}\right\}} \\ &= \frac{\exp\left\{\frac{|x - \theta_0|}{\sigma}\right\}}{1 + \frac{\sigma}{b} \exp\left\{\frac{|x - \theta_0|}{\sigma}\right\}} \end{aligned}$$

es creciente en  $x$  y converge a  $b/\sigma$  cuando  $x$  tiende a infinito.  $\square$

De manera que el p-valor y el ínfimo de la probabilidad final de  $H_0^*$  son infinitésimos del mismo orden. Ahora bien, si, por ejemplo,  $\sigma = 1$  y  $b = 0.2$ , el p-valor será, asintóticamente, cinco veces mayor que el ínfimo de la probabilidad final.

La conclusión que obtenemos a la vista de los dos últimos casos tratados es que cuando las observaciones provienen de una distribución de Cauchy o de una doble exponencial, los p-valores son mayores que las cotas inferiores obtenidas para la probabilidad final de la hipótesis nula puntual. Este resultado puede, tal vez, relacionarse con la clasificación de estas distribuciones respecto a observaciones atípicas.

# Capítulo 4

## Distribuciones iniciales $\varepsilon$ -contaminadas

### 4.1 Introducción

Como ya se ha indicado en los capítulos precedentes, para efectuar un análisis bayesiano relativo a un parámetro desconocido,  $\theta$ , es preciso explicitar las opiniones iniciales mediante una distribución de probabilidad que recoja, de alguna manera, la información disponible acerca de  $\theta$ .

Una forma interesante de describir esa información inicial es considerar la clase de distribuciones  $\varepsilon$ -contaminadas, que se describe de la siguiente manera

$$G = \{\pi = (1 - \varepsilon)\pi_0 + \varepsilon q, q \in Q\} \quad (4.1)$$

donde  $\pi_0$  es la distribución que se utilizaría en un análisis bayesiano con una única distribución a priori. Es decir, la distribución inicial en la que uno realmente cree.  $Q$  es la clase de distribuciones de probabilidad que representa las posibles desviaciones (contaminaciones) de  $\pi_0$ . Mientras que  $\varepsilon$ ,  $0 < \varepsilon < 1$ , representa el nivel de

contaminación que se quiere introducir en  $\pi_0$ .

La clase de las distribuciones  $\varepsilon$ -contaminadas es una clase que se muestra sensible al proceso de elicitación de la distribución inicial. Al finalizar el proceso de elección de una distribución inicial se puede llegar, de forma razonable, a concretar una distribución,  $\pi_0(\theta)$ , que represente nuestra opinión inicial acerca del comportamiento del parámetro  $\theta \in \Theta$ . Ahora bien, a lo largo del proceso que se ha llevado a cabo para determinar  $\pi_0$ , se habrán realizado determinados juicios probabilísticos acerca del comportamiento de subconjuntos de  $\Theta$ . Parece razonable pensar que estos juicios tendrán un cierto error,  $\varepsilon$ . Es por ello igualmente razonable hacer que la clase de las distribuciones iniciales a considerar,  $G$ , recoja las distribuciones con esos posibles errores cometidos al determinar  $\pi_0$ .

No todas las clases de distribuciones iniciales que han sido consideradas en la literatura son sensibles a esa posibilidad de error. Por ejemplo la clase de las distribuciones iniciales conjugadas es demasiado limitada, sobre todo porque es poco flexible respecto al comportamiento de las colas (Berger, 1985). Igualmente, las clases que exigen restricciones acerca de los momentos fuerzan a su vez importantes restricciones sobre las colas de las distribuciones iniciales permitidas.

Por otra parte, con la clase  $G$  definida en (4.1) es relativamente cómodo trabajar y resulta muy flexible tanto desde la variedad de clases de contaminaciones que se pueden considerar como de la cantidad de error que se quiera introducir.

Para la clase  $Q$  se pueden contemplar distintas posibilidades. Berger(1994) propone varias clases de contaminaciones, entre otras clases de distribuciones iniciales, para efectuar un análisis bayesiano robusto. Nosotros limitaremos nuestro estudio a tres tipos de contaminaciones:

1) La clase  $Q_T = \{\text{Todas las distribuciones de probabilidad}\}$ , clase utilizada por Huber(1973) para calcular el rango de la probabilidad final de un conjunto. Asimismo, Sivaganesan(1988) emplea esta clase para determinar el rango de la media y la

varianza de la distribución final.

La clase de todas las distribuciones de probabilidad resulta interesante porque contiene toda distribución “próxima” a  $\pi_0$  y, además, es sencilla de manejar. Sin embargo, puede parecer poco razonable en la práctica por ser demasiado amplia y contener distribuciones muy “alejadas” de  $\pi_0$ .

II) La clase  $Q_U = \{ \text{Todas las distribuciones de probabilidad que son unimodales con la misma moda que } \pi_0 \}$ , que se considera más razonable que  $Q_T$  cuando  $\pi_0$  es unimodal. Ver, entre otros, Berger y Berliner(1986), Berger(1985) y Sivaganesan y Berger(1989). Y

III) La clase  $Q_{US} = \{ \text{Todas las densidades de la forma } q(|\theta - \theta_0|), q \text{ no creciente} \}$ , esto es, la clase de las contaminaciones unimodales y simétricas. Clase que se considera particularmente razonable cuando  $\pi_0$  es, igualmente, unimodal y simétrica.

Las clases II y III permiten una variación muy amplia tanto en la forma funcional como en las colas de toda distribución  $\pi(\theta)$  de la clase  $G$  en (4.1), al mismo tiempo que mantienen las principales características de la distribución base  $\pi_0(\theta)$ . Conviene observar, además, que si la clase de contaminaciones es  $Q_{US}$  y ponemos  $\varepsilon = 1$ , se obtiene la clase  $G_{US}$ , utilizada en el capítulo 3.

Además de estas, otras clases de contaminaciones que han sido tratadas en la literatura son, por ejemplo, la clase  $Q = \{ q(\theta), \pi(\theta) = (1 - \varepsilon)\pi_0(\theta) + \varepsilon q(\theta) \text{ unimodal, con moda } \theta_0 \text{ y } \pi(\theta_0) \leq \pi_0(\theta_0) \}$ , introducida por Berger y Berliner(1986) y utilizada por Sivaganesan(1989) para calcular el rango de variación de la esperanza de la distribución final. Moreno y Cano(1991) fijan las probabilidades de conjuntos  $\Theta_i$  ( $i \geq 1$ ) que forman una partición finita o numerable del espacio paramétrico con la clase  $Q = \{ \text{medidas de probabilidad } q, q(\Theta_i) = p_i, i \geq 1 \}$ . Ellos utilizan esta clase para obtener cotas de la probabilidad final de cualquier conjunto medible.

Algunas clases de contaminaciones combinan información sobre los cuantiles con el aspecto de la función de densidad. Así, Moreno y Pericchi(1991) y Sivagane-

san(1991a) analizan el caso de contaminaciones unimodales con cuantiles especificados mientras que Sivaganesan(1991b) examina la robustez de la media, la mediana y la moda de la distribución final cuando las contaminaciones son unimodales con moda y mediana especificadas.

Consideremos, ahora, el contraste de hipótesis nula puntual

$$H_0^* : \theta = \theta_0 \quad \text{frente a} \quad H_1^* : \theta \neq \theta_0, \quad (4.2)$$

para cuya resolución nos apoyamos en la observación de una variable aleatoria,  $X$ , con densidad  $f(x - \theta)$  continua en  $\theta_0$ . Supongamos que la probabilidad asignada a  $\theta = \theta_0$  es  $p > 0$ .

Por otra parte, una hipótesis precisa se puede representar, ver Berger y Delampady(1987), como

$$H_0 : |\theta - \theta_0| \leq b \quad \text{frente a} \quad H_1 : |\theta - \theta_0| > b, \quad (4.3)$$

donde  $b$  es “pequeño”.

En el contraste (4.3) podemos utilizar una distribución inicial continua,  $\pi(\theta)$ , que representa nuestras opiniones iniciales acerca de  $\theta$ . De esta forma, como para efectuar el contraste (4.2) no es posible utilizar  $\pi(\theta)$ , aproximamos (4.2) por (4.3) eligiendo un valor adecuado de  $b$ , valor que dependerá del problema que estemos tratando, y calculamos  $p$ , probabilidad asignada a la hipótesis nula puntual en la distribución inicial mixta, como

$$p = \int_{|\theta - \theta_0| \leq b} \pi(\theta) d\theta \quad (4.4)$$

Utilizando  $p$  como indica (4.4) resulta que la distancia entre  $\pi(\theta)$  y la distribución mixta tiende a cero cuando  $b$  tiende a cero (ver capítulo 3).

Nuestro objetivo es encontrar cotas inferiores para la probabilidad final de la hipótesis nula puntual,  $H_0^*$ , sobre la clase de las distribuciones  $\varepsilon$ -contaminadas (4.1)

utilizando las clases de contaminaciones,  $Q$ , mencionadas anteriormente. Para ello necesitamos introducir algunas notaciones y fórmulas.

Designaremos por  $f(x|\theta)$  a la función de verosimilitud, función de  $\theta$  dada la observación  $x$ . La distribución marginal de  $X$  con respecto a la distribución inicial  $\pi \in G$  la notaremos por  $m(x|\pi)$ ,

$$m(x|\pi) = \int f(x|\theta)\pi(\theta)d\theta. \quad (4.5)$$

Entonces, si  $\pi \in G$ , como  $\pi(\theta) = (1 - \varepsilon)\pi_0(\theta) + \varepsilon q(\theta)$ , tenemos que

$$m(x|\pi) = (1 - \varepsilon)m(x|\pi_0) + \varepsilon m(x|q), \quad (4.6)$$

suponiendo que todas las cantidades involucradas existan.

En estas condiciones, la distribución final de  $\theta$ , dado  $x$ , con respecto a  $\pi$  viene dada por

$$\begin{aligned} \pi(\theta|x) &= \frac{f(x|\theta)\pi(\theta)}{m(x|\pi)} = \frac{f(x|\theta)((1 - \varepsilon)\pi_0(\theta) + \varepsilon q(\theta))}{m(x|\pi)} \\ &= \frac{(1 - \varepsilon)f(x|\theta)\pi_0(\theta) + \varepsilon f(x|\theta)q(\theta)}{m(x|\pi)} \\ &= \frac{1}{m(x|\pi)}(1 - \varepsilon)\frac{f(x|\theta)\pi_0(\theta)}{m(x|\pi_0)}m(x|\pi_0) + \varepsilon\frac{f(x|\theta)q(\theta)}{m(x|q)}m(x|q) \\ &= (1 - \varepsilon)\frac{m(x|\pi_0)}{m(x|\pi)}\pi_0(\theta|x) + \varepsilon\frac{m(x|q)}{m(x|\pi)}q(\theta|x), \end{aligned} \quad (4.7)$$

suponiendo que  $\pi_0(\theta|x)$ , densidad final de  $\pi_0$  dado  $x$ , y  $q(\theta|x)$ , densidad final de  $q$  dado  $x$ , existan.

De esta manera,  $\pi(\theta|x)$  se puede escribir

$$\pi(\theta|x) = \lambda(x)\pi_0(\theta|x) + (1 - \lambda(x))q(\theta|x) \quad (4.8)$$

donde  $\lambda(x) \in [0, 1]$  viene dado por

$$\lambda(x) = (1 - \varepsilon) \frac{m(x|\pi_0)}{m(x|\pi)}$$

que se puede considerar como la probabilidad final de que  $\pi_0$  sea la verdadera distribución final si se pensó que  $\pi_0$  o  $q$  eran las distribuciones iniciales correctas con probabilidades  $1 - \varepsilon$  y  $\varepsilon$  respectivamente.

Finalmente, de (4.4) tenemos que

$$p = (1 - \varepsilon)p_0 + \varepsilon q_0, \tag{4.9}$$

siendo

$$p_0 = \int_{|\theta - \theta_0| \leq b} \pi_0(\theta) d\theta, \quad y \quad q_0 = \int_{|\theta - \theta_0| \leq b} q(\theta) d\theta. \tag{4.10}$$

Debemos observar que, en principio, se puede suponer que  $p$  y  $p_0$  no están demasiado separadas, lo que significa que podríamos especificar  $p$  por  $p_0$  más (menos) un cierto "error".

La sección 4.2 se dedica a la obtención de una cota inferior para la probabilidad final de la hipótesis nula puntual y a la comparación de esta cota inferior con la medida clásica del  $p$ -valor correspondiente a las observaciones realizadas cuando la clase de contaminaciones esta formada por todas las distribuciones de probabilidad. En las secciones 4.3 y 4.4 estudiamos el mismo problema cuando la clase de contaminaciones está formada por todas las distribuciones unimodales y unimodales y simétricas, respectivamente.

## 4.2 Contaminaciones arbitrarias

En esta sección calculamos una cota inferior para la probabilidad final de la hipótesis nula en (4.2) cuando la distribución inicial pertenece a la clase de las distribuciones  $\varepsilon$ -contaminadas, (4.1), y  $\mathcal{Q}$  es la clase de todas las distribuciones de probabilidad.

Si consideramos que  $\pi_0(\theta)$  es nuestra densidad inicial, es decir aquella que representa nuestras opiniones iniciales, parece natural considerar como clase de contaminaciones la clase

$$Q_T = \{\text{Todas las distribuciones de probabilidad}\}$$

con las ventajas e inconvenientes que se indicaron en la introducción de este capítulo. En estas condiciones, obtenemos una cota inferior para la probabilidad final de  $H_0^*$  cuando la distribución inicial varia sobre la clase  $G_T = \{\pi = (1-\varepsilon)\pi_0 + \varepsilon q, q \in Q_T\}$ .

**Teorema 4.1** *Para el contraste de hipótesis nula puntual (4.2), si tomamos  $p$ , masa inicial asignada a la hipótesis nula, como se indica en (4.9) con  $q$  cualquier distribución de probabilidad, entonces*

$$Pr(H_0^*|x) \geq \left(1 + \frac{1 - (1 - \varepsilon)p_0}{(1 - \varepsilon)p_0} r(x)\right)^{-1}, \quad (4.11)$$

donde

$$r(x) = (1 - \varepsilon) \frac{m(x|\pi_0)}{f(x|\theta_0)} + \varepsilon \frac{\sup_{\theta \neq \theta_0} f(x|\theta)}{f(x|\theta_0)}$$

DEMOSTRACIÓN:

La probabilidad final de  $H_0^*$  se puede escribir

$$Pr(H_0^*|x) = \frac{f(x|\theta_0)}{f(x|\theta_0) + \frac{1-p}{p}m(x|\pi)}. \quad (4.12)$$

Obtener una cota inferior de (4.12) sobre la clase  $G_T$  equivale a calcular una cota superior para  $\frac{1-p}{p}m(x|\pi)$  cuando  $\pi \in G_T$ . Ahora bien,

$$\frac{1-p}{p}m(x|\pi) = \left[ \frac{1}{(1-\varepsilon)p_0 + \varepsilon q_0} - 1 \right] [(1-\varepsilon)m(x|\pi_0) + \varepsilon m(x|q)]$$

de manera que, si ponemos  $I_b = (\theta_0 - b, \theta_0 + b)$ , una cota superior se obtiene cuando

- (i)  $q_0 = 0$ . Esto es,  $q$  asigna probabilidad 0 al intervalo  $I_b$  y 1 a  $I_b^c$ . Y
- (ii)  $m(x|q)$  sea máxima para  $q \in Q$ .

Pero

$$m(x|q) = \int_{\Theta} f(x|\theta)q(\theta)d\theta \leq \sup_{\theta \neq \theta_0} f(x|\theta) \int_{\Theta} q(\theta)d\theta = \sup_{\theta \neq \theta_0} f(x|\theta).$$

Y, de aquí se obtiene inmediatamente (4.11).  $\square$

**Ejemplo 4.1.** Supongamos que  $X|\theta$  se distribuye según una  $N(\theta, \sigma^2)$ , con  $\sigma^2$  conocida, y que  $\pi_0(\theta)$  es  $N(\mu, \tau^2)$  con  $\mu$  y  $\tau^2$  conocidas. Consideremos  $X_1, \dots, X_n$  una muestra aleatoria simple de la población  $X$ . Entonces, sabemos que  $\bar{X}$  sigue una distribución  $N(\theta, \sigma^2/n)$ , que  $m(\bar{x}|\pi_0)$  es  $N(\mu, \sigma^2/n + \tau^2)$  y resulta que

$$\sup_{\theta \neq \theta_0} f(\bar{x}|\theta) = \frac{1}{\sigma\sqrt{2\pi}},$$

ya sea  $x = \theta_0$  o  $x \neq \theta_0$ . Además

$$p_0 = \int_{|\theta - \theta_0| \leq b} \pi_0(\theta) d\theta = \Phi\left(\frac{\theta_0 + b - \mu}{\tau}\right) - \Phi\left(\frac{\theta_0 - b - \mu}{\tau}\right), \quad (4.13)$$

siendo  $\Phi$  la función de distribución de la  $N(0, 1)$ . De esta manera, la probabilidad final de la hipótesis nula queda acotada inferiormente por (4.11), esto es

$$Pr(H_0^*|\bar{x}) \geq \left\{ 1 + \left( \frac{1}{(1-\varepsilon)p_0} - 1 \right) r(\bar{x}) \right\}^{-1} \quad (4.14)$$

donde

$$r(\bar{x}) = (1 - \varepsilon) \sqrt{\frac{\sigma^2}{\sigma^2 + n\tau^2}} \exp(A(\bar{x}) + B(\bar{x})) + \varepsilon \exp(B(\bar{x})) \quad (4.15)$$

con

$$A(\bar{x}) = -\frac{n(\bar{x} - \mu)^2}{2(\sigma^2 + n\tau^2)}, \quad \text{y} \quad B(\bar{x}) = \frac{n(\bar{x} - \theta_0)^2}{2\sigma^2}.$$

**Tabla 4.1:** Cotas inferiores para la probabilidad final de  $H_0^*$ . Verosimilitud Normal.

|         | t      |        |        |        |
|---------|--------|--------|--------|--------|
|         | 1.645  | 1.960  | 2.596  | 3.291  |
| $b=0.1$ | 0.0325 | 0.0189 | 0.0046 | 0.0006 |
| $b=0.2$ | 0.0657 | 0.0387 | 0.0097 | 0.0013 |
| $b=0.3$ | 0.0994 | 0.0595 | 0.0151 | 0.0021 |
| $b=0.4$ | 0.1335 | 0.0812 | 0.0209 | 0.0028 |
| $b=0.5$ | 0.1678 | 0.1037 | 0.0272 | 0.0038 |

La tabla 4.1 muestra, con  $\sigma^2 = 1$ ,  $\tau^2 = 2$ ,  $\theta_0 = \mu = 0$  y  $n = 10$  los valores de la cota inferior dada por (4.14) para algunos valores específicos de  $t = \sqrt{n}(\bar{x} - \theta_0)/\sigma$  y algunos  $b$ .  $\square$

Desde el punto de vista clásico, si existe un estadístico apropiado,  $T(X)$ , el p-valor (o nivel de significación) de los datos observados,  $x$ , es

$$p(x) = Pr(|T(X)| \geq |T(x)| | \theta = \theta_0).$$

En esta situación, existe un valor de  $b$ , digamos  $b^*$ , de tal manera que para ese valor de  $b$ , el p-valor coincide con la cota inferior de la probabilidad final de la hipótesis nula puntual,  $H_0^*$ , dada por (4.11). Para calcular este  $b^*$  será suficiente igualar el p-valor,  $p(x)$  a dicha cota inferior. Esto es,

$$p(x) = \left( 1 + \frac{1 - (1 - \varepsilon)p_0}{(1 - \varepsilon)p_0} r(x) \right)^{-1},$$

y entonces obtenemos

$$p_0 = \frac{1}{1 - \varepsilon} \left[ 1 + \frac{1 - p(x)}{p(x)r(x)} \right]^{-1}. \tag{4.16}$$

A partir de (4.16) podemos obtener, mediante cálculo numérico, el valor de  $b^*$ .

Además, como la cota inferior dada por (4.11) es función continua de  $b$ , si elegimos un valor de  $b$  próximo a  $b^*$ , los valores de dicha cota para  $b$  y  $b^*$  quedarán

próximos y, por tanto, aproximadamente igual al p-valor. Es decir, eligiendo un valor del radio del intervalo,  $b$ , que sea cercano a  $b^*$  obtendremos que las respuestas clásica y bayesiana al problema del contraste de hipótesis (4.2) serán numéricamente iguales.

Por otra parte, el ínfimo de la probabilidad final de  $H_0$ , en el contraste (4.3) sobre  $G$  viene dado por (Huber, 1973)

$$\inf_{\pi \in G_T} Pr(H_0|x) = \frac{Pr(H_0|x, \pi_0)}{1 + s(b)}, \quad (4.17)$$

donde

$$s(b) = \frac{\varepsilon \sup_{|\theta - \theta_0| > b} f(x|\theta)}{1 - \varepsilon m(x|\pi_0)}$$

y  $Pr(H_0|x, \pi_0)$  es la probabilidad final de  $H_0$  cuando la distribución inicial es  $\pi_0(\theta)$ , esto es

$$Pr(H_0|x, \pi_0) = \int_{|\theta - \theta_0| \leq b} \pi_0(\theta|x) d\theta.$$

**Ejemplo 4.1.** (Continuación). En este caso, el p-valor viene dado por

$$p(t) = 2(1 - \Phi(t)),$$

siendo  $\Phi$  la función de distribución de la  $N(0, 1)$ . Mientras que  $p_0$  y  $r(\bar{x})$  vienen dadas por (3.3) y (4.15) respectivamente.

La ecuación (4.16) proporciona los valores de  $b^*$  que hacen que la cota inferior de la probabilidad final de la hipótesis nula puntual sea igual al p-valor correspondiente a la observación realizada. La tabla 4.2 muestra dichos valores para  $n = 10$ ,  $\sigma^2 = 1$ ,  $\tau^2 = 2$ ,  $\theta_0 = \mu = 0$  ( $\theta_0$  y  $\mu$  no tienen por qué ser necesariamente iguales) y  $\varepsilon = 0.2$  para algunos valores específicos de  $t = \sqrt{n}(|\bar{x} - \theta_0|/\sigma)$ .

Además, para poner de manifiesto la equivalencia entre los contrastes (4.2) y (4.3), podemos calcular el ínfimo de  $Pr(H_0|x)$  sobre la clase  $G_T$  a partir de (4.17).

Es sencillo verificar que la distribución final  $\pi_0(\theta|x)$  es

$$N\left(\bar{x} - \frac{\sigma^2}{\sigma^2 + n\tau^2}(\bar{x} - \mu), \frac{\sigma^2\tau^2}{\sigma^2 + n\tau^2}\right),$$

con lo que la probabilidad final de  $H_0$  dados  $\pi_0$  y  $\bar{x}$  resulta

$$\begin{aligned} Pr(H_0|\bar{x}, \pi_0) &= \int_{|\theta - \theta_0| \leq b} \pi_0(\theta|\bar{x}) d\theta = \\ &\Phi\left(\frac{\sigma^2(\theta_0 + b - \mu) + n\tau^2(\theta_0 + b - \bar{x})}{\sigma\tau\sqrt{\sigma^2 + n\tau^2}}\right) - \Phi\left(\frac{\sigma^2(\theta_0 - b - \mu) + n\tau^2(\theta_0 - b - \bar{x})}{\sigma\tau\sqrt{\sigma^2 + n\tau^2}}\right). \end{aligned} \quad (4.18)$$

En cuanto al  $\sup_{|\theta - \theta_0| > b} f(x|\theta)$  tenemos que

(i) Si  $x \in I_b$  el supremo se alcanzará en  $\theta = \theta_0 - b$  o en  $\theta = \theta_0 + b$ . En este caso

$$\sup_{|\theta - \theta_0| > b} f(\bar{x}|\theta) = \frac{\sqrt{n}}{\sigma\sqrt{2\pi}} \exp -n \frac{(|x - \theta_0| - b)^2}{2\sigma^2}$$

(ii) Si  $x \in I_b^c$ , el supremo se alcanzará en  $\theta = x$  y entonces

$$\sup_{|\theta - \theta_0| > b} f(x|\theta) = \frac{\sqrt{n}}{\sigma\sqrt{2\pi}}.$$

De manera que de (4.17) obtenemos

$$\inf_{\pi \in G_T} Pr(H_0|\bar{x}) = \begin{cases} \frac{Pr(H_0|\bar{x}, \pi_0)}{1 + A(\varepsilon, \bar{x})}, & \text{si } \bar{x} \in I_b^c \\ \frac{Pr(H_0|\bar{x}, \pi_0)}{1 + A(\varepsilon, \bar{x}) \exp\left(-n \frac{(|\bar{x} - \theta_0| - b)^2}{2\sigma^2}\right)}, & \text{si } \bar{x} \in I_b \end{cases} \quad (4.19)$$

donde

$$A(\varepsilon, \bar{x}) = \frac{\varepsilon}{1 - \varepsilon} \sqrt{1 + \frac{n\tau^2}{\sigma^2} \exp\left\{\frac{n(\bar{x} - \mu)^2}{2(\sigma^2 + n\tau^2)}\right\}}$$

**Tabla 4.2:** Valores de  $b^*$  que hacen que el  $p$ -valor sea igual a  $\underline{Pr}(H_0|x)$ ,  $\underline{Pr}(H_0^*|x)$ . Infimo sobre  $G_T$ .

| t     | $p$ valor = $\underline{Pr}(H_0^* t)$ | $b^*$ | $\underline{Pr}(H_0 t)$ | $\underline{Pr}(H_0^* t, p = \frac{1}{2})$ |
|-------|---------------------------------------|-------|-------------------------|--|
| 1.645 | 0.100                                 | 0.302 | 0.1076                  | 0.4154                                     |
| 1.960 | 0.050                                 | 0.255 | 0.0564                  | 0.2896                                     |
| 2.596 | 0.010                                 | 0.207 | 0.0113                  | 0.0898                                     |
| 3.291 | 0.001                                 | 0.156 | 0.0012                  | 0.0131                                     |

y  $Pr(H_0|\bar{x}, \pi_0)$  viene dada por (4.19).

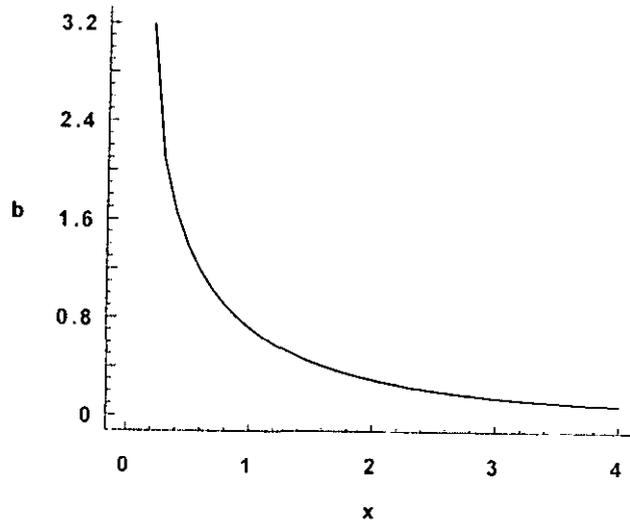
En la tabla 4.2 podemos observar que los valores del ínfimo de la probabilidad final de  $H_0$ , calculados a partir de (4.19), son aproximadamente iguales a los de la cota inferior de la probabilidad final de  $H_0^*$  obtenida de (4.11).

Obsérvese que si  $t = 2.596$ ,  $\underline{Pr}(H_0^*|t) = 0.010035$ , el  $p$ -valor es 0.01 y si se quiere contrastar  $H_0 : \theta \in I_{0.207}(0)$  frente a  $H_1 : \theta \in I_{0.207}^c(0)$  entonces las aproximaciones clásica y bayesiana coinciden, mientras que con  $p = \frac{1}{2}$  obtenemos  $\underline{Pr}(H_0^*|t, p = \frac{1}{2}) = 0.0898$ , lo que supone una fuerte discrepancia entre ambas aproximaciones.

Los valores de  $b^*$  dependen de las observaciones realizadas, a medida que estas crecen, los valores de  $b^*$ , radio del intervalo de la hipótesis nula en (4.3) se hacen más pequeños. La situación se describe gráficamente en la figura 4.1.

### 4.3 Contaminaciones unimodales

Elegir la clase  $Q_T$ , de todas las distribuciones, como clase de contaminaciones permite efectuar un análisis sencillo del problema pero, como ya mencionamos en la introducción de este capítulo, puede presentar algunos inconvenientes debido, sobre todo, a que algunas distribuciones de  $G_T$  pueden resultar poco realistas en la práctica.



**Figura 4.1:** Longitud del intervalo,  $b^*$ , como función de las observaciones

Cuando  $\Theta$  es un subconjunto de los números reales y  $\pi_0$  es unimodal resulta razonable considerar aquellas contaminaciones que sean, asimismo, unimodales. De este modo, en esta sección consideraremos la clase de las distribuciones  $\varepsilon$ -contaminadas (4.1), que notaremos por  $G_U$ , con la clase de contaminaciones

$$Q_U = \{ \text{Todas las distribuciones de probabilidad que son unimodales} \\ \text{con la misma moda que } \pi_0 \}$$

En esta ocasión, calcular el ínfimo de  $Pr(H_0^*|x)$  sobre  $G_U$  es igual que hacerlo sobre la clase (ver Sivaganesan y Berger(1989))

$$G'_U = \{ \pi(\theta) = (1 - \varepsilon)\pi_0(\theta) + \varepsilon q(\theta), q(\theta) \in U(\theta_0, \theta_0 + k) \text{ ó } q(\theta) \in U(\theta_0 - k, \theta_0), \\ \text{para algún } k > 0 \}.$$

Entonces, de acuerdo con la demostración del teorema 4.1, bastaría con encontrar el supremo de  $M(k) = \frac{1-p}{p} m(x|\pi)$  sobre la clase  $G'_U$ .

Suponiendo  $b < k$ , tenemos que

$$p = \int_{|\theta - \theta_0| \leq b} \pi(\theta) d\theta = (1 - \varepsilon)p_0 + \varepsilon \frac{b}{k}, \quad (4.20)$$

con  $p_0$  dado por (4.10), mientras

$$m(x|q) = \int f(x|\theta)q(\theta)d\theta = \frac{1}{k} \int_{\theta_0}^{\theta_0+k} f(x|\theta)d\theta \quad (4.21)$$

si  $q(\theta)$  es uniiiforme en el intervalo  $(\theta_0, \theta_0 + k)$ , y

$$m(x|q) = \frac{1}{k} \int_{\theta_0-k}^{\theta_0} f(x|\theta)d\theta \quad (4.22)$$

si  $q(\theta)$  es uniforme en  $(\theta_0 - k, \theta_0)$ .

Entonces, teniendo en cuenta (4.6), (4.21) y (4.22), tenemos que

$$M(k) = \begin{cases} \frac{1-p}{p} \left( A + \frac{\varepsilon}{k} \int_{\theta_0}^{\theta_0+k} f(x|\theta)d\theta \right) & \text{si } q(\theta) \in U(\theta_0, \theta_0 + k) \\ \frac{1-p}{p} \left( A + \frac{\varepsilon}{k} \int_{\theta_0-k}^{\theta_0} f(x|\theta)d\theta \right) & \text{si } q(\theta) \in U(\theta_0 - k, \theta_0). \end{cases} \quad (4.23)$$

donde  $A = (1 - \varepsilon)m(x|\pi_0)$  y  $p$  viene dado por (4.20). Expresión que, fijado  $b$  y observado  $x$ , depende sólo de  $k$ . Para encontrar el valor de  $k$  que maximiza (4.23), obtenemos  $M'(k)$  e igualamos a cero. Es relativamente sencillo comprobar que se obtienen las siguientes ecuaciones

$$(1 - \varepsilon)bm(x|\pi_0) + kp(1 - p)f(x|\theta_0 + k) + [p^2 - (1 - \varepsilon)p_0] \int_{\theta_0}^{\theta_0+k} f(x|\theta)d\theta = 0 \quad (4.24)$$

si  $q(\theta)$  es uniforme en  $(\theta_0, \theta_0 + k)$ , y

$$(1 - \varepsilon)bm(x|\pi_0) + kp(1 - p)f(x|\theta_0 + k) + [p^2 - (1 - \varepsilon)p_0] \int_{\theta_0-k}^{\theta_0} f(x|\theta)d\theta = 0 \quad (4.25)$$

si  $q(\theta)$  es uniforme en  $(\theta_0 - k, \theta_0)$ .

Las ecuaciones (4.24) y (4.25) tendrán solución  $k \geq 0$  sólo si

(i)  $p^2 - (1 - \varepsilon)p_0 < 0$ , y

(ii)  $\left| [p^2 - (1 - \varepsilon)p_0] \int_{\theta_0}^{\theta_0+k} f(x|\theta) d\theta \right| > (1 - \varepsilon)bm(x|\pi_0)$ .

Debe observarse que la condición (i) es razonable puesto que si se considera que  $p$  es aproximadamente igual a  $p_0$ , dicha condición equivale a exigir que  $p$  sea menor que  $1 - \varepsilon$  que parece adecuado sobre todo si se piensa en una cantidad de error que no sea elevada. No obstante si  $\varepsilon = 1$  o resulta que  $p^2 - (1 - \varepsilon)p_0 \geq 0$ , tendríamos que  $M'(k)$  sería positiva en todo caso y, por lo tanto, el supremo se alcanzará cuando  $k$  tienda a infinito.

De (4.24) ó (4.25), obtendremos el valor de  $k$  que maximiza  $M(k)$  y, de esta manera, encontraremos el valor de  $\underline{Pr}(H_0^*|x) = \inf_{\pi \in G_U} Pr(H_0^*|x)$ . Por ejemplo, de (4.24), el valor de  $k$  que maximiza  $M(k)$  viene dado por la solución de ( $k > 0$ )

$$k = \frac{[(1 - \varepsilon)p_0 - p^2] \int_{\theta_0}^{\theta_0+k} f(x|\theta) d\theta - (1 - \varepsilon)bm(x|\pi_0)}{p(1 - p)f(x|\theta_0 + k)}, \tag{4.26}$$

ecuación que se aproxima numéricamente.

**Ejemplo 4.2.** Sea  $X_1, \dots, X_n$  una muestra aleatoria simple de una población  $N(\theta, \sigma^2)$ , con  $\sigma^2$  conocida. Supongamos que  $\pi_0(\theta)$  es  $N(\mu, \tau^2)$ , con  $\mu$  y  $\tau^2$  conocidas. Entonces  $m(\bar{x}|\pi_0)$ , la distribución marginal de  $\bar{X}$  dado  $\pi_0$ , es  $N(\mu, \tau^2 + \sigma^2/n)$ . Además

$$p_0 = \int_{\theta_0-b}^{\theta_0+b} \pi_0(\theta) d\theta = \Phi\left(\frac{\theta_0 + b - \mu}{\tau}\right) - \Phi\left(\frac{\theta_0 - b - \mu}{\tau}\right),$$

$$\int_{\theta_0}^{\theta_0+k} f(\bar{x}|\theta) d\theta = \Phi\left(\frac{\theta_0 + k - \bar{x}}{\sigma} \sqrt{n}\right) - \Phi\left(\frac{\theta_0 - \bar{x}}{\sigma} \sqrt{n}\right),$$

$$f(\bar{x}|\theta_0 + k) = \frac{\sqrt{n}}{\sigma\sqrt{2\pi}} \exp\left(-\frac{n(x - \theta_0 - k)^2}{2\sigma^2}\right)$$

**Tabla 4.3:** Valores de  $k$  en los que se alcanza el ínfimo de  $Pr(H_0^*|t)$  sobre  $G_U$  para algunos valores de  $t$  y  $b$ .

| $t$   | $b$ | $k$   | $\sup_{\pi \in G_U} M(k)$ | $\underline{Pr}(H_0 \bar{x}, G_U)$ |
|-------|-----|-------|---------------------------|------------------------------------|
| 1.645 | 0.1 | 1.081 | 5.52449                   | 0.05573                            |
|       | 0.2 | 1.112 | 2.58070                   | 0.11217                            |
|       | 0.3 | 1.150 | 1.60304                   | 0.16902                            |
| 1.960 | 0.1 | 1.175 | 5.43849                   | 0.03286                            |
|       | 0.5 | 1.201 | 2.54499                   | 0.06770                            |
|       | 1.0 | 1.233 | 1.58374                   | 0.10450                            |
| 2.596 | 0.1 | 1.375 | 5.14862                   | 0.00836                            |
|       | 0.2 | 1.394 | 2.41652                   | 0.01764                            |
|       | 0.3 | 1.420 | 1.50861                   | 0.02664                            |
| 3.291 | 0.1 | 1.598 | 4.74991                   | 0.00118                            |
|       | 0.2 | 1.614 | 2.23486                   | 0.00250                            |
|       | 0.3 | 1.632 | 1.39895                   | 0.00399                            |

y  $p$  viene dado por (4.20).

En estas condiciones, si  $n = 10$ ,  $\varepsilon = 0.2$ ,  $\sigma^2 = 1$ ,  $\tau^2 = 2$  y  $\theta_0 = \mu = 0$ , la tabla 4.3 muestra, para valores concretos de  $t = \sqrt{n}|\bar{x} - \theta_0|/\sigma$  y  $b$ , los valores de  $k$ , obtenidos de (4.26), para los cuales se alcanza el ínfimo de la probabilidad final de la hipótesis nula puntual.

La última columna se obtiene de

$$\underline{Pr}(H_0^*|\bar{x}, G_U) = \left[ 1 + \frac{\sup_{\pi \in G_U} M(k)}{f(\bar{x}|\theta_0)} \right]^{-1}. \tag{4.27}$$

Podemos obtener, ahora, los valores de  $b$  y  $k$ , digamos  $b^*$  y  $k^*$ , de manera que el  $\inf_{\pi \in G_U} P(H_0^*|\bar{x})$  coincida con el correspondiente  $p$ -valor. En efecto, igualando el  $p$ -valor,  $p(\bar{x})$ , a (4.27) obtenemos que

$$\frac{f(x|\theta_0)}{\sup_{\pi \in G_U} M(k)} = \frac{p(\bar{x})}{1 - p(\bar{x})}.$$

**Tabla 4.4:** Valores de  $b$  y  $k$  que hacen que el  $p$ -valor sea igual a  $\underline{Pr}(H_0|\bar{x})$ ,  $\underline{Pr}(H_0^*|\bar{x})$ , ínfimo sobre  $G_U$ .

| $t$   | $b^*$  | $k^*$ | $\underline{Pr}(H_0^* \bar{x}) = p\text{-valor}$ | $\underline{Pr}(H_0 \bar{x})$ | $\underline{Pr}(H_0 \bar{x}, p = \frac{1}{2})$ |
|-------|--------|-------|--|-------------------------------|--|
| 1.645 | 0.1785 | 1.122 | 0.1  | 0.10298                       | 0.36708  |
| 1.960 | 0.1498 | 1.180 | 0.05   | 0.05310                       | 0.27710  |
| 2.596 | 0.1184 | 1.381 | 0.01   | 0.01081                       | 0.11419  |
| 3.291 | 0.0855 | 1.589 | 0.001  | 0.00107                       | 0.02584  |

En la tabla (4.4) se muestran los valores de  $b^*$  y  $k^*$  para algunos  $t$ . Además, observamos nuevamente que  $\underline{Pr}(H_0^*|\bar{x}, G_U)$  es aproximadamente igual al  $\underline{Pr}(H_0|\bar{x}, G_U)$ .

Este último obtenido a partir de

$$Pr(H_0|\bar{x}) = \int_{\theta_0-b}^{\theta_0+b} \pi(\theta|\bar{x}) d\theta,$$

donde  $\pi(\theta|\bar{x})$  viene dado por (4.8). Mientras que  $m(\bar{x}|\pi_0)$  es una distribución  $N(\mu, \tau^2 + \sigma^2/n)$ ,  $\pi_0(\theta|\bar{x})$  se distribuye

$$N\left(\bar{x} - \frac{\sigma^2}{\sigma^2 + n\tau^2}(\bar{x} - \mu), \frac{\sigma^2\tau^2}{\sigma^2 + n\tau^2}\right),$$

y

$$m(\bar{x}|q) = \frac{1}{k} \left[ \Phi\left(\frac{\theta_0 + k - \bar{x}}{\sigma} \sqrt{n}\right) - \Phi\left(\frac{\theta_0 - \bar{x}}{\sigma} \sqrt{n}\right) \right]. \square$$

Lo mismo que sucedía en el caso de contaminaciones arbitrarias, la cota inferior obtenida para la probabilidad final de la hipótesis nula puntual es aproximadamente igual al  $p$ -valor cuando se elige  $b$  próximo a  $b^*$ . En esta ocasión, la cota inferior se obtiene cuando la clase de contaminaciones es uniforme en  $(\theta_0, \theta_0 + k^*)$  ó  $(\theta_0 - k^*, \theta_0)$ , con  $k^*$  obtenido como solución de la ecuación (4.26). Nuevamente se observa que si se elige  $p = \frac{1}{2}$  el ínfimo de la probabilidad final de  $H_0^*$  es considerablemente mayor que el  $p$ -valor.

## 4.4 Contaminaciones unimodales y simétricas

En esta sección trataremos el problema del contraste de hipótesis precisa cuando la distribución inicial pertenece a la clase de las distribuciones  $\varepsilon$ -contaminadas (4.1) con la clase de contaminaciones formada por

$$Q = \{ \text{Todas distribuciones unimodales y simétricas con la misma moda que } \pi_0 \}$$

Llamaremos  $G_{US}$  a esta clase de distribuciones  $\varepsilon$ -contaminadas.

Esta clase de distribuciones parece apropiada, sobre todo, cuando la distribución inicial  $\pi_0$  es, igualmente, unimodal y simétrica. De esta manera se evitan contaminaciones que den excesivo peso a valores alejados de  $\theta_0$ .

Utilizaremos el resultado conocido de que toda densidad unimodal y simétrica se puede expresar como mixtura de densidades uniformes y simétricas.

Sivaganesan y Berger(1989) calculan el ínfimo de la probabilidad final de la hipótesis nula puntual con este tipo de distribuciones iniciales. Ellos asignan una masa inicial de  $1/2$  a  $\theta_0$  y distribuyen la otra mitad de acuerdo con una distribución inicial,  $\pi(\theta)$ , de la clase  $G_{US}$ .

En las mismas condiciones que en la sección anterior, el ínfimo de la  $Pr(H_0^*|x)$  se obtendrá calculando el supremo de  $\frac{1-p}{p}m(x|\pi)$ . Y, como ya se ha adelantado, maximizar esta cantidad sobre la clase  $G_{US}$  es lo mismo que hacerlo sobre la clase

$$G'_{US} = \{ \pi = (1 - \varepsilon)\pi_0 + \varepsilon q, q \text{ uniforme en } (\theta_0 - k, \theta_0 + k), k \geq 0 \}.$$

En estas condiciones, tenemos que la probabilidad inicial asignada a la hipótesis nula,  $p$ , dada por (4.9) es

$$p = (1 - \varepsilon) \int_{\theta_0 - b}^{\theta_0 + b} \pi_0(\theta) d\theta + \varepsilon \frac{b}{k}, \quad (4.28)$$

donde suponemos  $b \leq k$ . Y la distribución marginal de  $X$  dado  $\pi$  es

$$\begin{aligned} m(x|\pi) &= (1 - \varepsilon)m(x|\pi_0) + \varepsilon m(x|q) \\ &= (1 - \varepsilon)m(x|\pi_0) + \frac{\varepsilon}{2k} \int_{\theta_0 - k}^{\theta_0 + k} f(x|\theta) d\theta. \end{aligned} \quad (4.29)$$

De manera que, de (4.28) y (4.29), tenemos

$$M(k) = \left( \frac{1}{p} - 1 \right) m(x|\pi) = \quad (4.30)$$

$$\frac{k(1 - \varepsilon)m(x|\pi_0) + \frac{\varepsilon}{2} \int_{\theta_0 - k}^{\theta_0 + k} f(x|\theta) d\theta}{k(1 - \varepsilon)p_0 + \varepsilon b} - (1 - \varepsilon)m(x|\pi_0) - \frac{\varepsilon}{2k} \int_{\theta_0 - k}^{\theta_0 + k} f(x|\theta) d\theta.$$

Expresión que, fijado el radio del intervalo,  $b$ , y dado  $x$ , depende sólo de  $k$ . De manera que para obtener el valor de  $k$  que maximiza (4.30) calculamos la derivada de  $M(k)$  e igualamos a cero, resultando la siguiente ecuación

$$\begin{aligned} (1 - \varepsilon)bm(x|\pi_0) + \frac{1}{2}kp(1 - p)(f(x|\theta_0 + k) + f(x|\theta_0 - k)) \\ + \frac{1}{2}(p^2 - (1 - \varepsilon)p_0) \int_{\theta_0 - k}^{\theta_0 + k} f(x|\theta) d\theta = 0 \end{aligned} \quad (4.31)$$

siendo

$$p = (1 - \varepsilon)p_0 + \varepsilon \frac{b}{k} \quad \text{y} \quad p_0 = \int_{\theta_0 - b}^{\theta_0 + b} \pi_0(\theta) d\theta. \quad (4.32)$$

ecuación que es similar a la obtenida en la Sección 4.3 para el caso de contaminaciones unimodales. A partir de ella obtenemos el valor de  $k$  que maximiza  $M(k)$  y, por tanto, aquel donde se alcanza el ínfimo de la probabilidad final de la hipótesis nula puntual.

**Ejemplo 4.3.** Consideremos, al igual que en los ejemplos anteriores, que  $X$  es una variable aleatoria con distribución  $N(\theta, \sigma^2)$ , con  $\sigma^2$  conocida. Supongamos que  $\pi_0(\theta)$  es  $N(\mu, \tau^2)$ , con  $\mu$  y  $\tau^2$  conocidas. Sea  $X_1, \dots, X_n$  una muestra aleatoria simple de

**Tabla 4.5:** Valores de  $k$  para los que se alcanza el ínfimo de la probabilidad final de  $H_0^*$  para algunos valores de  $t$  y  $b$ . Ínfimo sobre  $G_{US}$

| $t$   | $b$ | $k$       | $\sup M(k)$ | $\underline{Pr}(H_0^* \bar{x})$ |
|-------|-----|-----------|-------------|---------------------------------|
| 1.645 | 0.1 | 1.570     | 4.40095     | 0.06898                         |
|       | 0.2 | $+\infty$ | 2.08649     | 0.13503                         |
|       | 0.3 | $+\infty$ | 1.32996     | 0.19689                         |
| 1.960 | 0.1 | 1.546     | 4.31036     | 0.04111                         |
|       | 0.2 | 5.073     | 2.03466     | 0.08327                         |
|       | 0.3 | $+\infty$ | 1.29449     | 0.12493                         |
| 2.596 | 0.1 | 1.634     | 4.08511     | 0.01051                         |
|       | 0.2 | 1.782     | 1.92363     | 0.02206                         |
|       | 0.3 | 4.477     | 1.21042     | 0.03462                         |
| 3.291 | 0.1 | 1.785     | 3.78266     | 0.00148                         |
|       | 0.2 | 1.846     | 1.78289     | 0.00314                         |
|       | 0.3 | 1.963     | 1.11884     | 0.00499                         |

X. Entonces, sabemos que  $m(\bar{x}|\pi_0)$  se distribuye según  $N(\mu, \tau^2 + \sigma^2/n)$ . Además, la probabilidad inicial viene dada por (4.32) con

$$p_0 = \Phi\left(\frac{\theta_0 + b - \mu}{\tau}\right) - \Phi\left(\frac{\theta_0 - b - \mu}{\tau}\right).$$

De esta manera, la ecuación (4.31) proporciona soluciones en  $k$  para valores concretos de  $\bar{x}$  y  $b$ . La tabla 4.5 recoge los valores de  $k$  en los que se alcanza el ínfimo de la probabilidad final de  $H_0^*$  para algunos valores específicos de  $t = \sqrt{n}|\bar{x} - \theta_0|/\sigma$  y diferentes valores de  $b$ . Para los cálculos se han tomado  $\epsilon = 0.2$ ,  $\sigma^2 = 1$ ,  $\tau^2 = 2$ ,  $\theta_0 = \mu = 0$  y  $n = 10$ .

La última columna de la tabla 4.5 se obtiene mediante

$$\underline{Pr}(H_0^*|\bar{x}, G_{US}) = \left(1 + \frac{\sup_{G_{US}} M(k)}{f(\bar{x}|\theta_0)}\right)^{-1}$$

**Tabla 4.6:** Valores de  $b$  que hacen que el  $p$ -valor sea igual a  $\underline{Pr}(H_0|\bar{x})$ ,  $\underline{Pr}(H_0^*|\bar{x})$ , ínfimo sobre  $G_{US}$ .

| $t$   | $b^*$  | $k^*$ | $\underline{Pr}(H_0^* \bar{x}) = p\text{-valor}$ | $\underline{Pr}(H_0 \bar{x})$ | $\underline{Pr}(H_0 \bar{x}, p = \frac{1}{2})$ |
|-------|--------|-------|--|-------------------------------|--|
| 1.645 | 0.1461 | 3.85  | 0.1  | 0.10811                       | 0.51364  |
| 1.960 | 0.1213 | 1.60  | 0.05   | 0.05115                       | 0.38817  |
| 2.596 | 0.0954 | 1.63  | 0.01   | 0.01014                       | 0.14231  |
| 3.291 | 0.0688 | 1.77  | 0.001  | 0.00104                       | 0.02357  |

Un par de observaciones pueden hacerse acerca de los resultados que muestra la tabla 4.5. En primer lugar, la distribución  $q(\theta)$  para la que se alcanza el ínfimo depende del valor fijado para  $b$  y de la observación realizada obteniéndose, en algunos casos, la cota inferior cuando la contaminación es la uniforme impropia. En segundo lugar, cuando  $b$  toma valores moderadamente pequeños el ínfimo de la probabilidad final es muy parecido al  $p$ -valor correspondiente a la observación realizada.

En consecuencia, al igual que sucedía en las secciones anteriores, es posible determinar un valor de  $b$ , al que nuevamente llamaremos  $b^*$ , de tal manera que el  $p$ -valor coincida con el ínfimo de la probabilidad final de la hipótesis nula puntual. Este valor de  $b$  depende, actualmente, del tamaño muestral, cuanto mayor es  $n$  más pequeño tiene que ser  $b$  para que las dos aproximaciones coincidan. La tabla (4.6) muestra los valores de  $b^*$  y los correspondientes de  $k$ , que hemos llamado  $k^*$ . En la tabla se recogen, también, los valores del ínfimo de la probabilidad final de la hipótesis de intervalo,  $H_0$ , que se han obtenido de

$$Pr(H_0|\bar{x}) = \int_{\theta_0-b}^{\theta_0+b} \pi(\theta|\bar{x}) d\theta,$$

donde  $\pi(\theta|\bar{x})$  viene dada por (4.7), de tal manera que

$$Pr(H_0|\bar{x}) = \frac{1 - \epsilon}{m(\bar{x}|\pi)} \int_{\theta_0-b}^{\theta_0+b} f(\bar{x}|\theta)\pi_0(\theta) d\theta + \frac{\epsilon}{m(\bar{x}|\pi)} \int_{\theta_0-b}^{\theta_0+b} \frac{f(\bar{x}|\theta)}{-2k} d\theta.$$

con  $f(\bar{x}|\theta)$  densidad de la distribución  $N(\bar{x}, \sigma^2/n)$  y  $m(\bar{x}|\pi)$  dada por (4.8).  $\square$

De manera que en el caso de contaminaciones unimodales y simétricas, también es posible aproximar las medidas de evidencia clásicas, mediante el  $p$ -valor, y bayesianas, mediante una cota inferior de la probabilidad final de la hipótesis nula puntual, siempre que la distribución inicial mixta sea elegida de acuerdo con nuestro procedimiento.

Debe observarse, por último que a medida que exigimos condiciones más restrictivas a la clase de contaminaciones, los valores de  $b$  deben ser más pequeños para que los contrastes puntual y de intervalo sean equivalentes.

# Bibliografía

- BERGER, J.O. (1985). *Statistical Decision Theory and Bayesian Analysis*. Springer Verlag. New York.
- BERGER, J.O. (1986). *Are  $p$ -values reasonable measures of accuracy*. In Pacific Statistical Congress (I. S. Francis et al. eds.). North Holland, Amsterdam.
- BERGER, J.O. (1994). *An overview over Robust Bayesian analysis* (with discussion). *Test*, **3**, 1, 5–124.
- BERGER, J.O. AND BERLINER, L.M. (1986), *Robust Bayes and empirical Bayes analysis with  $\varepsilon$ -contaminated priors*. *Ann. Statist.*, **14**, 2, 461–486.
- BERGER, J.O. AND DELAMPADY, M. (1987). *Testing Precise Hypotheses*. *Statistical Science*, **2**, **3**, 317–352.
- BERGER, J.O. AND MORTERA, J. (1994). *Robust Bayesian hypothesis testing in the presence of nuisance parameters*. *J. Statist. Planning and Infer.*, **40**, 357–373.
- BERGER, J.O. AND SELLKE, T. (1987). *Testing a Point Null Hypothesis: The Irreconcilability of  $P$ -values and Evidence* (with discussion). *J. Amer. Statist. Assoc.* **82**, 112–132.

- BERNARDO, J.M. (1980). *A Bayesian analysis of classical hypothesis testing* (with discussion). (J.M. Bernardo, M.H. De Groot, D.V. Lindley and A.F.M. Smith eds.). Valencia: University Press, 605–618.
- BOX, G.E.P. AND TIAO, G.C. (1992). *Bayesian Inference in Statistical Analysis*. Wiley, New York.
- 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**, 106–111.
- DELAMPADY, M. (1989a). *Lower bounds on Bayes factors for interval hypotheses*. *J. Amer. Statist. Assoc.*, **84**, 120–124.
- DELAMPADY, M. (1989b). *Lower bounds on Bayes factors for invariant testing situations*. *J. Multivariate Anal.*, **28**, 227–246.
- DELAMPADY, M. AND BERGER, J.O. (1990). *Lower bounds on posterior probabilities for multinomial and chi-squared tests*. *Ann. Statist.* **18**, 1295–1316.
- DEGROOT, M. H. (1973). *Doing What Comes Naturally: Interpreting a Tail Area as a Posterior Probability or as a Likelihood Ratio*. *J. Amer. Statist. Assoc.* **68**, 966–969.
- DEMPSTER, A.P. (1973). *The direct use of likelihood for significance testing*. In Proc. of the Conference on Foundational Questions in Statistical Inference (O. Barndorff-Nielsen et al., eds.) 335–352. Dept. Theoretical Statistics, Univ. Aarhus.
- DICKEY, J.M. (1971). *The weighted likelihood ratio, linear hypothesis on normal location parameters*. *Ann. Math. Statist.*, **42**, 204–305.
- DICKEY, J.M. (1973). *Scientific reporting*. *J. Roy. Statist. Soc. Ser. B*, **35**, 285–305.

- DICKEY, J.M. (1974). *Bayesian Alternatives to the F-Test and Least Squares Estimate in the Normal Linear Model*. In *Studies in Bayesian Econometrics and Statistics*, eds. S.E. Fienberg and A. Zellner, Amsterdam: North Holland, 515-554.
- DICKEY, J. M. (1977). *Is the Tail Area Useful as an Approximate Bayes Factor?* J. Amer. Statist. Assoc., **72**, 138-142.
- EDWARDS, W., LINDMANN, H., AND SAVAGE, L.J. (1963). *Bayesian Statistical Inference for Psychological Research*. Psychological Review, **70**, 193-242. Reprinted (1984) in: J.B. Kadane, ed., *Robustness of Bayesian Analysis*, North-Holland, Amsterdam, 1-62.
- FELLER, W. (1975). *Introducción a la Teoría de Probabilidades y sus aplicaciones*. Vol. I. Limusa, México.
- GÓMEZ-VILLEGAS, M. A. Y DE LA HORRA, J. (1984). *Aproximación de factores Bayes*. Cuad. Bioestadist. 2, 355-361.
- GÓMEZ-VILLEGAS, M. A. AND GÓMEZ SÁNCHEZ-MANZANO, E. (1992). *Bayes Factor in Testing Precise Hypotheses*. Commun. Statist. - Theory Meth., 21(6), 1707-1715.
- GÓMEZ-VILLEGAS, M.A. AND MAÍN, P. (1992). *The influence of prior and likelihood tail behaviour on the posterior distribution*. Bayesian Statistics, 4 (J.M. Bernardo et al. eds.), Oxford: University Press, 661-667.
- GÓMEZ-VILLEGAS, M.A. AND SANZ, L. (1994). *Reconciling Bayesian and frequentist evidence in the point null testing problem*. Technical Report, 2, Dpto. de Estadística e Invest. Operativa, Fac. de Matemáticas. Universidad Complutense de Madrid.
- GOOD, I.J. (1965). *The Estimation of Probabilities: An Essay on Modern Bayesian Methods*. M.I.T. Press. Cambridge, Mass.

- GOOD, I.J. (1967). *A Bayesian significance test for the multinomial distribution*. J. Roy. Statist. Soc. Ser. B, **29**, 399–341.
- GOOD, I.J. (1985). *Weight of evidence: A brief survey*. In Bayesian Statistics 2 (J.M. Bernardo, M.H. DeGroot, D.V. Lindley and A.F.M. Smith, eds.) 249–270. North-Holland. Amsterdam.
- GOOD, I.J. (1986). *A flexible Bayesian model for comparing two treatments*. J. Statist. Comput. Simulation, **26**, 301–305
- HUBER, P.J. (1973). *The use of Choquet capacities in statistics*. Bull. Internat. Statist. Inst., **45**, 181–191.
- JEFFREYS, H. (1957). *Scientific Inference*. Cambridge Univ. Press, Cambridge.
- JEFFREYS, H. (1961). *Theory of Probability*, 3rd ed. Oxford: University Press. London.
- KASS, R.E. AND RAFTERY, A.E. (1995). *Bayes factors (with discussion)*. J. Amer. Statist. Assoc., **90**, 773–795.
- LEE, P.M. (1989). *Bayesian Statistics: An introduction*. London, Charles Griffin.
- LEHMAN, E.L. (1959). *Testing Statistical Hypotheses*. New York, John Wiley and Sons Inc. (2nd ed., 1986).
- LINDLEY, D.V. (1957). *A Statistical Paradox*. Biometrika, **44**, 187–192.
- LINDLEY, D.V. (1961). *The use of prior probability distributions in statistical inference and decisions*. Proc. Fourth Berkeley Symp. Math. Statist. Probab., **1**, 453–468. Univ. California Press.
- LINDLEY, D.V. (1965). *An Introduction to Probability and Statistics from a Bayesian Viewpoint* **1, 2**. Cambridge Univ. Press. Cambridge.

- LINDLEY, D.V. (1977). *A problem in forensic science*. Biometrika, **64**, 207–213.
- LINDLEY, D.V. (1988). *Statistical Inference Concerning Hardy–Weinberg equilibrium*. Bayesian Statistics 3 (J.M. Bernardo et al. eds.) Oxford: University Press, 307–326.
- MAÍN, P. (1988). *Prior and posterior tail comparisons*. Bayesian Statistics 3 (J.M. Bernardo et al. eds.) Oxford: University Press, 669–675.
- MORENO, E. AND CANO, J.A. (1989). *Testing a point null hypothesis: asymptotic robust Bayesian analysis with respect to the priors given in a subsigma field*. Int. Statist. Rev., **57**, 221–232.
- MORENO, E. AND CANO, J.A. (1991). *Robust Bayesian analysis with  $\epsilon$ -contaminations partially known*. J. Roy. Statist. Soc. Ser B, **53**, 143–155.
- MORENO, E. AND PERICCHI, (1991). *Robust Bayesian analysis for  $\epsilon$ -contaminations with shape and quantile constraints*. Proc. of the Fifth International Symposium on Applied Stochastic Models and Data Analysis. World Scientific Publisher, 454–470.
- PRATT, J. W. (1965). *Bayesian Interpretation of Standard Inference Statements* (with discussion). J. Royal Statist. Soc., Ser. B, **27**, 169–203.
- SHAFER, G. (1982). *Lindley's paradox* (with discussion). J. Amer. Statist. Assoc., **77**, 325–351.
- SIVAGANESAN, S. (1988). *Range of the posterior measures for priors with arbitrary contaminations*. Commun. Statist. - Theory Methods, **17**, 1591–1612.
- SIVAGANESAN, S. (1991a). *Sensitivity of some posterior summaries when the prior is unimodal with specified quantiles*. Canadian Journal of Statistics, **19**, 57–65.

- SIVAGANESAN, S. (1991b). *Sensitivity of some standard Bayesian estimates to prior uncertainty: A comparison*. J. Statist. Plan. and Infer., **27**, 85–103.
- SIVAGANESAN, S. AND BERGER, J.O. (1989). *Ranges of posterior measures for priors with unimodal contaminations*. Ann. Statist, **17**, 2, 868–889.
- SMITH, C.A.B. (1965). *Personal probability and Statistical Analysis*. J. Roy. Statist. Soc. Ser. A, **128**, 469–499.
- SMITH, A.F.M. AND SPIEGELHALTER, B.J. (1980). *Bayes factors and choice criteria for linear models*. J. Roy. Statist. Soc. Ser. B, **42**, 213–220.
- SOAL, AND BATEMAN, (1954). *Modern Experiments in Telepathy*. Faber and Faber. London.
- ZELLNER, A. (1971). *An Introduction to Bayesian Inference in Econometrics*. New York: John Wiley.
- ZELLNER, A. (1984). *Posterior Odds Ratios for regression Hypotheses: General Considerations and Some Specific Results*. In Basic Issues in Econometrics, Chicago: University of Chicago Press, 275–305.
- ZELLNER, A. AND SIOW, A. (1980). *Posterior Odds Ratios for Selected Regression Hypotheses*. Bayesian Statistics, eds. J. M. Bernardo, M. H. DeGroot, D. V. Lindley and A. F. M. Smith, Valencia: University Press, pp. 586–603.