

UNIVERSIDAD DE JAÉN

Departamento de Estadística e I.O.

Aproximación Bayesiana a la metodología de la Verosimilitud Empírica

Memoria presentada para
optar al grado de doctor por
Raúl Amor Pulido.

Vº Bº del Director

Vº Bº del Tutor

Fdo. Dr. Carlos Sánchez González
Universidad de Granada

Fdo. Dr. Juan Carlos Ruiz Molina
Universidad de Jaén

Jaén, Diciembre de 2001.

Editor: Editorial de la Universidad de Granada
Autor: Raúl Amor Pulido
D.L.: GR 2010-2011
ISBN: 978-84-694-1434-7

Antes de comenzar el desarrollo del trabajo, quiero expresar mi agradecimiento a los doctores D. Agustín Hernández Bastida y D. Carlos Sánchez González, ambos del Departamento de Economía Aplicada de la Universidad de Granada por las propuestas, sugerencias y correcciones que han hecho posible la realización de esta memoria. Pero sobre todo quiero agradecerles la confianza que depositaron en mí al proponerme la realización de una Tesis Doctoral. Además, quiero agradecer al Departamento de Estadística e I.O. de la Universidad de Jaén la confianza depositada en mí como profesor y, en especial, a todos mis compañeros por las muestras de apoyo manifestadas. Finalmente, quiero agradecer el apoyo y comprensión recibido por mi familia y por mi futura esposa, Chelo. Sin el apoyo de todas estas personas no hubiera sido posible la realización de este trabajo.

Índice:

Prólogo.....	1
1. El Problema de la Verosimilitud.....	3
1.1. Introducción.....	3
1.2. Generalidades sobre el problema de un subparámetro.....	4
1.3. Verosimilitud Empírica.....	5
1.3.1. Verosimilitud Empírica generalizada.....	13
1.4. El concepto de Curva de verosimilitud más probable.....	16
1.5. Verosimilitud Modificada.....	17
1.6. Relación entre la verosimilitud empírica generalizada y la verosimilitud modificada.....	22
1.7. La función de verosimilitud elegida y sus propiedades.....	30
1.7.1. Dependencia de los datos.....	31
1.7.2. Continuidad.....	31
1.7.3. Posibilidad de utilización en el teorema de Bayes.....	33
1.7.4. Suficiencia.....	34
1.7.5. Función de verosimilitud en muestras con la misma media y varianza muestral.....	35

2. Distribución a priori y Robustez Bayesiana.....	37
2.1. Introducción.....	37
2.2. Aplicaciones previas y utilización de la distribución Uniforme como distribución a priori.....	38
2.2.1. Auditoría de cuentas.....	38
2.2.2. Ecología. Distribución a priori de máxima entropía.....	40
2.2.3. Utilización de la distribución Uniforme como distribución a priori.....	41
2.3. Método de la máxima verosimilitud tipo II para la elección de la distribución a priori.....	42
2.3.1. Consideración de clases paramétricas usuales.....	43
2.4. Robustez Bayesiana.....	45
2.5. Robustez Bayesiana Global para clases de distribuciones a priori.....	47
2.5.1. Clases de contaminación.....	48
2.5.2. Clases de Bandas de Medidas de Probabilidad.....	74
3. Inferencia estadística y análisis de robustez con una función de verosimilitud empírica.....	83
3.1. Introducción.....	83
3.2. Estimación puntual.....	84
3.3. Obtención de regiones de confianza.....	85
3.4. Contrastes de hipótesis.....	86
3.5. Análisis de Robustez con una función de verosimilitud empírica.....	88
3.5.1. Elección de la función de verosimilitud del problema.....	89
3.5.2. Elección de la distribución a priori del parámetro media.....	92
3.5.3. Obtención de conclusiones a posteriori.....	103

ÍNDICE

4. Conclusiones y líneas de investigación futura.....	109
Apéndice.....	113
A.1. Obtención de la verosimilitud del problema, elección de la distribución a priori y obtención de la distribución a posteriori.....	113
A.1.1. Introducción de datos y cálculos iniciales.....	113
A.1.2. Elección de la distribución a priori.....	115
A.1.3. Cálculo de la verosimilitud en los valores de theta calculados y la distribución a posteriori en esos puntos.....	116
A.1.4. Cálculo de medidas a posteriori.....	120
A.2. Robustez Bayesiana.....	122
A.2.1. Clase formada por todas las distribuciones.....	122
A.2.2. Clase formada por todas las distribuciones unimodales en θ_0 .	123
A.2.3. Clase formada por todas las distribuciones unimodales en θ_0 y simétricas.....	124
Bibliografía.....	127

Prólogo:

Esta memoria pretende realizar un estudio relativo a una característica de la población. Dicho estudio se centrará en una medida que pretende resumir el comportamiento de dicha característica de la población. Es decir, dada una característica de la población, que para nosotros será una variable aleatoria X , vamos a estudiar lo que sucede con un parámetro relativo a dicha variable aleatoria, que va a ser su media, aunque se puede generalizar a cualquier otro parámetro poblacional de dicha variable aleatoria.

Para dicho estudio, vamos a considerar el punto de vista bayesiano, pues consideramos que es el que mejor se ajusta al problema que pretendemos resolver. Por tanto, como en cualquier problema de tipo bayesiano, vamos a necesitar una verosimilitud dependiente de la muestra escogida y del parámetro estudiado, y una distribución a priori relativa al parámetro, para poderlas combinar y así obtener la distribución a posteriori del parámetro respecto de la muestra escogida. Una vez que tenemos la distribución a posteriori, podemos realizar inferencia sobre dicho parámetro y obtener conclusiones relativas a él.

Para ello, consideramos una muestra aleatoria procedente de dicha variable aleatoria.

El primer problema que se nos plantea es la elección de la función de verosimilitud para el problema, aspecto que abordaremos en el Capítulo 1 de la memoria. La función de verosimilitud que escogemos, que depende de la muestra seleccionada y del parámetro estudiado, es una generalización de la función de

PRÓLOGO

verosimilitud empírica estudiada en la literatura, que puede ser combinada mediante el teorema de Bayes con cualquier distribución a priori para obtener la distribución a posteriori correspondiente. Versiones para espacios paramétricos discretos de dicha modificación han sido aplicadas en varios campos (auditoría de cuentas, ecología), pero nosotros aportamos su aplicación a casos de espacios paramétricos continuos, puesto que demostramos que la función de verosimilitud empírica es continua en un espacio paramétrico continuo. Además vemos algunas propiedades interesantes de dicha función.

El segundo problema que se nos presenta en nuestro estudio es la elección de la distribución a priori del parámetro estudiado, que lo vamos a estudiar en el Capítulo 2 de la memoria. Por lo general, la distribución a priori va a depender de la información que posea el investigador sobre dicho parámetro. Si se conoce la distribución y sus parámetros, perfecto. Pero, por lo general, todo lo anterior es desconocido, y quizás se conozca alguna característica de la distribución. Entonces elegimos clases de distribuciones a priori a partir de la información del investigador y estudiamos su Robustez Bayesiana. Si la clase elegida es robusta, elegimos como distribución a priori una distribución singular de la clase, que suele ser la distribución de máxima verosimilitud tipo II de dicha clase. Este capítulo también muestra una revisión bastante amplia de resultados sobre análisis de robustez bayesiana.

Una vez elegidas convenientemente la función de verosimilitud y la distribución a priori, podemos combinarlas utilizando el teorema de Bayes, obteniendo la distribución a posteriori del parámetro respecto la muestra escogida, que era nuestro objetivo. A partir de este momento ya se pueden realizar estudios relativos al parámetro estudiado.

En el Capítulo 3 se pone de manifiesto la combinación de la función de verosimilitud empírica con una distribución a priori, para obtener una distribución a posteriori, a partir de la cual estudiamos diversos aspectos relativos a la inferencia bayesiana sobre el parámetro, aplicando la teoría bayesiana e incluiremos el desarrollo de una aplicación del modelo propuesto a lo largo de la memoria, con la utilización de un programa informático que se presenta en el Apéndice.

En el Capítulo 4 comentamos las conclusiones obtenidas en la memoria, así como las posibles líneas de investigación por las que podemos continuar trabajando.

Capítulo 1:

El problema de la Verosimilitud.

1.1. Introducción.

En este primer capítulo de la memoria vamos a estudiar la función de verosimilitud que utilizaremos en el estudio de la distribución a posteriori del parámetro utilizado.

Comenzaremos viendo lo que sucede en el caso del estudio de un subparámetro. Continuaremos mostrando dos antecedentes de la utilización de la función de verosimilitud que vamos a elegir posteriormente: la *función de verosimilitud empírica* y la *función de verosimilitud más probable*. Estudiaremos la verosimilitud modificada. Demostraremos que estas tres verosimilitudes que hemos nombrado anteriormente son equivalentes. Terminaremos justificando la elección de la función de verosimilitud con la que vamos a trabajar durante el resto de la memoria y estudiaremos algunas propiedades de dicha función, como son su continuidad, y la obtención de estadísticos suficientes.

1.2. Generalidades sobre el problema de un subparámetro.

Planteémonos el problema siguiente:

Dada una variable aleatoria X , queremos realizar inferencia sobre su media, para lo cual realizaremos un estudio bayesiano sobre dicho parámetro.

Para ello obtenemos una muestra de tamaño n para una variable aleatoria que puede tomar k valores distintos, cada uno de ellos con frecuencia muestral n_i ($i=1, \dots, k$). La función de verosimilitud que vamos a considerar depende de los parámetros p_1, \dots, p_k , donde p_i es la probabilidad del valor poblacional i -ésimo. La información disponible a priori se refiere solo a la media (μ), que es una función lineal de p_1, \dots, p_k , ya que

$$\mu = \sum_{i=1}^k p_i \cdot x_i .$$

Por tanto, estamos en un caso en el que se incluyen situaciones en las que el parámetro media de los datos tiene un fuerte significado intuitivo del que se tienen conocimientos relevantes y extramuestrales, aunque no coincidan con los parámetros de la función de verosimilitud.

Por consiguiente, tenemos una situación en la que es posible especificar un conjunto de categorías a las que puede pertenecer cada una de las observaciones muestrales. Esto nos va a permitir modelizar la verosimilitud mediante una distribución multinomial. Generalmente, el número de parámetros de la multinomial va ser elevado y no será fácil especificar una distribución a priori sobre dichos parámetros. Además, en nuestro caso disponemos de información a priori del parámetro media de la variable aleatoria, que hemos visto que es una combinación lineal de los parámetros de la multinomial. Este escenario se denomina *situación con información a priori parcial*, puesto que la información a priori es demasiado pobre para poderse combinar con la verosimilitud a través del teorema de Bayes.

Nuestra aproximación consistirá en realizar una modificación de la verosimilitud para hacerla compatible con la información a priori disponible, de manera que se pueda combinar con la distribución a priori para obtener la distribución a posteriori de la media de la variable aleatoria estudiada. Así, obtendremos conclusiones aprovechando toda la información disponible y que es relevante en el problema.

1.3. Verosimilitud Empírica.

Vamos a estudiar un antecedente de la utilización de una función similar a la que nosotros vamos a escoger como función de verosimilitud de nuestro problema. Este antecedente es la denominada Verosimilitud Empírica.

El método de la verosimilitud empírica para construir intervalos de confianza fue introducido por Owen en 1988, como alternativa a los métodos bootstrap de verosimilitud, y consiste en calcular el perfil de verosimilitud de una distribución general multinomial cuya masa de probabilidad se encuentra localizada en los datos puntuales. De esta manera, la forma de las regiones de confianza se fundamenta directamente en el conjunto de datos observados y las regiones tienden a concentrarse en lugares donde la densidad del parámetro estimado es la más grande.

Analizaremos el concepto de verosimilitud empírica y para ello vamos a contemplar el caso de la construcción de regiones de confianza para parámetros que se pueden expresar como funciones suaves de la media, aunque en nuestro caso, lo que nos interesa finalmente es estudiarla exclusivamente para la media de una variable unidimensional.

Vamos a suponer que el parámetro r -variante de interés es $\vec{\theta}_{r \times 1} = g(\mu)$, siendo $\mu_{s \times 1}$ la media de la población de la cual se obtuvo la muestra aleatoria s -variante de tamaño n , con valores $\bar{X}_{1 \times s}, \dots, \bar{X}_{n \times s}$, y sea:

$$g: \mathcal{R}^s \rightarrow \mathcal{R}^r$$

$$\bar{\mu}_{s \times 1} \rightarrow \vec{\theta}_{r \times 1}$$

una función suave de la media.

Para la definición formal de la verosimilitud empírica, consideramos:

Sea $\vec{\theta}$ alguna característica de una población (como la media), para la que queremos construir una región de confianza (no se requiere que $\vec{\theta}$ sea un escalar, si $\vec{\theta}$ es de dimensión 2, hablaríamos de regiones de confianza en el plano). Sea $\bar{X}_1, \dots, \bar{X}_n$ una muestra aleatoria, que puede ser de vectores, independiente y con distribución común F_0 , obtenida de la población. Sea $\vec{p} = (p_1, \dots, p_n)$ un vector tal que cada $p_i \geq 0$ y $\sum_{i=1}^n p_i = 1$. Sea $\vec{\theta}(\vec{p})$ el valor del parámetro $\vec{\theta}$ cuando la población

CAPÍTULO 1: EL PROBLEMA DE LA VEROSIMILITUD

es discreta con probabilidad p_i en el punto \bar{X}_i (por ejemplo, si $\bar{\theta}$ es la media poblacional, $\bar{\theta}(\bar{p}) = \sum_{i=1}^n p_i \bar{X}_i$).

Entonces, podemos definir la función de verosimilitud empírica para una función suave de la media como:

Definición 1.1. (Hall y La Scala (1990))

La función de verosimilitud empírica para $\bar{\theta}(\bar{p})$ evaluada en $\bar{\theta} = \bar{\theta}_1$ se define como:

$$L(\bar{\theta}_1) = \underset{\substack{\bar{p}: \bar{\theta}(\bar{p}) = \bar{\theta}_1 \\ \sum_{i=1}^n p_i = 1 \\ p_i \geq 0}}{\text{Max}} \prod_{i=1}^n p_i .$$

Por tanto, la verosimilitud empírica es claramente una verosimilitud de perfil multinomial, y el número de parámetros en la multinomial es $k-1$, debido a la condición $\sum_{i=1}^n p_i = 1$. El objetivo que buscamos es obtener los parámetros de la multinomial, p_1, \dots, p_n que maximizan esta función, para así tener el valor de función de verosimilitud empírica del parámetro $\bar{\theta}$ buscado, que es función suave de la media μ .

Sujeto sólo a las restricciones $\sum_{i=1}^n p_i = 1$ y $p_i \geq 0 \quad \forall i = 1, \dots, n$ el producto se maximiza cuando $p_i = n^{-1}$ para $1 \leq i \leq n$, es decir, cuando cada extracción de la muestra tiene la misma probabilidad de ser obtenida.

Para esta elección de \bar{p} , tenemos $\bar{\theta}(\bar{p}) = \hat{\theta}$, el llamado estimador *bootstrap*; donde, por ejemplo, cuando $\bar{\theta}$ es la media poblacional, $\bar{\theta}(n^{-1}, \dots, n^{-1}) = n^{-1} \sum_{i=1}^n \bar{X}_i$, es la media muestral. Este ejemplo nos ayudará a desarrollar un paralelismo entre la verosimilitud paramétrica clásica y la verosimilitud empírica. En este caso, tenemos que $L(\hat{\theta}) = n^{-n}$, por lo que definimos:

APROXIMACIÓN BAYESIANA A LA METODOLOGÍA DE LA VEROSIMILITUD EMPÍRICA

Definición 1.2. (Hall y La Scala (1990))

La razón de verosimilitud empírica para $\vec{\theta}(\vec{p})$ evaluada en $\vec{\theta} = \vec{\theta}_1$ como:

$$R(\vec{\theta}_1) = \frac{L(\vec{\theta}_1)}{L(\hat{\vec{\theta}})} = \underset{\vec{p}: \vec{\theta}(\vec{p})=\vec{\theta}_1}{\text{Max}} \prod_{i=1}^n n p_i \cdot \sum_{i=1}^n p_i = 1, p_i \geq 0$$

Definición 1.3. (Hall y La Scala (1990))

El logaritmo de la razón de verosimilitud empírica para $\vec{\theta}(\vec{p})$ evaluado en $\vec{\theta} = \vec{\theta}_1$ está dado por:

$$l_E(\vec{\theta}_1) = -2 \cdot \log(R(\vec{\theta}_1)) = -2 \cdot \log \frac{L(\vec{\theta}_1)}{L(\hat{\vec{\theta}})} = -2 \cdot \log \left(\underset{\vec{p}: \vec{\theta}(\vec{p})=\vec{\theta}_1}{\text{Max}} \prod_{i=1}^n n p_i \cdot \sum_{i=1}^n p_i = 1, p_i \geq 0 \right)$$

Estamos interesados en conjuntos de la forma $\mathcal{R}_c = \{ \vec{\theta} : l_E(\vec{\theta}) \leq c \}$ que puedan utilizarse como regiones de confianza para $\vec{\theta}_0$. Bajo estas condiciones, un contraste de la hipótesis nula $\vec{\theta}_0 = t$ se rechazará cuando $t \notin \mathcal{R}_c$, es decir, cuando no exista una distribución \vec{p} con $\vec{\theta}(\vec{p}) = t$ que verifique:

$$l_E(t) \leq c \Leftrightarrow -2 \cdot \log(R(t)) \leq c \Leftrightarrow \log(R(t)) \geq -c/2 \Leftrightarrow R(t) \geq e^{-c/2}.$$

Estudiamos ahora el desarrollo del problema de optimización de la función de verosimilitud empírica para la media, problema que sería análogo en el caso de una función suave de la media:

Lema 1.1. (de optimización de la función de verosimilitud empírica)

Los valores de p_i que maximizan la función de verosimilitud empírica de la media evaluada en $\vec{\mu} = \vec{\mu}_1$

$$L(\vec{\mu}_1) = \underset{\vec{p}: \vec{\mu}(\vec{p})=\vec{\mu}_1}{\text{Max}} \prod_{i=1}^n p_i \cdot \sum_{i=1}^n p_i = 1, p_i \geq 0$$

CAPÍTULO 1: EL PROBLEMA DE LA VEROSIMILITUD

son:

$$p_i = n^{-1} \left(1 + \vec{t}^T (\bar{X}_i - \bar{\mu}_1) \right)^{-1}, \quad \forall i = 1, \dots, n,$$

donde los valores de \vec{t} son vectores que dependen de $\bar{\mu}_1$ ($\vec{t} = \lambda(\bar{\mu}_1)$), y que satisfacen:

$$\left\{ \sum_{i=1}^n p_i (\bar{X}_i - \bar{\mu}_1) = \sum_{i=1}^n n^{-1} \left(1 + \vec{t}^T (\bar{X}_i - \bar{\mu}_1) \right)^{-1} (\bar{X}_i - \bar{\mu}_1) = \vec{0}, \right.$$

y, por tanto, tenemos que:

$$L_{\mu}(\bar{\mu}_1) = n^{-n} \prod_{i=1}^n \left(1 + \vec{t}^T (\bar{X}_i - \bar{\mu}_1) \right)^{-1} \text{ y}$$

$$l_{\mu}(\bar{\mu}_1) = 2 \sum_{i=1}^n \log \left(1 + \vec{t}^T (\bar{X}_i - \bar{\mu}_1) \right).$$

Demostración:

Incluimos la demostración, porque en ella vamos a ver cómo se resuelve el problema para cada valor de la media.

El desarrollo lo vamos a hacer mediante multiplicadores de Lagrange.

Teniendo en cuenta que la restricción:

$$\bar{\mu}(\vec{p}) = \sum_{i=1}^n p_i \bar{X}_i = \bar{\mu}_1 \text{ es equivalente a } \sum_{i=1}^n (\bar{X}_i - \bar{\mu}_1) p_i = \vec{0}_{r \times 1},$$

tenemos:

$$\begin{aligned} & \underset{p_i}{\text{Max}} \prod_{i=1}^n p_i \\ & \text{sujeto a } \left\{ \begin{array}{l} \sum_{i=1}^n p_i = 1 \\ \sum_{i=1}^n (\bar{X}_i - \bar{\mu}_1) p_i = \vec{0}_{r \times 1} \\ p_i \geq 0 \quad \forall i = 1, \dots, n \end{array} \right. \end{aligned}$$

que es equivalente, tomando logaritmos, a:

$$\underset{p_i}{\text{Max}} \sum_i n_i \log(p_i)$$

APROXIMACIÓN BAYESIANA A LA METODOLOGÍA DE LA VEROSIMILITUD EMPÍRICA

$$\text{sujeto a } \begin{cases} \sum_{i=1}^n p_i = 1 \\ \sum_{i=1}^n (\bar{X}_i - \bar{\mu}_1) p_i = \bar{0}_{r \times 1} \\ p_i \geq 0 \quad \forall i = 1, \dots, n \end{cases}$$

Construimos la función lagrangiana, tomando, para facilitar los cálculos, $n \cdot h$ y $n \cdot \bar{t}^T$ como multiplicadores de Lagrange:

$$\text{Max}_{p_i} \left[\sum_i \log(p_i) - n \cdot h \left(\sum_i p_i - 1 \right) - n \cdot \bar{t}^T \left(\sum_i (\bar{X}_i - \bar{\mu}_1) p_i \right) \right] = L(\bar{p}, \bar{t}).$$

Derivamos respecto de las distintas probabilidades y de los multiplicadores e igualamos a cero:

$$\begin{cases} \frac{\partial L(\bar{p}, \bar{t})}{\partial p_i} = \frac{1}{p_i} - n \cdot h - n \cdot \bar{t}^T (\bar{X}_i - \bar{\mu}_1) = 0, \quad \forall i = 1, \dots, n \\ \frac{\partial L(\bar{p}, \bar{t})}{\partial h} = 0 \Rightarrow \sum_i p_i = 1 \\ \frac{\partial L(\bar{p}, \bar{t})}{\partial \bar{t}} = 0 \Rightarrow \sum_i (\bar{X}_i - \bar{\mu}_1) p_i = 0 \end{cases}$$

De la primera ecuación tenemos:

$$\frac{1}{p_i} = n(h + \bar{t}^T (\bar{X}_i - \bar{\mu}_1)),$$

de donde obtenemos el valor de cada una de las probabilidades que maximizan la función de verosimilitud empírica:

$$p_i = n^{-1} (h + \bar{t}^T (\bar{X}_i - \bar{\mu}_1))^{-1}, \quad \forall i = 1, \dots, n.$$

Sustituyendo ahora este valor en las restricciones:

$$\begin{cases} \bar{0} = \sum_i (\bar{X}_i - \bar{\mu}_1) p_i = \sum_i (\bar{X}_i - \bar{\mu}_1) n^{-1} (h + \bar{t}^T (\bar{X}_i - \bar{\mu}_1))^{-1} = \sum_i (\bar{X}_i - \bar{\mu}_1) (h + \bar{t}^T (\bar{X}_i - \bar{\mu}_1))^{-1} \\ 1 = \sum_i p_i = \sum_i n^{-1} (h + \bar{t}^T (\bar{X}_i - \bar{\mu}_1))^{-1} \Rightarrow 1 = \sum_i n^{-1} (h + \bar{t}^T (\bar{X}_i - \bar{\mu}_1))^{-1} \end{cases}$$

Como $\frac{1}{p_i} = n(h + \bar{t}^T (\bar{X}_i - \bar{\mu}_1))$, tenemos que $\frac{1}{n} = (h + \bar{t}^T (\bar{X}_i - \bar{\mu}_1)) p_i$.

Si sumamos los k valores, tenemos:

$$1 = \sum_i (h + \bar{t}^T (\bar{X}_i - \bar{\mu}_1)) p_i = h \sum_i p_i + \bar{t}^T \sum_i (\bar{X}_i - \bar{\mu}_1) p_i = h.$$

Por tanto, tenemos que $h = 1$, por lo que $p_i = n^{-1} (1 + \bar{t}^T (\bar{X}_i - \bar{\mu}_1))^{-1}$, $\forall i = 1, \dots, n$.

CAPÍTULO 1: EL PROBLEMA DE LA VEROSIMILITUD

Entonces, los valores de \bar{t} son valores que dependen de $\bar{\mu}_1$ ($\bar{t} = \lambda(\bar{\mu}_1)$) y satisfacen:

$$\left\{ \sum_i p_i(\bar{X}_i - \bar{\mu}_1) = \sum_i n^{-1} (1 + \bar{t}^T (\bar{X}_i - \bar{\mu}_1))^{-1} (\bar{X}_i - \bar{\mu}_1) = 0, \right.$$

es decir:

$$\left\{ \sum_i p_i(\bar{X}_i - \bar{\mu}_1) = \sum_i (1 + \bar{t}^T (\bar{X}_i - \bar{\mu}_1))^{-1} (\bar{X}_i - \bar{\mu}_1) = 0. \right.$$

Y la función de verosimilitud empírica tomará el valor:

$$L_\mu(\bar{\mu}_1) = \prod_i p_i = \prod_i n^{-1} (1 + \bar{t}^T (\bar{X}_i - \bar{\mu}_1))^{-1} = n^{-n} \prod_i (1 + \bar{t}^T (\bar{X}_i - \bar{\mu}_1))^{-1}.$$

$$\begin{aligned} l_\mu(\bar{\mu}_1) &= -2 \cdot \log \left(\frac{L_\mu(\bar{\mu}_1)}{L_\mu(\bar{X})} \right) = -2 \cdot \log \left(\frac{n^{-n} \prod_i (1 + \bar{t}^T (\bar{X}_i - \bar{\mu}_1))^{-1}}{n^{-n}} \right) = \\ &= -2 \sum_i \log (1 + \bar{t}^T (\bar{X}_i - \bar{\mu}_1))^{-1} = 2 \sum_i \log (1 + \bar{t}^T (\bar{X}_i - \bar{\mu}_1)) \end{aligned}$$

con la restricción de

$$\sum_i n^{-1} (1 + \bar{t}^T (\bar{X}_i - \bar{\mu}_1))^{-1} (\bar{X}_i - \bar{\mu}_1) = 0,$$

que es el resultado que queríamos probar.

La Estadística utiliza funciones de la razón de verosimilitud paramétrica para construir regiones de confianza y realizar contrastes, aunque en algunas su obtención es complicada y más aún encontrar su distribución. Wilks, en 1.938, prueba que, bajo condiciones de regularidad, $-2 \cdot \log(R_0)$ sigue una distribución asintótica χ^2 , donde R_0 es el máximo de la función de razón de verosimilitud sujeto a ciertas hipótesis:

Ahora, $L(\theta)$ es la verosimilitud clásica de una muestra, $\hat{\theta}$ el estimador de máxima verosimilitud usual, y $l(\theta) = -2 \cdot \log\{L(\theta)/L(\hat{\theta})\}$ el logaritmo de la razón de verosimilitud; y θ_0 denota el verdadero valor de θ , suponemos que no hay parámetros de ruido, y que t es el rango de la matriz de varianzas asintóticas de $n^{1/2} \cdot \hat{\theta}$. Con estas hipótesis, el teorema de Wilks establece que, bajo condiciones de regularidad apropiadas, $l(\theta_0)$ tiene una distribución asintótica $\chi^2(t)$.

APROXIMACIÓN BAYESIANA A LA METODOLOGÍA DE LA VEROSIMILITUD EMPÍRICA

Este resultado es clave para construir las regiones de confianza basadas en la verosimilitud paramétrica, de la siguiente forma: Se encuentra en las tablas el valor de c tal que $P(\chi^2 \leq c) = 1-\alpha$, donde $1-\alpha$ es el nivel de confianza deseado para la región. Entonces, $\mathcal{R}_c = \{\theta | l(\theta) \leq c\}$ es la región apropiada. A la vista del teorema, la cobertura asintótica de \mathcal{R}_c es igual a $1-\alpha$:

$$P(\theta_0 \in \mathcal{R}_c) = P\{l(\theta_0) \leq c\} \rightarrow 1-\alpha \text{ con } n \rightarrow \infty.$$

Una propiedad importante de la verosimilitud empírica es que admite una versión no paramétrica del teorema de Wilks, por lo que la anterior forma de construir una región de confianza se va a emplear en el caso de la verosimilitud empírica.

Para enunciar un teorema análogo al de Wilks para la verosimilitud empírica en el caso de una función suave de la media, consideramos lo siguiente:

Sea $\bar{\mu}_0 = \bar{\mu}$, de dimensión s , el verdadero valor de la media poblacional, sea la función $g = (g^{(1)}, \dots, g^{(r)})^T: \mathcal{R}^s \rightarrow \mathcal{R}^r$ con derivadas continuas en un entorno de $\bar{\mu}_0$, sea $\bar{\theta}_0 = g(\bar{\mu}_0)$ el verdadero valor de un parámetro $\bar{\theta} = g(\bar{\mu})$; y consideremos la razón

de verosimilitud empírica $R(\bar{\theta}_1) = \frac{L(\bar{\theta}_1)}{L(\hat{\theta})} = \text{Max}_{p: \substack{\sum_{i=1}^n p_i = 1 \\ p_i \geq 0}} \prod_{i=1}^n n \cdot p_i$ y el logaritmo de la

razón de verosimilitud empírica $l_E(\bar{\theta}_1) = -2 \cdot \log\left(L(\bar{\theta}_1) / L(\hat{\theta})\right)$. Sea c un valor real,

tal que $P(\chi^2 \leq c) = 1-\alpha$, y sea $\mathcal{R}_c = \{\bar{\theta} : l_E(\bar{\theta}) \leq c\}$ la región de confianza del parámetro $\bar{\theta}$. Como g tiene derivadas continuas en un entorno de $\bar{\mu}_0$, entonces la

matriz de varianzas asintótica de $n^{1/2} \cdot \hat{\theta}$ es:

$$V = v_0 \Sigma v_0^T,$$

donde v_0 es la matriz $(r \times s)$ cuyos elementos vienen definidos por:

$$v_0^{(ij)} = \partial g^{(i)}(\bar{\mu}) / \partial \mu^{(j)} |_{\bar{\mu}=\bar{\mu}_0}$$

y $\Sigma = E\{(\bar{X} - \bar{\mu}_0)(\bar{X} - \bar{\mu}_0)^T\}$ es la matriz de varianzas poblacional.

Teorema 1.1. (análogo al de Wilks para la verosimilitud empírica)

(Hall y La Scala (1.990, pg.113))

Sea \vec{X} una variable aleatoria s -variante con varianza finita,
 g una función con derivadas continuas en un entorno de $\vec{\mu}_0$,
 $t \leq \text{Min}(r,s)$ el rango de V y
 l_E el logaritmo de la función de razón de verosimilitud empírica.
 Entonces,

$$l_E(\vec{\theta}_0) = -2 \cdot \log(R(\vec{\theta}_0)) = -2 \cdot \log \frac{L(\vec{\theta}_0)}{L(\hat{\vec{\theta}})} = -2 \cdot \log \left(\underset{\substack{\vec{p}: \\ \vec{\theta}(\vec{p}) = \vec{\theta}_0}}{\text{Max}} \prod_{i=1}^n n p_i \right)$$

$$\left(\begin{array}{c} \sum_{i=1}^n p_i = 1 \\ p_i \geq 0 \end{array} \right)$$

tiene una distribución asintótica $\chi^2(t)$.

Es decir, dado c un número positivo cualquiera, entonces el conjunto

$$\mathfrak{R}_c = \{ \vec{\theta} : l_E(\vec{\theta}) \leq c \}$$

verifica que:

$$\lim_{n \rightarrow \infty} P(\vec{\theta}_0 = g(\vec{\mu}_0) \in \mathfrak{R}_c) = \lim_{n \rightarrow \infty} P(l_E(\vec{\theta}_0) \leq c) = P(\chi_t^2 \leq c).$$

El caso en el que el parámetro estudiado $\vec{\theta}$ sea la media poblacional, que es el que nos interesa, la verosimilitud empírica adoptará una formulación sencilla y, además, cumple la propiedad de que las regiones de verosimilitud empírica para un valor medio dado (univariante o multivariante) son siempre convexas.

De acuerdo con el lema 1.1., el logaritmo de la razón de verosimilitud empírica para la media viene dado por:

$$l_E(\vec{\mu}) = -2 \sum_{i=1}^n \log(n p_i(\vec{\mu})) = 2 \sum_{i=1}^n \log(1 + \vec{t}^T (\vec{X}_i - \vec{\mu})).$$

A partir de aquí, Hall y La Scala (1.990, pg.114) prueban el siguiente teorema:

Teorema 1.2. (Hall y La Scala (1.990, pg.114))

- 1°. Una región de confianza de verosimilitud empírica para una media poblacional es siempre convexa.
- 2°. Una región de confianza de verosimilitud empírica para una función continua de una media poblacional es siempre conexa.
- 3°. Sin embargo, excepto en el caso donde la función es escalar, la región de confianza de verosimilitud empírica para una función continua de una media poblacional puede ser no convexa con probabilidad positiva.

1.3.1. Verosimilitud Empírica generalizada.

Vamos a considerar la siguiente generalización de la función de Verosimilitud Empírica:

Sea $\vec{\theta}$ alguna característica de una población (como la media), para la que queremos construir una región de confianza. Sea $(\vec{X}_1, \dots, \vec{X}_k)$ con frecuencias $n_i > 0$ y $\sum_{i=1}^k n_i = n$, una muestra aleatoria, que puede ser de vectores, independiente y con distribución común F_{θ} , obtenida de la población. Sea $\vec{p} = (p_1, \dots, p_k)$ un vector tal que cada $p_i \geq 0$ y $\sum_{i=1}^k p_i = 1$. Sea $\vec{\theta}(\vec{p})$ el valor del parámetro $\vec{\theta}$ cuando la población es discreta con probabilidad p_i en el punto \vec{X}_i (por ejemplo, si $\vec{\theta}$ es la media poblacional, $\vec{\theta}(\vec{p}) = \sum_{i=1}^k p_i \vec{X}_i$).

Entonces, podemos definir la verosimilitud empírica generalizada para una función suave de la media como:

Definición 1.4.

La función de verosimilitud empírica generalizada para $\vec{\theta}(\vec{p})$ evaluada en $\vec{\theta} = \vec{\theta}_1$ se define como:

$$L^*(\vec{\theta}_1) = \underset{\substack{\vec{p}: \\ \sum_{i=1}^k p_i = 1 \\ p_i \geq 0}}{\text{Max}} \prod_{i=1}^k p_i^{n_i} .$$

CAPÍTULO 1: EL PROBLEMA DE LA VEROSIMILITUD

Como podemos observar, la definición es similar a la de función de verosimilitud empírica, con la diferencia de que se pueden considerar valores muestrales repetidos (puede existir algún $n_i > 1$), de manera que si todas las frecuencias absolutas muestrales son iguales a uno $n_i = 1 \quad \forall i = 1, \dots, k$, tenemos que ambas definiciones son la misma. De nuevo, el objetivo que buscamos es obtener los parámetros de la multinomial, p_1, \dots, p_k que maximizan esta función, para así tener el valor de la función de verosimilitud empírica generalizada del parámetro $\vec{\theta}$ buscado, que es función suave de la media μ .

Sujeto solo a las restricciones $\sum_{i=1}^k p_i = 1$ y $p_i \geq 0 \quad \forall i = 1, \dots, k$ el producto se maximiza cuando $p_i = n_i \cdot n^{-1}$ para $1 \leq i \leq k$. Para esta elección de \vec{p} , tenemos $\vec{\theta}(\vec{p}) = \hat{\theta}$, el llamado estimador *bootstrap*; donde, por ejemplo, cuando $\vec{\theta}$ es la media poblacional, $\vec{\theta}(n_1 \cdot n^{-1}, \dots, n_k \cdot n^{-1}) = n^{-1} \sum_{i=1}^k n_i \bar{X}_i$, es, de nuevo, la media muestral. En este caso, tenemos que $L^*(\hat{\theta}) = n^{-n} \prod_{i=1}^k n_i^{n_i}$. Por consiguiente, podemos definir:

Definición 1.5.

La razón de la verosimilitud empírica generalizada para $\vec{\theta}(\vec{p})$ evaluada en $\vec{\theta} = \vec{\theta}_1$ está dada por:

$$R^*(\vec{\theta}_1) = \frac{L^*(\vec{\theta}_1)}{L^*(\hat{\theta})} = \frac{1}{\prod_{i=1}^k n_i^{n_i}} \underset{\substack{p: \\ \sum_{i=1}^k p_i = 1 \\ p_i \geq 0}}{\text{Max}} \prod_{i=1}^k n p_i^{n_i} .$$

Definición 1.6.

El logaritmo de la razón de verosimilitud empírica generalizada para $\vec{\theta}(\vec{p})$ evaluado en $\vec{\theta} = \vec{\theta}_1$ está dado por:

$$l_E^*(\bar{\theta}_1) = -2 \cdot \log(R^*(\bar{\theta}_1)) = -2 \cdot \log \frac{L^*(\bar{\theta}_1)}{L^*(\hat{\theta})} = -2 \cdot \log \left(\frac{1}{\prod_{i=1}^k n_i} \underset{\substack{p: \theta(p)=\bar{\theta}_1 \\ \sum_{i=1}^k p_i=1 \\ p_i \geq 0}}{\text{Max}} \prod_{i=1}^k n p_i^{n_i} \right).$$

Estudiamos ahora el desarrollo del problema de optimización de la función de verosimilitud empírica generalizada para la media, que sería análogo para una función suave de la media:

Lema 1.2. (de optimización de la verosimilitud empírica generalizada)

Los valores de p_i que maximizan la función de verosimilitud empírica generalizada de la media evaluada en $\bar{\mu} = \bar{\mu}_1$

$$L^*(\bar{\mu}_1) = \underset{\substack{p: \mu(p)=\bar{\mu}_1 \\ \sum_{i=1}^k p_i=1 \\ p_i \geq 0}}{\text{Max}} \prod_{i=1}^k p_i^{n_i}$$

son:

$$p_i = n_i \cdot n^{-1} \left(1 + \bar{t}^T (\bar{X}_i - \bar{\mu}_1) \right)^{-1}, \quad \forall i = 1, \dots, k,$$

donde los valores de \bar{t} son vectores que dependen de $\bar{\mu}_1$ ($\bar{t} = \lambda(\bar{\mu}_1)$), y que satisfacen:

$$\left\{ \sum_{i=1}^k p_i (\bar{X}_i - \bar{\mu}_1) = \sum_{i=1}^k n_i \cdot n^{-1} \left(1 + \bar{t}^T (\bar{X}_i - \bar{\mu}_1) \right)^{-1} (\bar{X}_i - \bar{\mu}_1) = \bar{0}, \right.$$

y, por tanto, tenemos que:

$$L_\mu^*(\bar{\mu}_1) = n^{-n} \prod_{i=1}^k \left(n_i \left(1 + \bar{t}^T (\bar{X}_i - \bar{\mu}_1) \right)^{-1} \right)^{n_i} \quad \text{y}$$

$$l_\mu^*(\bar{\mu}_1) = -2 \cdot \log \left(\frac{n^{-n} \prod_{i=1}^k \left(n_i \left(1 + \bar{t}^T (\bar{X}_i - \bar{\mu}_1) \right)^{-1} \right)^{n_i}}{n^{-n} \prod_{i=1}^k n_i^{n_i}} \right) =$$

$$\begin{aligned} &= -2 \sum_{i=1}^k n_i \log \left(n_i \left(1 + \bar{t}^T (\bar{X}_i - \bar{\mu}_1) \right)^{-1} \right) + 2 \sum_{i=1}^k n_i \log(n_i) = \\ &= -2 \sum_{i=1}^k n_i \cdot \log(n_i) + 2 \sum_{i=1}^k n_i \log \left(1 + \bar{t}^T (\bar{X}_i - \bar{\mu}_1) \right) + 2 \sum_{i=1}^k n_i \log(n_i) = \\ &= 2 \sum_{i=1}^k n_i \log \left(1 + \bar{t}^T (\bar{X}_i - \bar{\mu}_1) \right). \end{aligned}$$

Demostración:

Es totalmente análoga a la del lema 1.1. ya estudiado.

Puesto que las expresiones del logaritmo de la razón de verosimilitud empírica y del logaritmo de la razón de verosimilitud empírica generalizada para la media son iguales, podemos aplicar los teoremas 1.1. y 1.2. para la verosimilitud empírica generalizada para la media sin ningún problema, por lo que en esta extensión para la media, que es lo que realmente nos va a interesar, podemos hacer inferencia de la misma forma que en la función de verosimilitud empírica.

1.4. El concepto de Curva de verosimilitud más probable.

Vamos a estudiar otro antecedente de la utilización de una función similar a la que nosotros vamos a escoger como verosimilitud del problema. Este antecedente es la Curva de verosimilitud más probable.

Introducida y desarrollada en distintos trabajos por McCray (1984, 1997), fundamentalmente en el marco de auditoría de cuentas, define esta verosimilitud para realizar inferencia sobre el parámetro estudiado, a partir de una información muestral dada.

Procede del siguiente modo:

Dada una muestra aleatoria de tamaño n con k observaciones distintas (X_1, \dots, X_k) cada una de ellas con frecuencia absoluta n_i , donde $n_i > 0$ para $i=1, \dots, k$ y

$\sum_{i=1}^k n_i = n$; sea $\bar{p} = (p_1, \dots, p_k)$ con $p_i > 0$ para $i=1, \dots, k$ y $\sum_{i=1}^k p_i = 1$, el vector formado

por las proporciones poblacionales desconocidas para cada observación muestral;

APROXIMACIÓN BAYESIANA A LA METODOLOGÍA DE LA VEROSIMILITUD EMPÍRICA

sea $\bar{\theta}$ el vector de parámetros desconocidos de la población, que son función de la muestra y del vector \bar{p} .

Entonces, la función de verosimilitud, a la que nombra como función de verosimilitud más probable, para un vector de parámetros dado es:

$$L(\bar{p} | \bar{\theta}) = C \cdot \text{Max} \prod_{i=1}^k p_i^{n_i},$$

donde $\bar{\theta} = f(\bar{X}, \bar{p})$ y C es una constante normalizadora.

En nuestro caso, tenemos que el parámetro desconocido $\bar{\theta}$ será la media μ y la función $f(\bar{X}, \bar{p}) = \sum_{i=1}^k p_i \cdot x_i$.

A partir de esta verosimilitud se realizan contrastes de hipótesis sobre el parámetro, utilizando la razón de verosimilitudes entre la hipótesis nula y la alternativa; y se obtienen regiones de confianza con nivel de confianza $1-\alpha$ sobre el mismo parámetro, de manera que $P[L(p|\theta) \geq c] = 1-\alpha$.

Como se puede observar, en este caso la función de verosimilitud es exactamente la misma que en la verosimilitud empírica, pero la forma de realizar inferencia varía.

1.5. Verosimilitud Modificada.

Desarrollada por McCray, Hernández y Vázquez (1995), estudia una adaptación de una función de tipo multinomial para que pueda ser combinada con información a priori sobre el parámetro desconocido (en nuestro caso la media) para un espacio paramétrico discreto a través del teorema de Bayes.

Consideremos un experimento en el que el resultado de cada realización pertenece a una de las k categorías alternativas en que se ha dividido el espacio muestral. Sea $p=(p_1, \dots, p_k)$ un vector de probabilidades, es decir, $0 \leq p_i \leq 1$, para todo $i=1, \dots, k$ y $\sum_{i=1}^k p_i = 1$, donde cada p_i designa la probabilidad de que el resultado de una realización particular del experimento pertenezca a la categoría i -ésima.

Consideramos n realizaciones independientes del experimento, y sea (n_1, \dots, n_k) un vector aleatorio donde n_i indica el número de realizaciones en las que se presenta la

CAPÍTULO 1: EL PROBLEMA DE LA VEROSIMILITUD

categoría i -ésima y se verifica que $\sum_{i=1}^k n_i = n$. Entonces, la distribución del vector aleatorio (n_1, \dots, n_k) es una distribución multinomial de parámetros (p_1, \dots, p_k) , es decir:

$$P[n_1, \dots, n_k | n, p] = \frac{n!}{n_1! \dots n_k!} \prod_{i=1}^k p_i^{n_i},$$

donde $n_i = 0, 1, \dots, n$ y $\sum_{i=1}^k n_i = n$.

Designaremos a la distribución multinomial con los parámetros indicados por $M(p_1, \dots, p_k)$ o por $M(\vec{p})$.

Si tenemos un problema estadístico con esta función de verosimilitud y queremos realizar un análisis bayesiano, sería necesario disponer de información a priori sobre los parámetros de dicha función, es decir, (p_1, \dots, p_k) , para poder aplicar el teorema de Bayes. Pero este no es nuestro caso, puesto que nuestra información a priori se refiere a la media de la variable estudiada, que es función de los parámetros de la función de verosimilitud. Es decir, si notamos por Θ el espacio paramétrico de los parámetros (p_1, \dots, p_k) y por Λ el espacio paramétrico de la media μ de la variable estudiada, tenemos una función de la que sí disponemos de información:

$$\begin{aligned} \varphi: \Theta &\rightarrow \Lambda \\ (p_1, \dots, p_k) &\rightarrow \varphi(p_1, \dots, p_k) = \sum_{i=1}^k p_i \cdot x_i = \mu \end{aligned}$$

Esta función φ es inyectiva y medible, y el conjunto Λ es discreto.

Por tanto, disponemos de información a priori sobre los valores de μ , que notaremos como $\pi(\mu)$.

Pero, como ya hemos comentado anteriormente, el teorema de Bayes no permite la combinación de la verosimilitud $M(p_1, \dots, p_k)$ con la distribución a priori $\pi(\mu)$. Por tanto, no es posible, en principio, realizar un análisis bayesiano, ya que, entre otros problemas, la función de verosimilitud del problema no es integrable respecto de la medida de probabilidad a priori disponible.

A continuación vamos a modificar la verosimilitud original del problema para que la nueva verosimilitud modificada sea compatible con la distribución a priori disponible, en el sentido de poderle aplicar el teorema de Bayes con la información a priori disponible. Utilizamos, para ello, la siguiente notación:

APROXIMACIÓN BAYESIANA A LA METODOLOGÍA DE LA VEROSIMILITUD EMPÍRICA

Sea Θ un espacio métrico separable, \mathcal{A} el σ -álgebra natural sobre Θ y \mathcal{B} un sub- σ -álgebra de \mathcal{A} .

Sean $A_b^+(p)$ el conjunto de todas las funciones real-valuadas $f(p)$, que son no negativas, acotadas y \mathcal{A} -medibles; π una medida de probabilidad sobre (Θ, \mathcal{B}) ; $\int_{\Theta}^* f(p)\pi(dp)$, para $f(p) \in A_b^+(p)$ la integral superior de $f(p)$ respecto de π , e I_C la función indicadora del conjunto C .

El siguiente teorema nos da la verosimilitud modificada que nos interesa para restablecer el teorema de Bayes.

Teorema 1.3. (Cano, Hernández y Moreno (1987a,b,c))

Sea $f(p)$ perteneciente a $A_b^+(p)$. Entonces:

1º. Existe una función no negativa, finita y \mathcal{B} -medible $f_{\pi}^B(p)$ definida salvo una π -equivalencia por:

$$\int_F f_{\pi}^B(p)\pi(dp) = \int_F f(p)\pi(dp), \quad F \in \mathcal{B}$$

2º. Para cada $f(p)$ perteneciente a $A_b^+(p)$, existe una versión $\hat{f}_{\pi}^B(p)$ tal que:

$$\hat{f}_{\pi}^B(p) \geq f(p) \quad \text{y} \quad f_{\pi}^B(p) \geq f(p), \quad \text{casi seguro } [\pi]$$

3º. Algunas propiedades básicas de la transformación $f \rightarrow f_{\pi}^B(p)$ son las siguientes:

a) Si $f(p) \leq K$ para todo $p \in \Theta$, entonces $f_{\pi}^B(p) \leq K$, casi seguro $[\pi]$

b) $[a \cdot f(p)]_{\pi}^B = a \cdot f_{\pi}^B(p)$ casi seguro $[\pi]$, siendo a una constante.

c) Dadas dos funciones $f(p)$ y $g(p)$, entonces:

$$[f(p) + g(p)]_{\pi}^B \leq f_{\pi}^B(p) + g_{\pi}^B(p), \quad \text{casi seguro } [\pi]$$

d) Si las funciones $f(p)$ y $g(p)$ satisfacen $f(p) \geq g(p)$, entonces:

$$f_{\pi}^B(p) \geq g_{\pi}^B(p), \quad \text{casi seguro } [\pi]$$

e) $E_{\pi}[f_{\pi}^B(p)] = \int_{\Theta}^* f(p)\pi(dp)$

f) Si $f(p)$ es una función \mathcal{B} -medible, entonces $f_{\pi}^B(p) = f(p)$, casi seguro $[\pi]$

CAPÍTULO 1: EL PROBLEMA DE LA VEROSIMILITUD

4º. Si B está generado por una partición numerable $\{B_n\}$, entonces, para cada $f(p)$ perteneciente a $A_b^+(p)$, la función de salto definida por

$$f^B(p) = \sum_n \sup_{p \in B_n} f(p) \cdot I_{B_n}(p)$$

es una versión común de $\{f_\pi^B(p) : \pi \in H\}$, donde H es la clase de todas las medidas de probabilidad sobre B.

Demostración:

Este teorema es una versión condensada y simplificada, a nuestros intereses en esta memoria, de resultados que aparecen en Cano, Hernández y Moreno (1987a, b y c).

La función $f_\pi^B(p)$ del Teorema 1.3. es la verosimilitud modificada que nos interesaba. El uso de esta función para definir una distribución a posteriori bajo información a priori parcial puede verse en Cano, Hernández y Moreno(1986).

Con la terminología del Teorema 1.3., podemos reescribir nuestro problema de la siguiente forma:

Disponemos de una función de verosimilitud $M(p_1, \dots, p_k)$ que será A-medible, donde A es el σ -álgebra usual de Borel.

Disponemos de una información a priori dada por una medida de probabilidad sobre Λ con el σ -álgebra usual.

Como Λ es discreto, los conjuntos formados por un punto $\{\mu\}$ son átomos de su σ -álgebra, por lo que conjuntos de la forma:

$$\varphi^{-1}(\{\mu\}) = \{(p_1, \dots, p_k) \in \Theta : \varphi(p_1, \dots, p_k) = \mu\}$$

pertenecen al σ -álgebra A.

Sea B el sub- σ -álgebra de A inducido por φ sobre Θ ; si definimos la probabilidad de un conjunto de la forma $\varphi^{-1}(\{\mu\})$ por la probabilidad de $\{\mu\}$, tendremos una medida de probabilidad, que seguimos notando por π , sobre el sub- σ -álgebra B.

El problema es que la verosimilitud $M(p_1, \dots, p_k)$ y la medida de probabilidad π no son compatibles para utilizar el teorema de Bayes. No son compatibles porque $M(p_1, \dots, p_k)$ es A-medible y π está definida sobre B, donde B es un sub- σ -álgebra estricto de A y $M(p_1, \dots, p_k)$ no es B-medible. El teorema 1.3. aporta una solución a este problema transformando la verosimilitud.

APROXIMACIÓN BAYESIANA A LA METODOLOGÍA DE LA VEROSIMILITUD EMPÍRICA

En este caso el sub- σ -álgebra está generado por una partición numerable, por lo que la verosimilitud modificada admite una expresión explícita de la forma:

$$M^B(p_1, \dots, p_k) = \sum_{\mu} \sup_{(p_1, \dots, p_k) \in \varphi^{-1}(\{\mu\})} M(p_1, \dots, p_k) \cdot I_{\varphi^{-1}(\{\mu\})},$$

observando que esta función es constante sobre cada conjunto del tipo $\varphi^{-1}(\{\mu\})$, se puede escribir como

$$M^B(\mu) = \sum_{\mu} \sup_{(p_1, \dots, p_k) \in \varphi^{-1}(\{\mu\})} M(p_1, \dots, p_k) \cdot I_{\varphi^{-1}(\{\mu\})}$$

La verosimilitud modificada es una función B-medible, por lo que es compatible con la distribución a priori π . Por consiguiente, podemos utilizar el teorema de Bayes para establecer la distribución a posteriori

$$\pi(\cdot | \text{datos}) = \frac{M^B(\mu) \cdot \pi(\mu)}{\sum_{\Lambda} M^B(\mu) \cdot \pi(\mu)}.$$

Por tanto, en nuestro caso, la verosimilitud modificada va a ser:

$$M^B(\mu) = \underset{\substack{\vec{p}: \\ p_i \geq 0 \\ \sum_{i=1}^k p_i = 1 \\ \sum_{i=1}^k p_i \cdot x_i = \mu}}{\text{Max}} \prod_{i=1}^k p_i^{n_i},$$

y, por definición, consideramos el valor 0^0 como 1, es decir:

$$M^B(\mu) = \underset{\substack{\vec{p}: \\ p_i \geq 0 \\ \sum_{i=1}^k p_i = 1 \\ \sum_{i=1}^k p_i \cdot x_i = \mu}}{\text{Max}} \prod_{i \in I} p_i^{n_i},$$

donde I es el conjunto de valores que toma la muestra, es decir, aquellos valores con $n_i > 0$.

1.6. Relación entre la verosimilitud empírica generalizada y la verosimilitud modificada.

Veamos la relación existente entre la función de verosimilitud empírica generalizada y la verosimilitud modificada para la media de una variable aleatoria:

Para ello, en primer lugar, vamos a ver que los dos problemas siguientes son equivalentes:

Dada una muestra de tamaño n procedente de una variable aleatoria con espacio muestral dividido en k clases de una partición, en la cual se obtienen r valores muestrales distintos, cada uno con frecuencia n_i , de manera que

$$\sum_{i=1}^r n_i = n.$$

- **Problema número 1:**

La verosimilitud modificada:

$$\text{Max}_{\{p_i; i \in I\}} \prod_{i \in I} p_i^{n_i}$$

sujeto a:

$$\left\{ \begin{array}{l} \sum_{i \in I} p_i + \sum_{j \in \bar{I}} p_j = 1 \\ \sum_{i \in I} x_i p_i + \sum_{j \in \bar{I}} x_j p_j = \mu \\ 0 \leq p_i \leq 1; \quad i \in I \cup \bar{I} \end{array} \right\}, \text{ donde } \left\{ \begin{array}{l} I \cup \bar{I} = \{1, 2, \dots, k\} \\ I = \{1, 2, \dots, r\}; \quad r < k \end{array} \right\},$$

con I el conjunto de valores observados en la muestra y \bar{I} el resto de posibles valores que puede tomar la variable aleatoria estudiada.

Suponemos todos los x_i positivos (si no lo fueran haríamos un cambio de variable para conseguirlo).

- **Problema número 2:**

$$\text{Max}_{\{p_i; i \in I\}} \prod_{i \in I} p_i^{n_i}$$

sujeto a:

$$\left\{ \begin{array}{l} \sum_{i \in I} p_i \leq 1 \\ \sum_{i \in I} x_i p_i \leq \mu \\ 0 \leq p_i \leq 1; \quad i \in I \end{array} \right\}.$$

APROXIMACIÓN BAYESIANA A LA METODOLOGÍA DE LA VEROSIMILITUD EMPÍRICA

Para realizar la demostración vamos a resolver el primer problema, viendo que cumple las condiciones de Kuhn-Tucker:

Las condiciones de Kuhn-Tucker para el problema:

$Min f(x)$

sujeto a:

$$\left\{ \begin{array}{l} g_1(x) \leq 0 \\ \vdots \\ g_k(x) \leq 0 \\ g_{k+1}(x) = 0 \\ \vdots \\ g_m(x) = 0 \end{array} \right\} \text{son:}$$

1. $\nabla f(x) + \sum_{i=1}^m \lambda_i \nabla g_i(x) = \vec{0}$.

2. Ortogonalidad:

$$\lambda_1 g_1(x) = 0$$

\vdots

$$\lambda_k g_k(x) = 0$$

3. $\lambda_1, \dots, \lambda_k \geq 0$

4. $g_i(x) \leq 0 \quad i = 1, \dots, k.$

5. $g_j(x) = 0 \quad j = k + 1, \dots, m$

Para el caso de maximizar solo cambia que los $\lambda_i \leq 0$.

En nuestro caso, tenemos, tras tomar logaritmos:

$$Max_{\{p_i, i \in I\}} \sum_{i=1}^r n_i \cdot \log(p_i)$$

sujeto a:

$$\left\{ \begin{array}{l} \sum_{i \in I \cup \bar{I}} p_i = 1 \\ \sum_{i \in I \cup \bar{I}} x_i p_i = \mu \\ 0 \leq p_i \quad (-p_i \leq 0) \end{array} \right\}$$

Sea $\vec{p} = (p_1, p_2, \dots, p_k)$

Entonces, como:

CAPÍTULO 1: EL PROBLEMA DE LA VEROSIMILITUD

$$g_i(\vec{p}) = -p_i \quad i = 1, \dots, k$$

$$g_{k+1}(\vec{p}) = \sum_{i=1}^k p_i - 1 = 0$$

$$g_{k+2}(\vec{p}) = \sum_{i=1}^k p_i x_i - \mu = 0$$

tenemos:

$$\nabla f(\vec{p}) = \left(\frac{n_1}{p_1}, \dots, \frac{n_r}{p_r}, 0, \dots, 0 \right)$$

$$\nabla g_i(\vec{p}) = -e_i = (0, \dots, 0, -1^{(i)}, 0, \dots, 0)$$

$$\nabla g_{k+1}(\vec{p}) = (1, \dots, 1)$$

$$\nabla g_{k+2}(\vec{p}) = (x_1, \dots, x_k).$$

Entonces, las condiciones de Kuhn-Tucker son:

$$L(x, p) = \sum_{i=1}^r n_i \log(p_i) + \alpha_1 \left(\sum_{i=1}^k p_i - 1 \right) + \alpha_2 \left(\sum_{i=1}^k p_i \cdot x_i - \mu \right) - \sum_{i=1}^k \lambda_i \cdot p_i$$

$$1. \quad \frac{n_i}{p_i} - \lambda_i + \alpha_1 + x_i \alpha_2 = 0; \quad i = 1, \dots, r$$

$$- \lambda_j + \alpha_1 + x_j \alpha_2 = 0; \quad j = r+1, \dots, k$$

$$2. \quad \lambda_i p_i = 0; \quad i = 1, \dots, k$$

$$3. \quad \lambda_j \leq 0; \quad j = 1, \dots, k.$$

$$4. \quad -p_i \leq 0; \quad i = 1, \dots, k$$

$$5. \quad \sum_{i=1}^k p_i = 1$$

$$\sum_{i=1}^k x_i p_i = \mu$$

En el óptimo tiene que ocurrir que $p_i \neq 0$ para $i = 1, \dots, r$, para evitar problemas con el valor de la función objetivo ($\prod p_i = 0$), por lo que:

$$\lambda_i = 0 \quad \text{para } i = 1, \dots, r.$$

Veamos qué ocurre con $\lambda_{r+1}, \dots, \lambda_k$.

Supongamos que dos de ellos son iguales a cero:

$$\lambda_{j_1} = \lambda_{j_2} = 0$$

APROXIMACIÓN BAYESIANA A LA METODOLOGÍA DE LA VEROSIMILITUD EMPÍRICA

Entonces:

$$\alpha_1 + x_{j_1} \alpha_2 = 0 \Rightarrow -\frac{\alpha_1}{\alpha_2} = x_{j_1}$$

$$\alpha_1 + x_{j_2} \alpha_2 = 0 \Rightarrow -\frac{\alpha_1}{\alpha_2} = x_{j_2},$$

lo cual es imposible, ya que $x_{j_1} \neq x_{j_2}$;

por lo que, a lo sumo, existe un único λ_j nulo, por lo que, debido a $\lambda_i p_i = 0$ para $i = r+1, \dots, k$, existirá, a lo sumo, un único p_i no nulo para $i = r+1, \dots, k$.

Sea p_j el único no nulo para $j = r+1, \dots, k$.

Por tanto, tenemos la siguiente transformación del primer problema:

$$\text{Max}_{\{p_i: i \in I\}} \sum_{i=1}^r n_i \cdot \log(p_i)$$

sujeto a:

$$\left\{ \begin{array}{l} p_j + \sum_{i=1}^r p_i = 1 \\ p_j x_j + \sum_{i=1}^r p_i x_i = \mu \\ -p_i, -p_j \leq 0; i = 1, \dots, r \end{array} \right\},$$

que es un problema de maximización con función objetivo cóncava y restricciones convexas, por lo que un punto de Kuhn-Tucker será un punto óptimo.

Veamos ahora la equivalencia de los problemas primero y segundo:

Si el valor λ_j es no nulo, tenemos la equivalencia, ya que p_j sería nulo.

Si el valor λ_j es nulo:

Sea (p_1, \dots, p_r, p_j) una solución del problema primero. Veamos que (p_1, \dots, p_r) es solución del problema segundo:

CAPÍTULO 1: EL PROBLEMA DE LA VEROSIMILITUD

Supongamos que no lo es. Entonces, existe (p'_1, \dots, p'_r) solución del problema segundo, tal que:

$$\sum_{i=1}^r p'_i \leq 1$$

$$\sum_{i=1}^r p'_i x_i \leq \mu$$

$$\text{y además } \prod_{i=1}^r p_i^{n_i} > \prod_{i=1}^r p'_i^{n_i} .$$

Sea p''_j la solución del sistema

$$p''_j = 1 - \sum_{i=1}^r p'_i$$

$$p''_j x_j = \mu - \sum_{i=1}^r p'_i x_i .$$

Entonces, el vector $(p'_1, \dots, p'_r, p''_j)$ contradice que (p_1, \dots, p_r, p_j) sea solución del problema de verosimilitud modificada; por lo que tenemos que (p_1, \dots, p_r) es solución del problema segundo.

Sea (p_1, \dots, p_r) una solución del problema segundo. Veamos que (p_1, \dots, p_r, p_j) es solución del problema primero (donde $p_j = 1 - \sum_{i=1}^r p_i$):

Supongamos que no lo es. Entonces, existe $(p'_1, \dots, p'_r, p'_j)$ solución del problema primero, tal que:

$$\sum_{i=1}^r p'_i + p'_j = 1$$

$$\sum_{i=1}^r p'_i x_i + p'_j x_j = \mu$$

$$\text{y además } \prod_{i=1}^r p_i^{n_i} > \prod_{i=1}^r p'_i^{n_i} .$$

Entonces, el vector (p_1, \dots, p_r) no es solución del problema segundo; por lo que tenemos que (p_1, \dots, p_r, p_j) es solución del problema primero.

APROXIMACIÓN BAYESIANA A LA METODOLOGÍA DE LA VEROSIMILITUD EMPÍRICA

Una vez que hemos visto que ambos problemas son equivalentes, vamos a ver que la solución del segundo problema se obtiene siempre cuando las restricciones se saturan. Para ello, vamos a comenzar a resolver el problema:

Problema número 2:

$$\text{Max}_{\{p_i, i \in I\}} \prod_{i \in I} p_i^{n_i}$$

sujeto a:

$$\left\{ \begin{array}{l} \sum_{i \in I} p_i \leq 1 \\ \sum_{i \in I} x_i p_i \leq \mu \\ 0 \leq p_i \leq 1; \quad i \in I \end{array} \right\}.$$

En nuestro caso, tenemos, tras tomar logaritmos:

$$\text{Max}_{\{p_i, i \in I\}} \sum_{i=1}^r n_i \cdot \log(p_i)$$

sujeto a:

$$\left\{ \begin{array}{l} \sum_{i \in I} p_i \leq 1 \\ \sum_{i \in I} x_i p_i \leq \mu \\ 0 \leq p_i \quad (-p_i \leq 0) \end{array} \right\}$$

Sea $\vec{p} = (p_1, p_2, \dots, p_r)$

Entonces, como:

$$g_i(\vec{p}) = -p_i \quad i = 1, \dots, r$$

$$g_{r+1}(\vec{p}) = \sum_{i=1}^r p_i - 1 \leq 0$$

$$g_{r+2}(\vec{p}) = \sum_{i=1}^r p_i x_i - \mu \leq 0$$

tenemos:

$$\nabla f(\vec{p}) = \left(\frac{n_1}{p_1}, \dots, \frac{n_r}{p_r} \right)$$

$$\nabla g_i(\vec{p}) = -e_i = (0, \dots, 0, -1^{(i)}, 0, \dots, 0)$$

$$\nabla g_{r+1}(\vec{p}) = (1, \dots, 1)$$

$$\nabla g_{r+2}(\vec{p}) = (x_1, \dots, x_r).$$

CAPÍTULO 1: EL PROBLEMA DE LA VEROSIMILITUD

Entonces, las condiciones de Kuhn-Tucker son:

$$1. \quad \frac{n_i}{p_i} - \lambda_i + \alpha_1 + x_i \alpha_2 = 0; \quad i = 1, \dots, r$$

$$2. \quad \lambda_i p_i = 0; \quad i = 1, \dots, r$$

$$\alpha_1 \left(\sum_{i=1}^r p_i - 1 \right) = 0$$

$$\alpha_2 \left(\sum_{i=1}^r x_i p_i - \mu \right) = 0$$

$$3. \quad \lambda_j \leq 0; \quad j = 1, \dots, k.$$

$$\alpha_1, \alpha_2 \leq 0$$

$$4. \quad -p_i \leq 0; \quad i = 1, \dots, k$$

$$5. \quad \sum_{i=1}^r p_i - 1 \leq 0$$

$$\sum_{i=1}^r x_i p_i - \mu \leq 0$$

En el óptimo tiene que ocurrir que $p_i \neq 0$ para $i = 1, \dots, r$, para evitar problemas con el valor de la función objetivo ($\prod p_i = 0$), por lo que:

$$\lambda_i = 0 \quad \text{para } i = 1, \dots, r.$$

Para ver que la solución se obtiene cuando las restricciones se saturan, consideramos los siguientes casos:

- Supongamos que en ninguna de las restricciones se cumple la igualdad:

$$\sum_{i=1}^r p_i < 1 \quad , \quad \sum_{i=1}^r x_i p_i < \mu .$$

Entonces, por la condición 2. de Kuhn-Tucker, tenemos que $\alpha_1 = \alpha_2 = 0$.

Por tanto, por la condición 1., tenemos que $\frac{n_i}{p_i} = 0$ para todo valor $i = 1, \dots, r$,

lo cual es imposible.

Por consiguiente, en alguna restricción debe cumplirse la igualdad.

- Supongamos que se cumple la igualdad en la primera restricción y no en la segunda:

$$\sum_{i=1}^r p_i = 1 \quad , \quad \sum_{i=1}^r x_i p_i < \mu .$$

APROXIMACIÓN BAYESIANA A LA METODOLOGÍA DE LA VEROSIMILITUD EMPÍRICA

Entonces, en el caso equivalente de la verosimilitud modificada tenemos:

$$\begin{cases} \sum_{i=1}^r p_i + \sum_{i=r+1}^k p_i = 1 \\ \sum_{i=1}^r p_i x_i + \sum_{i=r+1}^k p_i x_i = \mu \end{cases} \Rightarrow \sum_{i=r+1}^k p_i = 0 \Rightarrow \sum_{i=r+1}^k p_i x_i = 0 \Rightarrow$$

$$\Rightarrow \begin{cases} \sum_{i=1}^r p_i + \sum_{i=r+1}^k p_i = 1 \\ \sum_{i=1}^r p_i x_i + \sum_{i=r+1}^k p_i x_i < \mu \end{cases},$$

lo cual es contradictorio, ya que la solución del problema inicial no lo es del problema de verosimilitud modificada.

- Supongamos ahora que se cumple la igualdad en la segunda restricción y no en la primera:

$$\sum_{i=1}^r p_i < 1, \quad \sum_{i=1}^r x_i p_i = \mu.$$

Entonces, en el caso equivalente de la verosimilitud modificada tenemos:

$$\begin{cases} \sum_{i=1}^r p_i + \sum_{i=r+1}^k p_i = 1 \\ \sum_{i=1}^r p_i x_i + \sum_{i=r+1}^k p_i x_i = \mu \end{cases} \Rightarrow \sum_{i=r+1}^k p_i x_i = 0 \quad (x_i > 0) \Rightarrow p_i = 0 \quad i=r+1, \dots, k \Rightarrow$$

$$\Rightarrow \begin{cases} \sum_{i=1}^r p_i + \sum_{i=r+1}^k p_i < 1 \\ \sum_{i=1}^r p_i x_i + \sum_{i=r+1}^k p_i x_i = \mu \end{cases},$$

lo cual es contradictorio, ya que la solución del problema inicial no lo es del problema de verosimilitud modificada.

Por tanto, la solución de este problema se alcanza cuando en las restricciones se saturan.

Por tanto, hemos demostrado el siguiente teorema:

Teorema 1.4.

El problema:

$$\text{Max} \prod_{\{p_i: i \in I\}} p_i^{n_i}$$

sujeto a:

$$\left\{ \begin{array}{l} \sum_{i \in I} p_i \leq 1 \\ \sum_{i \in I} x_i p_i \leq \mu \\ 0 \leq p_i \leq 1; \quad i \in I \end{array} \right\}$$

alcanza la solución cuando en las restricciones se tiene la igualdad, es decir, que este problema tiene la misma solución que el problema de la verosimilitud empírica generalizada:

$$\text{Max}_{\{p_i; i \in I\}} \prod_{i \in I} p_i^{n_i}$$

sujeto a:

$$\left\{ \begin{array}{l} \sum_{i \in I} p_i = 1 \\ \sum_{i \in I} x_i p_i = \mu \\ 0 \leq p_i \leq 1; \quad i \in I \end{array} \right\},$$

siempre que μ esté dentro del intervalo de valores de la muestra, es decir, siempre que el problema de la verosimilitud empírica generalizada tenga solución.

Por tanto, hemos visto que los problemas de la verosimilitud empírica generalizada y la verosimilitud modificada son equivalentes.

1.7. La función de verosimilitud elegida y sus propiedades.

Una vez estudiado todo lo anterior estamos en condiciones de elegir la función de verosimilitud del problema que estamos estudiando. Resumiendo:

Dada una muestra aleatoria de tamaño n con k observaciones distintas (X_1, \dots, X_k) cada una de ellas con frecuencia absoluta n_i , donde $n_i > 0$ para $i=1, \dots, k$ y

$\sum_{i=1}^k n_i = n$; sea $\vec{p} = (p_1, \dots, p_k)$ con $p_i > 0$ para $i=1, \dots, k$ y $\sum_{i=1}^k p_i = 1$, el vector formado

por las proporciones poblacionales desconocidas para cada observación muestral; y sea μ la media de la variable aleatoria, que es el parámetro desconocido de la

APROXIMACIÓN BAYESIANA A LA METODOLOGÍA DE LA VEROSIMILITUD EMPÍRICA

población, que es función de la muestra y del vector \bar{p} , y sobre el que vamos a realizar la inferencia.

La función de verosimilitud elegida para realizar el estudio a posteriori de la media de la variable aleatoria es:

$$f(x | \mu) = f^B(\mu) = \begin{cases} \underset{\bar{p}=(p_1, \dots, p_k)}{\text{Max}} \prod_{i=1}^k p_i^{n_i} \\ \text{s.a.} \left\{ \begin{array}{l} \sum_{i=1}^k p_i = 1 \\ \sum_{i=1}^k p_i \cdot x_i = \mu \\ p_i > 0 \quad i = 1, \dots, k \end{array} \right. & \mu \in (\text{Min}(x_i), \text{Max}(x_i)) \\ 0 & \mu \notin (\text{Min}(x_i), \text{Max}(x_i)) \end{cases}$$

Estudiamos ahora algunas de sus propiedades más importantes:

1.7.1. Dependencia de los datos.

Esta función de verosimilitud tiene una ventaja fundamental en su uso, y es que esta verosimilitud es una función que depende exclusivamente del conjunto de datos seleccionado y de la definición del parámetro elegido, no siendo necesarios supuestos adicionales acerca de la distribución de la variable aleatoria estudiada.

1.7.2. Continuidad.

Vamos a ver que la función de verosimilitud elegida es continua respecto de μ , lo cual va a ser de bastante utilidad para el posterior estudio de la robustez bayesiana:

Teorema 1.5.

La función de verosimilitud elegida es continua respecto de μ en la recta real \mathfrak{R} .

Demostración:

Sean $x_1, \dots, x_k \in \mathfrak{R}$ distintos entre sí y $n_1, \dots, n_k \in \mathfrak{N} - \{0\}$.

Consideremos la aplicación:

$$f : \Lambda \rightarrow \mathfrak{R} \\ f(\mu) = \text{Max} \phi(D_\mu), \text{ donde:}$$

CAPÍTULO 1: EL PROBLEMA DE LA VEROSIMILITUD

ϕ es la aplicación continua

$$\phi : \mathfrak{R}^k \rightarrow \mathfrak{R}$$

$$p = (p_1, \dots, p_k) \rightarrow \phi(p) = \prod_{i=1}^k p_i^{n_i}$$

$$\Lambda = \left(\text{Min}_i(x_i), \text{Max}_i(x_i) \right) \neq \emptyset.$$

$$D_\mu = \left\{ p = (p_1, \dots, p_k) \in \mathfrak{R}^k : \sum_{i=1}^k p_i = 1; \sum_{i=1}^k p_i \cdot x_i = \mu; p_1, \dots, p_k \geq 0 \right\}.$$

El hecho de que en D_μ admitamos $p_i \geq 0$ en lugar de $p_i > 0$ es indiferente, puesto que en los puntos $p = (p_1, \dots, p_k)$ en los que algún $p_i = 0$ se tiene que $\phi(p) = 0$.

Siendo:

$$\Delta^{k-1} = \{ p = (p_1, \dots, p_k) \in \mathfrak{R}^k : \sum_{i=1}^k p_i = 1; p_1, \dots, p_k \geq 0 \} \text{ (el simplex } (k-1)\text{-dimensional) y}$$

$$\forall \bar{\mu} \in \mathfrak{R}, \Pi_{\bar{\mu}} \equiv \sum_{i=1}^k p_i x_i = \bar{\mu} \text{ un hiperplano de } \mathfrak{R}^k,$$

es fácil observar que $D_\mu = \Pi_\mu \cap \Delta^{k-1}$ para todo $\mu \in I$.

Entonces, se verifica:

1. $\Pi_{\bar{\mu}}, \bar{\mu} \in \mathfrak{R}$ son todos ellos planos paralelos entre sí y, puesto que Δ^{k-1} es convexo, compacto de frontera regular, podemos afirmar que:

$$\forall \delta_1 > 0, \exists \delta_2 > 0 \text{ tal que si } |\mu - \bar{\mu}| \leq \delta_2, \text{ entonces:}$$

$$\forall p \in D_\mu, B(p, \delta_1) \cap D_{\bar{\mu}} \neq \emptyset$$

$$\forall p \in D_{\bar{\mu}}, B(p, \delta_1) \cap D_\mu \neq \emptyset,$$

donde $B(p, \delta)$ es la bola de centro p y radio δ

2. Δ^{k-1} es compacto y, puesto que ϕ es continua, será uniformemente continua sobre Δ^{k-1} , y por ello:

$$\text{Si } \forall \varepsilon > 0, \exists \delta > 0 : \|p_1 - p_2\| \leq \delta, \text{ entonces } |\phi(p_1) - \phi(p_2)| \leq \varepsilon$$

3. D_μ es compacto y ϕ es continua en D_μ , por lo que:

$$\forall \mu \in \Lambda, \exists p_\mu \in D_\mu \text{ tal que } f(\mu) = \text{Max} \phi(D_\mu) = \phi(p_\mu)$$

Sea entonces $\varepsilon \in \mathfrak{R}^+$ y aplicando 2., tenemos que:

$$\exists \delta_1 > 0 : \|p_1 - p_2\| \leq \delta_1, \text{ entonces } |\phi(p_1) - \phi(p_2)| \leq \varepsilon.$$

APROXIMACIÓN BAYESIANA A LA METODOLOGÍA DE LA VEROSIMILITUD EMPÍRICA

Aplicando 1. para δ_1 , existe $\delta_2 > 0$ tal que tal que si $|\mu - \bar{\mu}| \leq \delta_2$, entonces:

$$\begin{aligned} \forall p \in D_{\mu}, B(p, \delta_1) \cap D_{\bar{\mu}} &\neq \emptyset \\ \forall p \in D_{\bar{\mu}}, B(p, \delta_1) \cap D_{\mu} &\neq \emptyset \end{aligned}$$

Entonces, dados μ_1 y μ_2 pertenecientes a Λ tales que $|\mu_1 - \mu_2| \leq \delta_2$, por 1., tenemos que:

$$\begin{aligned} p_{\mu_1} \in D_{\mu_1} &\Rightarrow B(p_{\mu_1}, \delta_1) \cap D_{\mu_2} \neq \emptyset \Rightarrow \exists p_2 \in B(p_{\mu_1}, \delta_1) \cap D_{\mu_2} \\ p_{\mu_2} \in D_{\mu_2} &\Rightarrow B(p_{\mu_2}, \delta_1) \cap D_{\mu_1} \neq \emptyset \Rightarrow \exists p_1 \in B(p_{\mu_2}, \delta_1) \cap D_{\mu_1} \end{aligned}$$

y entonces:

$$\begin{cases} \|p_1 - p_{\mu_2}\| < \delta_1 \Rightarrow |\phi(p_1) - \phi(p_{\mu_2})| < \varepsilon \\ \|p_2 - p_{\mu_1}\| < \delta_1 \Rightarrow |\phi(p_2) - \phi(p_{\mu_1})| < \varepsilon \end{cases}$$

y como $f(\mu_2) = \phi(p_{\mu_2})$ y $f(\mu_1) = \phi(p_{\mu_1})$, entonces:

$$\begin{cases} f(\mu_2) - \varepsilon < \phi(p_1) \leq \phi(p_{\mu_1}) = f(\mu_1) \\ f(\mu_1) - \varepsilon < \phi(p_2) \leq \phi(p_{\mu_2}) = f(\mu_2) \end{cases}$$

Por tanto, $|f(\mu_1) - f(\mu_2)| < \varepsilon$,

y con ello demostramos que f es uniformemente continua en Λ , y por tanto, continua en Λ .

En el extremo inferior de Λ f tiene límite por la derecha igual a cero y en el extremo superior de Λ f tiene límite por la izquierda igual a cero, debido a su construcción, por lo que la función es continua en todo \mathfrak{R} .

1.7.3. Posibilidad de utilización en el teorema de Bayes.

Hemos visto en el apartado de la verosimilitud modificada, que ésta se puede combinar mediante el teorema de Bayes con información a priori dependiente exclusivamente de μ .

La gran diferencia que existe entre la función de verosimilitud elegida y la verosimilitud modificada vista con anterioridad es que podemos considerar un espacio paramétrico continuo para el parámetro estudiado, la media en nuestro caso, lo que nos ha permitido demostrar que la función de verosimilitud escogida es continua, lo cual es muy útil para estudiar la robustez bayesiana del problema; mientras que para la verosimilitud modificada dicho espacio paramétrico era discreto.

Como tenemos que la verosimilitud elegida es una función continua en un intervalo cerrado y acotado, tenemos que es una función integrable en dicho intervalo, por lo cual podemos aplicar el teorema de Bayes sin ningún problema para la función de verosimilitud elegida combinada con una distribución a priori relativa al parámetro media estudiado.

1.7.4. Suficiencia.

En este apartado vamos a estudiar la suficiencia, es decir vamos a buscar estadísticos suficientes de la distribución a posteriori.

El concepto de estadístico suficiente es de gran importancia para simplificar problemas estadísticos, y es una función de los datos que resume toda la información muestral disponible sobre el parámetro estudiado.

Vamos a estudiar los estadísticos suficientes respecto de la función de verosimilitud, puesto que en un estudio bayesiano, los estadísticos suficientes respecto de la distribución a posteriori son los estadísticos suficientes de la función de verosimilitud utilizada.

Es conocido que la distribución multinomial para k posibles valores de la variable aleatoria es una distribución exponencial $(k-1)$ -paramétrica, y de cuyo desarrollo se obtiene que su estadístico suficiente y completo y, por tanto, minimal suficiente está dado por la muestra obtenida de la variable aleatoria explicada.

En nuestro caso, tenemos que la verosimilitud, desarrollada para cada valor de μ vale:

$$f(x | \mu) = f^B(\mu) = n^{-n} \prod_i \left(n_i (1 + t(X_i - \mu))^{-1} \right)^{n_i} .$$

con la restricción de

$$\sum_i n_i \cdot n^{-1} (1 + t(X_i - \mu))^{-1} (X_i - \mu) = 0 ,$$

es decir, t siempre depende del parámetro μ desconocido, por lo cual al aplicar el teorema de factorización, siempre va a depender la función que contenga al parámetro μ directamente de la muestra, por la construcción de la verosimilitud.

Por consiguiente, podemos afirmar que el estadístico minimal suficiente que estamos buscando es la propia muestra.

1.7.5. Función de verosimilitud en muestras con la misma media y varianza muestral.

Se puede observar que la función de verosimilitud no depende exclusivamente de la media y la varianza de la distribución, sino que va a depender de los datos, pues si consideramos dos muestras con la misma media y varianza, pero con distinta asimetría, podemos comprobar que la función de verosimilitud obtenida es distinta en ambos casos.

Para ello, consideramos las dos muestras siguientes, que van a tener la misma media y varianza:

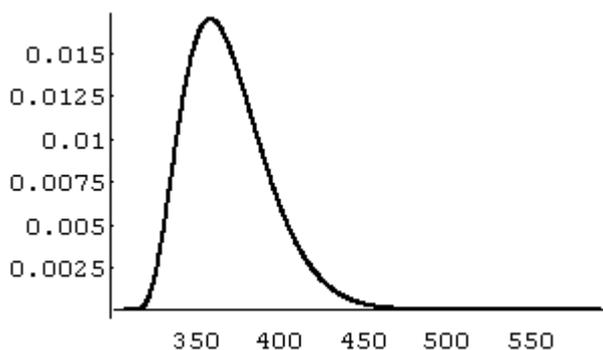
Muestra 1: 309, 310, 311, 312, 313, 315, 316, 321, 330, 350, 400, 500, 593.

Muestra 2: 411, 410, 409, 408, 407, 405, 404, 399, 390, 370, 320, 220, 127.

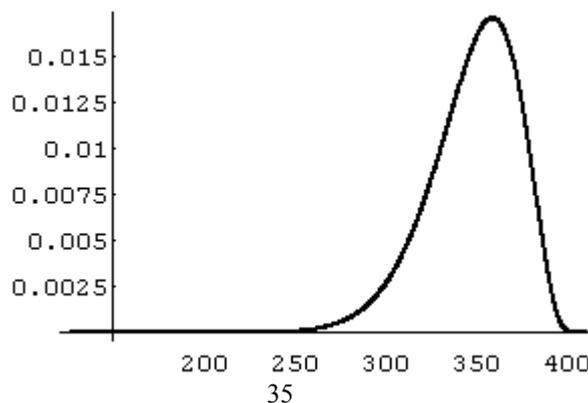
Podemos observar que ambas muestras tienen media igual a 360 y varianza muestral igual a 85.0276^2 .

La función de verosimilitud obtenida en ambos casos, podemos comprobar que es:

- En la primera muestra, tenemos:



- En la segunda muestra, tenemos:



CAPÍTULO 1: EL PROBLEMA DE LA VEROSIMILITUD

Por tanto, podemos ver que la verosimilitud elegida no va a depender exclusivamente de la media y varianza de la muestra obtenida.

Capítulo 2:

Distribución a priori y Robustez Bayesiana.

2.1. Introducción.

En este segundo capítulo de la memoria vamos a estudiar la elección de la distribución a priori correspondiente al parámetro desconocido de interés, la media de la variable aleatoria en nuestro caso, que vamos a utilizar para el estudio de la distribución a posteriori del parámetro respecto de la información muestral obtenida. Es importante tener en cuenta que la verosimilitud que hemos elegido en el Capítulo 1 es continua en un espacio paramétrico continuo y combinable mediante el teorema de Bayes con cualquier distribución a priori continua en dicho espacio paramétrico.

Por supuesto, si el investigador conoce por algún motivo la distribución a priori del parámetro estudiado, ésta sería la utilizada, pero nos vamos a centrar en el caso de que ésta sea desconocida para él. Comenzaremos analizando qué distribuciones a priori se han utilizado en aplicaciones previas de alguna función de verosimilitud parecida a la que hemos elegido en el capítulo anterior. Estas aplicaciones se refieren una al ámbito de la Auditoría de cuentas y otra al estudio de parques naturales en las Islas Canarias, en la que se utiliza la distribución a priori de máxima entropía. Comentaremos lo que sucede si la distribución a priori elegida es la distribución Uniforme.

CAPÍTULO 2: DISTRIBUCIÓN A PRIORI Y ROBUSTEZ BAYESIANA

Continuaremos dando un método de elección de la distribución a priori, que denominaremos método de máxima verosimilitud tipo II, aplicado a distintas distribuciones de probabilidad.

Al ser la elección de la distribución a priori, en cierto modo, subjetiva, vamos a utilizar la teoría desarrollada en la Robustez Bayesiana para determinarla, en la cual se utilizan clases de distribuciones a priori. Nos centraremos en el estudio de la robustez global, puesto que es la que mejor se adapta al caso que estamos estudiando, para lo cual es necesaria la continuidad de la verosimilitud utilizada, lo cual fue demostrado en el Capítulo 1 de la memoria.

2.2. Aplicaciones previas y utilización de la distribución Uniforme como distribución a priori.

Vamos a estudiar dos casos en los que la función de verosimilitud utilizada es la verosimilitud modificada estudiada en el capítulo anterior, que es un precedente de la función de verosimilitud que hemos elegido. Los casos estudiados se refieren a Auditoría de cuentas y Ecología, y en ellos se utilizan distintas distribuciones a priori. Además, estudiaremos el caso en el que la distribución a priori elegida sea la distribución Uniforme.

2.2.1. Auditoría de Cuentas.

Nos centraremos en los casos estudiados por Hernández y otros (1995) y Moreno Carretero y otros (1997), en los que, en el estudio de Auditorías de Cuentas utilizan como función de verosimilitud la verosimilitud modificada. Veamos dicho ejemplo:

En una auditoría de cuentas se quiere estudiar la cantidad total de error contable, que es una medida con fuerte significado intuitivo y de la que el auditor suele tener una información a priori relevante. Se supone un gran número de posibles estados de la naturaleza del error total igualmente espaciados. Consideramos que se utiliza muestreo DUS (Dollar-Unit Sampling) y que estamos interesados en utilizar esta observación muestral con información a priori para obtener una distribución a posteriori sobre el error total en la contabilidad.

APROXIMACIÓN BAYESIANA A LA METODOLOGÍA DE LA VEROSIMILITUD EMPÍRICA

Admitiendo errores de subvaloración y sobrevaloración, los posibles errores que se pueden presentar, expresados en porcentajes, varían desde -100 hasta 100 . Es decir, tenemos 201 categorías de fallos: $T_{-100}, T_{-99}, \dots, T_{-1}, T_0, T_1, \dots, T_{99}, T_{100}$, asociadas con los diferentes fallos del $-100\%, \dots, 100\%$. Sea p_i la probabilidad de que una unidad monetaria presente un fallo de categoría T_i ($i=-100, \dots, 100$).

Una muestra DUS de tamaño n proporciona la información muestral $(n_{-100}, \dots, n_i, \dots, n_{100})$, donde n_i es el número de unidades monetarias que han presentado un fallo del tipo T_i .

La magnitud relevante para el auditor, y sobre la cual suele tener información a priori, es la cantidad total de error, que viene dada por:

$$\lambda = K \cdot \sum_{i=-100}^{100} i \cdot p_i,$$

donde $K = \frac{RBV}{100}$, es decir, el valor registrado libro dividido entre 100.

Como se puede observar, la cantidad total del error es un parámetro similar a la media de la variable aleatoria estudiada, salvo que está multiplicado por una constante.

Utilizando el resultado estudiado en la verosimilitud modificada, considerando un gran número de posibles valores de la media igualmente espaciados, la verosimilitud del problema es:

$$M^B(\lambda) = \underset{\varphi^{-1}(\lambda)}{\text{Sup}} M(p_{-100}, \dots, p_{100}) \cdot I_{\varphi^{-1}(\lambda)} = \underset{p: \substack{p_i \geq 0 \\ \sum_{i=-100}^{100} p_i = 1 \\ K \cdot \sum_{i=-100}^{100} p_i \cdot i = \lambda}}{\text{Max}} L \cdot \prod_{i=-100}^{100} p_i^{n_i},$$

donde $\varphi^{-1}(\{\lambda\}) = \left\{ (p_{-100}, \dots, p_{100}) : p_i \geq 0, \sum_{i=-100}^{100} p_i = 1, K \cdot \sum_{i=-100}^{100} i \cdot p_i = \lambda \right\}$.

Esta verosimilitud modificada, como ya hemos visto, se puede combinar mediante el teorema de Bayes con cualquier distribución disponible por el auditor sobre la cantidad de error en la contabilidad, siempre que ésta haya sido discretizada para los valores de λ considerados.

Como distribuciones de probabilidad a priori se han utilizado por los autores anteriormente mencionados, las siguientes:

CAPÍTULO 2: DISTRIBUCIÓN A PRIORI Y ROBUSTEZ BAYESIANA

- La distribución Uniforme en el número de estados considerados de la magnitud estudiada, que estudiaremos más adelante.
- Una distribución Beta de parámetros 1 y β , donde β se puede calcular usando un error esperado λ_0 , es decir: $\lambda_0 = \frac{\alpha}{\alpha + \beta} = \frac{1}{1 + \beta}$; teniendo, para cada error esperado, una distribución a priori distinta.

2.2.2. Ecología. Distribución a priori de máxima entropía.

León y Vazquez-Polo (1998) proponen la utilización de la verosimilitud modificada combinada con la distribución a priori de máxima entropía en el estudio realizado sobre turistas europeos en las Islas Canarias referente al precio que están dispuestos a pagar para visitar los distintos parques naturales de las islas.

Dividimos el espacio muestral en un conjunto de k intervalos de precios, que es lo que ofrecemos a los r turistas encuestados. Sea B_i el punto medio del intervalo i -ésimo y p_i la probabilidad de que la respuesta esté dentro del intervalo i -ésimo.

El parámetro de interés considerado en el estudio es la cantidad media a pagar, ya que es función de las observaciones muestrales y de los parámetros del modelo. Entonces, un estimador para la media muestral está dado por:

$$\mu = \sum_{i=1}^k B_i \cdot p_i$$

Siguiendo el mismo razonamiento que en el caso de auditoría de cuentas, y considerando un gran número de posibles valores de la media igualmente espaciados, tenemos que la función de verosimilitud modificada es:

$$M^B(\mu) = \underset{\varphi^{-1}(\mu)}{\text{Sup}} M(p_1, \dots, p_k) \cdot I_{\varphi^{-1}(\mu)} = \underset{p: \substack{p_i \geq 0 \\ \sum_{i=1}^k p_i = 1 \\ \sum_{i=1}^k p_i \cdot B_i = \mu}}{\text{Max}} L \cdot \prod_{i=1}^k p_i^{n_i},$$

donde $\varphi^{-1}(\{\mu\}) = \left\{ (p_1, \dots, p_k) : p_i \geq 0, \sum_{i=1}^k p_i = 1, K \cdot \sum_{i=1}^k p_i \cdot B_i = \mu \right\}$.

Como distribución a priori, los autores consideran la distribución a priori de máxima entropía. Veamos brevemente cómo se elige en este caso la distribución a priori de máxima entropía de la media.

APROXIMACIÓN BAYESIANA A LA METODOLOGÍA DE LA VEROSIMILITUD EMPÍRICA

Las distribuciones a priori de máxima entropía son muy flexibles, en el sentido de que imponen pocas restricciones a la distribución a priori y solo necesitan una estimación del parámetro μ_0 y del rango de valores $[a,b]$ en el que está definido el parámetro. Si el espacio muestral es continuo, como es este caso, no existe una definición natural de entropía y los autores utilizan la definición de Jaynes (1968), según el cual la distribución a priori de máxima entropía para el parámetro media considerado, está dado por:

$$\pi(\mu) = \frac{\kappa \cdot e^{-\kappa\mu}}{e^{-\kappa a} - e^{-\kappa b}},$$

donde κ se obtiene resolviendo la siguiente ecuación no lineal:

$$\frac{\kappa(a \cdot e^{-\kappa a} - b \cdot e^{-\kappa b}) + e^{-\kappa a} - e^{-\kappa b}}{\kappa(e^{-\kappa a} - e^{-\kappa b})} = \mu_0,$$

pudiéndose obtener distintas distribuciones a priori de máxima entropía para distintos valores de μ_0 :

- Cuanto más próximo esté el valor de μ_0 a $(a+b)/2$, menos apuntada es la distribución.
- Si μ_0 es mayor que $(a+b)/2$, la distribución a priori de máxima entropía es asimétrica a la izquierda y tiene forma de J.
- Si μ_0 es menor que $(a+b)/2$, la distribución a priori de máxima entropía es asimétrica a la derecha y tiene forma de J invertida.

2.2.3. Utilización de la distribución Uniforme como distribución a priori.

Si no disponemos de información a priori sobre el parámetro, la media de la variable aleatoria de la que hemos tomado una muestra, o la información que disponemos sobre el parámetro es relativa a un comportamiento uniforme dentro de un intervalo de valores que contiene el comprendido entre el mayor y el menor valor muestral, tomaremos como distribución a priori del parámetro desconocido la distribución uniforme en el intervalo dado por el menor y el mayor valor muestral obtenido. Esta distribución tiene como función de densidad un valor constante dentro del intervalo de definición, lo cual nos va a dar un interesante resultado relativo a la distribución a posteriori.

Lema 2.1.

Si la distribución a priori del parámetro estudiado, la media, es la Uniforme en un intervalo que contiene al formado por el menor y el mayor valores muestrales, la distribución a posteriori del parámetro coincide con la función de verosimilitud más probable (o la función de verosimilitud empírica generalizada normalizada) de los datos.

Demostración:

Es consecuencia inmediata de la construcción de la distribución a posteriori.

Como consecuencia del lema 2.1., siempre que utilicemos esta distribución a priori, podemos obtener información a posteriori del parámetro aplicando toda la teoría relativa a la verosimilitud empírica y a la función de verosimilitud más probable, desarrollada en la literatura. Además, observaremos como estas dos corrientes son, simplemente, un caso particular del problema que vamos a abordar en esta memoria.

2.3. Método de la máxima verosimilitud tipo II para la elección de la distribución a priori.

Nos ocuparemos ahora del método que vamos a considerar para la elección de la distribución a priori. Este método va a ser el de máxima verosimilitud de tipo II.

Generalmente, en cada caso estudiado, vamos a tener distintas distribuciones como posibles alternativas de la distribución a priori. Este método nos permite contar con un criterio de elección de dicha distribución dentro de un conjunto de posibles distribuciones a priori.

Definición 2.1. (Berger (1985))

Sea Γ una clase de distribuciones a priori.

La distribución $\hat{\pi} \in \Gamma$ verifica para la muestra observada que:

$$m(x | \hat{\pi}) = \text{Sup}_{\pi \in \Gamma} m(x | \pi),$$

donde $m(x|\pi)$ es la densidad predictiva de la muestra dada la distribución a priori π , es decir, el denominador utilizado para obtener la distribución a posteriori al aplicar el teorema de Bayes:

APROXIMACIÓN BAYESIANA A LA METODOLOGÍA DE LA VEROSIMILITUD EMPÍRICA

$$m(x | \pi) = \int_{\Theta} f(x | \theta) \pi(\theta) d\theta$$

Entonces, decimos que $\hat{\pi} \in \Gamma$ es la distribución a priori de máxima verosimilitud tipo II de la clase Γ de distribuciones a priori.

Nosotros vamos a utilizar este criterio de elección siempre que tengamos una clase de posibles distribuciones a priori.

En nuestro caso, la densidad predictiva será de la forma:

$$m(x | \pi) = \int_{\Lambda} f^B(\mu) \cdot \pi(\mu) d\mu.$$

Esta distribución a priori depende claramente de la muestra obtenida, y la elección de una distribución a priori a partir de una muestra, es siempre algo subjetivo; sin embargo, puede justificarse la elección de una distribución concreta, ya que si $m(x|\pi)$ es pequeña, es muy improbable que π pueda ser válida, por lo que trabajar con π puede ser contraproducente.

2.3.1. Consideración de clases paramétricas usuales.

Vamos a estudiar cómo elegir los parámetros de una distribución a priori para el parámetro media de la variable aleatoria estudiada dentro de una clase formada por una única distribución de probabilidad conocida, pero cuyos parámetros son desconocidos:

- Distribuciones Normales:

Si sabemos que la distribución a priori del parámetro estudiado es Normal, pero desconocemos los parámetros, elegimos la clase de distribuciones a priori formada por todas las distribuciones Normales. La distribución a priori elegida podría ser la distribución de máxima verosimilitud tipo II de la clase. Entonces, la forma de buscarla sería encontrar la distribución ML-II de la clase formada por todas las distribuciones Normales, es decir, aquella distribución Normal con media y varianza tales que la densidad predictiva sea máxima:

$$\pi \sim N(\mu, \sigma) \text{ tal que } m(x|\pi(\mu, \sigma)) \text{ sea máximo.}$$

Para ello, utilizamos en la resolución el método de maximización no lineal denominado método del elipsoide, cuya idea básica consiste en generar una sucesión creciente de elipsoides que contienen al punto óptimo y cuyos centros convergen a él; el primer elipsoide se toma suficientemente grande como para que contenga al

CAPÍTULO 2: DISTRIBUCIÓN A PRIORI Y ROBUSTEZ BAYESIANA

conjunto factible del problema y a partir de él, se construye la sucesión partiendo por la mitad el anterior elipsoide y encerrando aquella mitad, que se sabe que contiene la solución óptima, en un nuevo elipsoide de la menor dimensión posible; repitiendo el proceso hasta obtener la convergencia. Pero los parámetros que se obtienen por este método son la media muestral para μ y cero para σ , es decir, obtenemos la distribución degenerada en la media muestral, cuyo uso carece de sentido. Por tanto, la distribución que vamos a elegir de la clase, va a ser aquella cuyos parámetros vienen dados por los estimadores de máxima verosimilitud para la distribución de probabilidad de la media de la variable aleatoria estudiada, obtenidos a partir de la muestra, es decir, la media muestral para μ y la desviación típica muestral dividida entre la raíz cuadrada del tamaño muestral para σ .

- **Distribuciones Gamma:**

Si sabemos que la distribución a priori del parámetro estudiado es una Gamma, pero desconocemos los parámetros, elegimos la clase de distribuciones a priori formada por todas las distribuciones Gamma. Para la elección de la distribución a priori, de nuevo intentamos utilizar el criterio de la máxima verosimilitud tipo II, pero los parámetros que se obtienen no son válidos, puesto que o bien α o bien β toman el valor infinito. La distribución que vamos a elegir de la clase, va a ser aquella cuyos parámetros vienen dados por los estimadores obtenidos a partir del método de los momentos, de manera que la media de la distribución coincida con la media muestral y la varianza de la distribución sea la varianza muestral dividida entre el tamaño muestral, es decir:

$$\begin{cases} \hat{\alpha} = \frac{\bar{x}^2 \cdot n}{S_x^2} \\ \hat{\beta} = \frac{S_x^2}{\bar{x} \cdot n} \end{cases} .$$

- **Distribuciones Beta:**

Si sabemos que la distribución a priori del parámetro estudiado es una Beta, pero desconocemos los parámetros, elegimos la clase de distribuciones a priori formada por todas las distribuciones Beta, considerándola en el intervalo formado por el menor y mayor valores muestrales obtenidos $[a,b]$, es decir, la variable considerada

es: $\frac{X - a}{b - a}$. Si utilizamos como criterio de elección de los parámetros el de máxima

APROXIMACIÓN BAYESIANA A LA METODOLOGÍA DE LA VEROSIMILITUD EMPÍRICA

verosimilitud tipo II, nos encontramos de nuevo con una distribución degenerada en la media. Por consiguiente, utilizamos de nuevo el método de los momentos en la elección de los parámetros, siendo:

$$\left\{ \begin{array}{l} \hat{\alpha} = \frac{n \cdot \bar{x} \left(\bar{x} - \bar{x}^2 - \frac{S_x^2}{n} \right)}{S_x^2} \\ \hat{\beta} = \frac{n(1 - \bar{x}) \left(\bar{x} - \bar{x}^2 - \frac{S_x^2}{n} \right)}{S_x^2} \end{array} \right.$$

- Otras distribuciones:

El procedimiento será análogo para cualquier otra clase formada por una distribución de probabilidad con parámetros desconocidos.

Si queremos obtener la distribución a priori a partir de la intersección de alguna de estas clases, obtenemos la distribución a priori de cada una de ellas y, de entre todas las obtenidas, la elegimos utilizando el método de máxima verosimilitud tipo II, es decir, nos quedamos con aquella que tenga una mayor densidad predictiva.

2.4. Robustez Bayesiana.

La elección de la distribución a priori, que tiene carácter subjetivo, va a estar basada en la información previa que ha ido acumulando el investigador. Pero, en la práctica, va a ser muy difícil elegir entre distintas distribuciones con características muy similares. Por ejemplo, si sabemos que la distribución de probabilidad a priori es simétrica y unimodal, puede ser complicado distinguir entre las distribuciones Normal y Cauchy. En otros casos, la elección de la distribución a priori depende de un grupo de investigadores, que pueden tener diferentes opiniones a priori. Por tanto, es conveniente utilizar distribuciones a priori dadas por clases, que nos van a ayudar a solucionar el problema de la elección de la distribución a priori original. Es decir, vamos a sustituir la especificación de una distribución a priori por toda una clase de distribuciones.

Por tanto, vamos a estudiar las herramientas básicas para el estudio de la propiedad de robustez respecto de la distribución a priori del análisis. El principal objetivo de la robustez bayesiana es cuantificar e interpretar la imprecisión inducida por

CAPÍTULO 2: DISTRIBUCIÓN A PRIORI Y ROBUSTEZ BAYESIANA

el conocimiento parcial de la distribución a priori. Además, se estudian métodos para reducir dicha imprecisión y alcanzar la robustez pretendida.

Entonces, vamos a elegir una clase de distribuciones a priori, de manera que la clase elegida esté de acuerdo con la información que disponemos relativa al parámetro estudiado. Propiedades deseables de la clase elegida deben ser: el cálculo de la robustez debe ser tan sencillo como sea posible, todas las distribuciones a priori razonables para ser utilizadas deben estar en la clase, las distribuciones que no sean razonables para ser utilizadas no deben estar en la clase y la clase se debe elegir a partir de la información a priori disponible sobre el parámetro estudiado.

En nuestro caso, el parámetro desconocido es la media de la variable aleatoria X estudiada, la verosimilitud es conocida y de la distribución a priori, por lo general, solo vamos a conocer características generales, que trataremos de resumir en una clase que contenga a todas las posibles distribuciones a priori del parámetro estudiado.

Esta imprecisión en la distribución a priori la vamos a medir a partir de una cantidad de interés (que puede ser la media a posteriori, la probabilidad a posteriori de un conjunto determinado, etcétera), calculando el rango en el que se mueve esta cantidad de interés a posteriori si la distribución a priori se encuentra dentro de la clase elegida. Si el rango es pequeño, diremos que el análisis es robusto, por lo que el análisis se puede realizar de manera satisfactoria; mientras que si esa diferencia es grande, diremos que existe carencia de robustez, por lo que deberemos intentar reducir la imprecisión, imponiendo nuevas restricciones a la clase de distribuciones a priori o tomando más datos muestrales.

La literatura ha desarrollado tres aproximaciones principales a la robustez bayesiana de la distribución a priori basadas en el estudio de una cantidad de interés a posteriori.

La primera es la *aproximación informal*, en la que se consideran unas pocas distribuciones a priori y se compara la cantidad de interés a posteriori estudiada que se obtiene con las distintas distribuciones a priori utilizadas. Esta aproximación es bastante simple y puede ser de ayuda, pero tiene el problema de que algunas distribuciones a priori compatibles con nuestra información relativa al parámetro no sean consideradas y éstas nos pueden dar resultados muy diferentes en la cantidad a posteriori estudiada.

La segunda es la *robustez global*, que considera la clase de todas las distribuciones a priori compatibles con la información disponible sobre el parámetro,

APROXIMACIÓN BAYESIANA A LA METODOLOGÍA DE LA VEROSIMILITUD EMPÍRICA

calculando el rango de valores entre los que se encuentra la cantidad de interés a posteriori estudiada según varía la distribución a priori dentro de la clase considerada.

La tercera es la *robustez local*, que está interesada en el índice de cambio en inferencias, con respecto a cambios en la distribución a priori, utilizando técnicas diferenciales para evaluar dicho índice.

Nosotros vamos a centrar nuestro estudio en la robustez global, puesto que la verosimilitud que estamos utilizando no nos va a permitir trabajar con la robustez local debido a la utilización de técnicas diferenciales. Respecto a la aproximación informal podemos comentar que ha sido ampliamente extendida por la robustez global, por lo que no vamos a entrar en su estudio.

2.5. Robustez Bayesiana global para clases de distribuciones a priori.

Como ya hemos comentado, vamos a elegir la distribución a priori del parámetro estudiado (la media en nuestro caso) a partir de una clase de distribuciones a priori que esté de acuerdo con la información previa disponible relativa a dicho parámetro. Consideraremos una cantidad de interés a posteriori y calcularemos el rango de valores entre los que se encuentra dicha cantidad de interés a posteriori estudiada según varía la distribución a priori dentro de la clase considerada. Además, para algunas clases de distribuciones a priori consideradas, daremos una distribución a priori que resume toda la clase, la distribución a priori de máxima verosimilitud tipo II de la clase, que utilizaremos como distribución a priori representativa de la clase si el análisis es robusto.

Veamos cómo podemos medir la robustez de un análisis:

Sea $\rho(x, \pi)$ la cantidad de interés a posteriori que queremos estudiar, que viene dada por la expresión:

$$\rho(x, \pi) = \int_{\Theta} h(\theta) \pi(\theta | x) d\theta = \frac{\int_{\Theta} h(\theta) f(x | \theta) \pi(\theta) d\theta}{\int_{\Theta} f(x | \theta) \pi(\theta) d\theta},$$

siendo $\pi(\theta|x)$ una distribución a posteriori obtenida a partir de una distribución a priori $\pi(\theta)$ perteneciente a la clase estudiada.

CAPÍTULO 2: DISTRIBUCIÓN A PRIORI Y ROBUSTEZ BAYESIANA

Las cantidades que vamos a considerar con más frecuencia son la media a posteriori del parámetro objeto de estudio (la media en nuestro caso) y la probabilidad a posteriori de un conjunto C , que se obtienen dando los siguientes valores a la función $h(\theta)$:

1. $h(\theta) = \theta$, para la media a posteriori.
2. $h(\theta) = I_C(\theta)$, para la probabilidad a posteriori de un conjunto C , donde $I_C(\theta)$ es la función indicadora sobre el conjunto C .

Calcularemos los rangos de la cantidad para cada clase, es decir, estudiaremos el comportamiento del intervalo:

$$\left(\inf_{\pi \in \Gamma} (\rho(x, \pi)), \sup_{\pi \in \Gamma} (\rho(x, \pi)) \right).$$

Para estudiar la existencia o no de robustez en un caso concreto, definimos la sensibilidad relativa (R.S.) de Sivaganesan (1991) como un factor normalizado, de manera que sus valores puedan ser comparados:

$$R.S. = \frac{\sup_{\pi \in \Gamma} (\rho(x, \pi)) - \inf_{\pi \in \Gamma} (\rho(x, \pi))}{2 \cdot \rho(x, \pi_0)} \cdot 100$$

Este factor puede ser entendido como una cantidad de variación, en porcentaje, de $E^\pi(h(\theta))$ cuando π varía sobre Γ .

Vamos a considerar dos tipos de clases de distribuciones a priori. La primera que vamos a considerar es la clase de contaminación, en la cual vamos a especificar una distribución a priori base presente en la clase, que tendrá un peso determinado, y va a ser contaminada por todas las posibles distribuciones a priori de la clase. La segunda, es la clase de distribuciones de probabilidad en una banda de medidas de probabilidad dada.

2.5.1. Clases de contaminación.

El primer tipo de clases de distribuciones a priori relativas al parámetro estudiado (media de la variable aleatoria estudiada) que vamos a estudiar es la clase de contaminación. Comenzaremos viendo su definición.

Definición 2.2. (Berger (1985))

Sea $\pi_0(\theta)$ la distribución a priori base relativa al parámetro desconocido, que va a ser elegida por el investigador.

Sea $\varepsilon \in [0,1]$ el grado de imprecisión en la distribución a priori inicial, que vamos a llamar grado de contaminación.

Entonces, definimos la clase de ε -contaminación como:

$$\Gamma_\varepsilon = \{\pi(\theta) = (1 - \varepsilon)\pi_0(\theta) + \varepsilon \cdot q(\theta) \mid q \in Z\},$$

donde Z es un conjunto de distribuciones de probabilidad, que se denomina clase contaminante. Posteriormente estudiaremos distintas elecciones de la clase contaminante.

El primer comentario que debemos realizar es relativo a la elección de la distribución a priori base elegida por el investigador. Esta distribución debe ser elegida de la clase de distribuciones que cumplen la información conocida por el investigador, es decir, debe formar parte de la clase contaminante seleccionada. Generalmente, se escoge debido sobre todo a su simplicidad matemática. Un método de selección a considerar es elegir la de máxima verosimilitud tipo II de entre un número de distribuciones a priori candidatas pertenecientes a la clase contaminante elegida (si tenemos que elegir una distribución dentro de una clase paramétrica la elegiremos utilizando el criterio ya comentado con anterioridad en el punto 2.3.1.).

Hay muchas razones para considerar estas clases de contaminación. En primer lugar, si tenemos que $\pi_0(\theta)$ es una distribución a priori creíble, pero en la que podemos cometer un error en su elección, en una cantidad ε . Así, las posibles distribuciones a priori que compensan este pequeño error se encuentran contenidas en la clase de distribuciones Γ_ε . Además, es sorprendentemente fácil trabajar con este tipo de clases de distribuciones y las clases son muy sensibles a la elección de la clase contaminante.

Este tipo de clases de distribuciones es ideal para analizar la robustez bayesiana del problema, ya que se puede ver que los posibles problemas de inferencia o decisión a tomar son esencialmente los mismos para cualquier distribución a priori de la clase. En realidad el problema se reduce esencialmente a encontrar máximos y mínimos de todas las distribuciones a priori de la clase, lo que nos va a permitir encontrar rangos para la

CAPÍTULO 2: DISTRIBUCIÓN A PRIORI Y ROBUSTEZ BAYESIANA

media a posteriori del parámetro estudiado, rangos para probabilidades a posteriori de conjuntos de confianza, etcétera.

Para elegir la distribución a priori con la que vamos a trabajar de cada clase de contaminación vamos a utilizar el criterio de máxima verosimilitud de tipo II; de esta forma, vamos a tener una distribución a priori concreta cuando consideremos una clase contaminante concreta y podremos realizar todos los cálculos con ella. Es decir, para $\pi(\theta) = (1 - \varepsilon)\pi_0(\theta) + \varepsilon \cdot q(\theta)$, $q \in Z$, tenemos que maximizar la densidad predictiva $m(x | \pi) = (1 - \varepsilon)m(x | \pi_0) + \varepsilon \cdot m(x | q)$ sobre π , lo cual, claramente, se hace maximizando $m(x|q)$ sobre $q \in Z$. Si dicho máximo se alcanza en $\hat{q} \in Z$, entonces la distribución a priori estimada, que va a depender de la muestra, está dada por:

$$\hat{\pi} = (1 - \varepsilon)\pi_0 + \varepsilon \cdot \hat{q}.$$

La elección de una distribución a priori dependiente de la muestra genera cierta controversia, pero se pueden dar muchas razones para hacerlo. Primero, si $m(x|\pi)$ es pequeña, es muy improbable que π pueda ser válida, por lo que trabajar con π puede ser contraproducente. Podemos recordar que cualquier $\pi \in \Gamma$ puede ser tratado como una representación razonable de la creencia a priori, así que $\hat{\pi}$ es simplemente la distribución a priori más posible, a la vista de la opinión a priori y los datos. Además, esta forma de elegir la distribución a priori, nos va a dar una serie de resultados razonables, como se puede observar en la literatura.

Antes de analizar distintas clases contaminantes, vamos introducir una nueva notación de la distribución, media y varianza a posteriori, que va a sernos bastante útil en el estudio de la robustez bayesiana:

Notación:

Dada una distribución a priori de la forma:

$$\pi(\theta) = (1 - \varepsilon)\pi_0(\theta) + \varepsilon \cdot q(\theta),$$

tenemos:

1°. Su densidad predictiva está dada por:

$$m(x | \pi) = (1 - \varepsilon)m(x | \pi_0) + \varepsilon \cdot m(x | q).$$

2°. Su distribución a posteriori correspondiente es:

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

APROXIMACIÓN BAYESIANA A LA METODOLOGÍA DE LA VEROSIMILITUD EMPÍRICA

donde $\lambda(x) \in [0,1]$ adopta la expresión:

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

3°. La media a posteriori del parámetro desconocido μ se puede escribir como:

$$\delta^\pi(x) = \lambda(x)\delta^{\pi_0}(x) + (1 - \lambda(x))\delta^q(x).$$

4°. La varianza a posteriori del parámetro desconocido es:

$$V^\pi(x) = \lambda(x)V^{\pi_0}(x) + (1 - \lambda(x))V^q(x) + \lambda(x)(1 - \lambda(x))(\delta^{\pi_0}(x) - \delta^q(x))^2$$

Gran parte de la bondad de la elección de estas clases contaminantes es la simplicidad de estas expresiones.

Veamos ahora distintas clases contaminantes, comenzando por la más simple, de manera que vamos a ir imponiendo sucesivamente nuevas restricciones para intentar alcanzar cada vez una mayor robustez. En el estudio, vamos a comenzar estudiando la robustez de cada clase y continuaremos obteniendo la distribución a priori de máxima verosimilitud tipo II de la clase considerada. Para ello, va a ser necesaria la condición de continuidad de la verosimilitud, lo cual es cierto gracias al teorema 1.5.

A) Todas las distribuciones.

Una clase contaminante que podemos considerar de forma natural es la clase formada por todas las distribuciones de probabilidad, dada una distribución a priori inicial π_0 , es decir, la clase de ε -contaminación:

$$\Gamma_1 = \{\pi : \pi = (1 - \varepsilon)\pi_0 + \varepsilon \cdot q, q \in Z_1\},$$

donde $Z_1 = \{\text{todas las distribuciones}\}$ es la clase contaminante.

Tiene el inconveniente de ser una clase de distribuciones a priori demasiado grande, e incluye muchas distribuciones a priori no razonables, pero proporciona resultados interesantes para empezar.

En este caso, así como en los casos de unimodalidad y unimodalidad y simetría, que estudiaremos a continuación, es fundamental considerar el siguiente lema:

Lema 2.2. (Sivaganesan y Berger (1987))

Si $A > 0$ y $f(x)$ y $g(x)$ son funciones continuas con $g(x) \geq 0$, entonces:

$$\sup_{dF(x)} \left(\inf_{dF(x)} \right) \frac{B + \int f(x) dF(x)}{A + \int g(x) dF(x)} = \sup_x \left(\inf_x \right) \frac{B + f(x)}{A + g(x)},$$

donde el supremo y el ínfimo están tomados sobre todas las distribuciones de probabilidad $dF(x)$, y $A, B, f(x)$ y $g(x)$ son tales que el supremo y el ínfimo de $\frac{B + f(x)}{A + g(x)}$ se alcanzan para algunos valores de x .

En el siguiente teorema vamos a ver que el rango de la media a posteriori de $h(\theta)$ sobre la clase de ε -contaminación se puede obtener encontrando los extremos de una función de una variable:

Teorema 2.1. (Sivaganesan (1988))

Sea $\rho(x, \pi)$ el valor esperado a posteriori de $h(\theta)$ con respecto a la distribución a priori π .

Entonces, para la clase de ε -contaminación Γ_1 , la clase contaminante Z_1 , y siempre que $\int_{\Theta} h(\theta) f(x | \theta) q(\theta) d\theta < \infty$, tenemos que:

$$\sup_{\pi \in \Gamma_1} \left(\inf_{\pi \in \Gamma_1} \right) \rho(x, \pi) = \sup_{\theta} \left(\inf_{\theta} \right) \frac{A_0 + h(\theta) f(x | \theta)}{A + f(x | \theta)},$$

donde $A = \frac{1 - \varepsilon}{\varepsilon} m(x | \pi_0)$ y $A_0 = A \cdot \rho(x, \pi_0)$

Veamos cómo podemos aplicar este resultado a nuestro caso: el parámetro θ de interés es la media μ de la variable aleatoria estudiada, la función $h(\mu)$ utilizada dependerá de la cantidad de interés que queramos estudiar y será función del parámetro de interés, y la función de verosimilitud $f(x | \theta)$ del teorema se corresponde con la función de verosimilitud $f^B(\mu)$ que hemos estudiado en el Capítulo 1 de esta memoria y que cumple la condición de continuidad impuesta por el lema 2.2. gracias al teorema 1.5.

APROXIMACIÓN BAYESIANA A LA METODOLOGÍA DE LA VEROSIMILITUD EMPÍRICA

Hemos estudiado la robustez bayesiana de la clase de distribuciones a priori formada por todas las distribuciones. En el caso en el que tengamos un grado de robustez suficiente, podemos utilizar una función de distribución de la clase como distribución a priori para obtener conclusiones a posteriori. Esta función de distribución es la distribución a priori de máxima verosimilitud tipo II. El siguiente teorema nos permite encontrar la distribución a priori de máxima verosimilitud tipo II y la correspondiente distribución a posteriori para la clase considerada:

Teorema 2.2. (Berger y Berliner (1986))

Consideramos una muestra procedente de una variable aleatoria X con verosimilitud $f(x|\theta)$. Supongamos que existe el estimador de máxima verosimilitud usual del parámetro θ , y es único $\hat{\theta}(x)$. Sea $\Gamma_1 = \{\pi : \pi = (1 - \varepsilon)\pi_0 + \varepsilon \cdot q, q \in Z_1\}$, donde $Z_1 = \{\text{todas las distribuciones}\}$ es la clase contaminante elegida.

Entonces:

1°. La distribución a priori de máxima verosimilitud tipo II está dada por:

$$\hat{\pi}(\theta) = (1 - \varepsilon)\pi_0(\theta) + \varepsilon \cdot \hat{q}_x(\theta),$$

donde \hat{q}_x toma probabilidad uno en el punto $\theta = \hat{\theta}(x)$, es decir, es la distribución degenerada en el estimador de máxima verosimilitud del parámetro.

2°. La distribución a posteriori generada está dada por:

$$\hat{\pi}(\theta | x) = \hat{\lambda}(x)\pi_0(\theta | x) + (1 - \hat{\lambda}(x))\hat{q}_x(\theta),$$

$$\text{donde } \hat{\lambda}(x) = \frac{(1 - \varepsilon)m(x | \pi_0)}{(1 - \varepsilon)m(x | \pi_0) + \varepsilon \cdot f(x | \hat{\theta}(x))}$$

Este teorema se puede trasladar sin ningún problema a nuestro caso, siendo la media μ de la variable aleatoria estudiada el parámetro desconocido θ , $f^B(\mu)$ la función de verosimilitud continua dada en el teorema por $f(x|\theta)$, y $\hat{\mu}(x)$ el estimador de máxima verosimilitud del parámetro, es decir, la media muestral.

B) Distribuciones unimodales.

Otra clase contaminante que puede resultar de interés es considerar la clase formada por contaminaciones unimodales, con la misma moda θ_0 que π_0 . Esta clase es más restrictiva que la anterior, puesto que consideramos la unimodalidad de la distribución a priori, por lo que eliminamos de la clase algunas distribuciones consideradas en la clase anterior.

Por tanto, consideramos la clase de ε -contaminación:

$$\Gamma_2 = \{\pi : \pi = (1 - \varepsilon)\pi_0 + \varepsilon \cdot q, q \in Z_2\},$$

donde $Z_2 = \{\text{distribuciones unimodales, con la misma moda } \theta_0 \text{ que } \pi_0\}$ es la clase contaminante.

Esta clase es particularmente razonable cuando la distribución a priori inicial π_0 es unimodal sobre θ_0 . La elección del valor θ_0 por parte del investigador se debe basar en la información previa que dispone relativa al parámetro estudiado; o bien, si conoce la existencia de la unimodalidad, pero desconoce la moda, utilizará la moda de la distribución a priori base elegida para el estudio.

La considerable simplicidad de trabajar con esta clase contaminante viene del hecho de que en gran parte del análisis Z_2 puede ser sustituida por:

$$Z_2' = \{\text{Uniformes}(\theta_0, \theta_0 + a), a \in \mathfrak{R}\},$$

donde la densidad cuando $a=0$ es la distribución degenerada en θ_0 . Por tanto, las optimizaciones afectarán solo a la variable a .

En este caso, es fundamental considerar el siguiente lema:

Lema 2.3. (Sivaganesan y Berger (1987))

Dada q una distribución de probabilidad perteneciente a la clase contaminante Z_2 y una función h tal que $\int_{\Theta} h(\theta) f(x | \theta) q(\theta) d\theta < \infty$, tenemos que:

$$\int_{\Theta} h(\theta) f(x | \theta) q(\theta) d\theta = \int_0^{\infty} H^h(z) dF(z),$$

donde F es alguna función de distribución, en concreto, la mixtura que produce q y

$$H^h(z) = \begin{cases} \frac{1}{z} \int_{\theta_0}^{\theta_0+z} h(\theta) f(x|\theta) d\theta & z \neq 0 \\ h(\theta_0) f(x|\theta_0) & z = 0 \end{cases}.$$

En el siguiente teorema vamos a ver como los extremos del rango de la media a posteriori de $h(\theta)$ sobre la clase de ε -contaminación se alcanzan sobre contaminaciones uniformes, sobre la clase:

$$\Gamma_2' = \{\pi(\theta) = (1-\varepsilon)\pi_0(\theta) + \varepsilon q(\theta) \mid q \approx U(\theta_0, \theta_0+z) \text{ ó } U(\theta_0-z, \theta_0)\} \subset \Gamma_2$$

Teorema 2.3. (Sivaganesan (1988))

Sea $\rho(x, \pi)$ el valor esperado a posteriori de $h(\theta)$ con respecto a la distribución a priori π . Entonces, para la clase de ε -contaminación Γ_2 , la clase contaminante Z_2 , y siempre que $\int_{\Theta} h(\theta) f(x|\theta) q(\theta) d\theta < \infty$, tenemos que:

$$\sup_{\pi \in \Gamma_2} \left(\inf_{\pi \in \Gamma_2} \right) \rho(x, \pi) = \sup_{\pi \in \Gamma_2} \left(\inf_{\pi \in \Gamma_2} \right) \rho(x, \pi) = \sup_z \left(\inf_z \right) \frac{A_0 + H^h(z)}{A + H^0(z)},$$

donde $A = \frac{1-\varepsilon}{\varepsilon} m(x|\pi_0)$, $A_0 = A \cdot \rho(x, \pi_0)$ y

$$H^0(z) = \begin{cases} \frac{1}{z} \int_{\theta_0}^{\theta_0+z} f(x|\theta) d\theta & z \neq 0 \\ f(x|\theta_0) & z = 0 \end{cases}$$

$$H^h(z) = \begin{cases} \frac{1}{z} \int_{\theta_0}^{\theta_0+z} h(\theta) f(x|\theta) d\theta & z \neq 0 \\ h(\theta_0) f(x|\theta_0) & z = 0 \end{cases}$$

En nuestro caso: el parámetro θ de interés es la media μ de la variable aleatoria estudiada, la función $h(\mu)$ utilizada dependerá de la cantidad de interés que queramos estudiar y será función del parámetro de interés, y la función de verosimilitud $f(x|\theta)$ del teorema corresponde a la función de verosimilitud continua $f^B(\mu)$ que hemos estudiado en el Capítulo 1 de esta memoria.

Hemos estudiado la robustez bayesiana de la clase de distribuciones a priori considerada. En el caso en el que tengamos un grado de robustez suficiente, podemos

CAPÍTULO 2: DISTRIBUCIÓN A PRIORI Y ROBUSTEZ BAYESIANA

utilizar una función de distribución de la clase como distribución a priori para obtener conclusiones a posteriori. Esta función de distribución es la distribución a priori de máxima verosimilitud tipo II. El siguiente teorema nos permite encontrar la distribución a priori de máxima verosimilitud tipo II y la correspondiente distribución a posteriori para la clase contaminante considerada:

Teorema 2.4. (Berger y Berliner (1986))

Consideramos una muestra procedente de una variable aleatoria X con verosimilitud $f(x|\theta)$. Sea $\Gamma_2 = \{\pi : \pi = (1 - \varepsilon)\pi_0 + \varepsilon q, q \in Z\}$, donde:

$$Z_2 = \{\text{distribuciones unimodales, con la misma moda } \theta_0 \text{ que } \pi_0\}$$

es la clase contaminante elegida.

Entonces:

La distribución a priori de máxima verosimilitud tipo II está dada por:

$$\hat{\pi}(\theta) = (1 - \varepsilon)\pi_0(\theta) + \varepsilon \hat{q}_x(\theta),$$

donde \hat{q}_x es una distribución Uniforme $(\theta_0, \theta_0 + \hat{a})$, donde \hat{a} es el valor de a que maximiza la función:

$$m(x|a) = \begin{cases} a^{-1} \int_{\theta_0}^{\theta_0+a} f(x|\theta) d\theta & a \neq 0 \\ f(x|\theta_0) & a = 0 \end{cases}.$$

Este teorema se puede trasladar igualmente a nuestro caso, siendo la media μ de la variable aleatoria estudiada el parámetro desconocido θ , y $f^B(\mu)$ la función de verosimilitud continua dada en el teorema por $f(x|\theta)$.

C) Distribuciones unimodales para las que la resultante a priori tiene la misma moda que la a priori base y $\pi(\theta_0) \leq \pi_0(\theta_0)$.

Una clase contaminante algo distinta de la anterior que puede resultar de interés es considerar la clase formada por todas las distribuciones de probabilidad tales que la distribución a priori resultante tiene la misma moda θ_0 que π_0 y, además, $\pi(\theta_0) \leq \pi_0(\theta_0)$.

Por tanto, consideramos la clase de ε -contaminación:

$$\Gamma_3 = \{\pi : \pi = (1 - \varepsilon)\pi_0 + \varepsilon q, q \in Z_3\}, \text{ donde}$$

APROXIMACIÓN BAYESIANA A LA METODOLOGÍA DE LA VEROSIMILITUD EMPÍRICA

$Z_3 = \{\text{distribuciones de probabilidad tales que } \pi \text{ es unimodal, con moda } \theta_0 \text{ (no necesariamente única y la misma que } \pi_0), \text{ y } \pi(\theta_0) \leq \pi_0(\theta_0)\}$ es la clase contaminante.

Esta clase es particularmente razonable cuando la distribución a priori inicial π_0 es unimodal sobre θ_0 . La elección del valor θ_0 por parte del investigador se debe basar en la información previa que dispone relativa al parámetro estudiado; o bien, si conoce la existencia de la unimodalidad, pero desconoce la moda, utilizará la moda de la distribución a priori base elegida para el estudio.

La utilización de esta clase exige que la verosimilitud elegida verifique que $1/f(x|\theta)$ sea convexa, lo que en nuestro caso podemos comprobar gráficamente, dibujando la función $1/f(x|\theta)$.

La considerable simplicidad de trabajar con esta clase contaminante viene del hecho de que en gran parte del análisis la clase de contaminación Γ_3 puede ser sustituida por la clase:

$$\tilde{\Gamma}_3 = \{\pi : \pi = (1 - \varepsilon)\pi_0 + \varepsilon \cdot q, \pi \text{ es de la forma } \tilde{\pi}\},$$

donde:
$$\tilde{\pi}(\theta) = \begin{cases} K & ; \theta \in B \\ (1 - \varepsilon)\pi_0(\theta) & ; \theta \notin B \end{cases}$$

para algún intervalo B y una constante K apropiada. Cuando $\theta_0 \notin B$, K es el valor de $(1 - \varepsilon)\pi_0(\theta)$ en el extremo más próximo a θ_0 ; y cuando $\theta_0 \in B$, K vale $(1 - \varepsilon)\pi_0(\theta_0)$. La longitud del intervalo B está determinada por la condición de $\tilde{\pi}$ unimodal con moda θ_0 y $\int_B (K - (1 - \varepsilon)\pi_0(\theta))d\theta = \varepsilon$.

Veamos la forma específica de $\tilde{\pi}$ para diferentes intervalos B .

Lema 2.4. (Sivaganesan (1989))

La forma específica de $\tilde{\pi}$ para diferentes intervalos B es:

i) Si $B = (\tilde{\theta}, w(\tilde{\theta})) \subset (\theta_0, \infty)$ para $\tilde{\theta} > \theta_0$, el correspondiente $\tilde{\pi}$ está dado por:

$$\tilde{\pi}(\theta) = \begin{cases} (1 - \varepsilon)\pi_0(\tilde{\theta}) & \theta \in [\tilde{\theta}, w(\tilde{\theta})] \\ (1 - \varepsilon)\pi_0(\theta) & \text{en otro caso} \end{cases}, \text{ donde}$$

$w(\theta) > \theta$ está definido por:

$$(1 - \varepsilon)\pi_0(\theta)(w(\theta) - \theta) - (1 - \varepsilon) \int_{\theta}^{w(\theta)} \pi_0(t)dt = \varepsilon.$$

CAPÍTULO 2: DISTRIBUCIÓN A PRIORI Y ROBUSTEZ BAYESIANA

ii) Si $B = (w'(\tilde{\theta}), \tilde{\theta}) \subset (-\infty, \theta_0)$ para $\tilde{\theta} < \theta_0$, el correspondiente $\tilde{\pi}$ está dado por:

$$\tilde{\pi}(\theta) = \begin{cases} (1-\varepsilon)\pi_0(\tilde{\theta}) & \theta \in [w'(\tilde{\theta}), \tilde{\theta}] \\ (1-\varepsilon)\pi_0(\theta) & \text{en otro caso} \end{cases}, \text{ donde}$$

$w'(\theta) < \theta$ está definido por:

$$(1-\varepsilon)\pi_0(\theta)(\theta - w'(\theta)) - (1-\varepsilon) \int_{w'(\theta)}^{\theta} \pi_0(t) dt = \varepsilon.$$

iii) Si θ_0 es un extremo del intervalo B , $\tilde{\pi}$ es de la forma:

$$\tilde{\pi}(\theta) = \begin{cases} (1-\varepsilon)\pi_0(\theta_0) & \theta \in [\theta_0, v(\rho)] \\ (1-\varepsilon)\pi_0(\theta) & \text{en otro caso} \end{cases}, \text{ para algún } 0 \leq \rho \leq \varepsilon,$$

donde $v(\rho) > \theta_0$ está definido por:

$$(1-\rho)\pi_0(\theta_0)(v(\rho) - \theta_0) - (1-\varepsilon) \int_{\theta_0}^{v(\rho)} \pi_0(\xi) d\xi = \varepsilon.$$

iv) Si θ_0 es un punto interior del intervalo B , $\tilde{\pi}$ es de la forma:

$$\tilde{\pi}(\theta) = \begin{cases} (1-\varepsilon)\pi_0(\theta_0) & \theta \in [\theta', \theta''] \\ (1-\varepsilon)\pi_0(\theta) & \text{en otro caso} \end{cases}, \text{ donde}$$

$\theta' < \theta''$ son soluciones de la ecuación:

$$(1-\varepsilon)\pi_0(\theta_0)(\theta'' - \theta') - (1-\varepsilon) \int_{\theta'}^{\theta''} \pi_0(\theta) d\theta = \varepsilon.$$

Una vez estudiada las forma de las distribuciones a priori de la clase de contaminación $\tilde{\Gamma}_3$ estudiaremos la obtención del rango de cantidades de interés a posteriori. En el siguiente teorema, estudiamos el rango de la media a posteriori:

Teorema 2.5. (Sivaganesan (1989))

Si $1/f(x|\theta)$ es convexa y la cantidad de interés es la media a posteriori, entonces:

$$\begin{aligned} \sup_{\pi \in \tilde{\Gamma}_3} (\inf) \delta^\pi(x) &= \sup_{\pi \in \tilde{\Gamma}_3} (\inf) \delta^\pi(x) = \\ &= \sup_{\pi \in \tilde{\Gamma}_3} (\inf) \frac{A \cdot \delta^{\pi_0}(x) + \int_B \theta \cdot f(x|\theta) (K - (1-\varepsilon)\pi_0(\theta)) d\theta}{A + \int_B f(x|\theta) (K - (1-\varepsilon)\pi_0(\theta)) d\theta}, \end{aligned}$$

APROXIMACIÓN BAYESIANA A LA METODOLOGÍA DE LA VEROSIMILITUD EMPÍRICA

$$\text{donde } A = \frac{(1 - \varepsilon)m(x | \pi_0)}{m(x | \tilde{\pi})}.$$

Como podemos observar, el rango de la media a posteriori en la clase Γ_3 se puede obtener maximizando y minimizando $\delta^\pi(x)$ en la clase $\tilde{\Gamma}_3$, lo que se puede hacer maximizando y minimizando $\delta^\pi(x)$ sobre todos los posibles intervalos B . Sivaganesan (1993) también establece el rango de la probabilidad a posteriori de un intervalo en la clase Γ_3 .

Estos resultados son directamente trasladables a nuestro caso, siempre que se cumpla la condición indicada, siendo la media μ de la variable aleatoria estudiada el parámetro desconocido θ , y $f^B(\mu)$ la función de verosimilitud dada en el teorema por $f(x|\theta)$.

Si queremos obtener la distribución a priori de máxima verosimilitud de tipo II, necesitamos que la verosimilitud $f(x|\theta)$ sea unimodal (como función de θ), con moda $\hat{\theta}$. Además, es conveniente considerar que tanto π_0 como f son no nulas y estrictamente monótonas a cada lado de las modas respectivas. Esto lo podemos comprobar directamente a partir de las gráficas de la verosimilitud y de la distribución a priori base. Supondremos también, sin pérdida de generalidad, que $\hat{\theta} \geq \theta_0$. La distribución va a depender de distintos casos, pero en todos ellos va a tener la forma:

$$\hat{\pi}(\theta) = \begin{cases} K & \theta \in B \\ (1 - \varepsilon)\pi_0(\theta) & \theta \notin B \end{cases}$$

donde K está implícitamente determinada por la restricción de que $\hat{\pi}$ tiene masa 1.

Veamos la distribución a priori de máxima verosimilitud tipo II en cada caso:

Teorema 2.6. (Berger y Berliner (1986))

Si definimos:

i) $v(\rho) \geq \theta_0$ para $-\varepsilon' \leq \rho \leq \varepsilon$ implícitamente como en el lema 2.4. y definimos:

$$V(\rho) = f(x | v(\rho))(v(\rho) - \theta_0) - \int_{\theta_0}^{v(\rho)} f(x | \theta) d\theta.$$

ii) $w(\theta) \geq \theta$ para $\theta_0 \leq \theta$ implícitamente como en el lema 2.4. y definimos:

CAPÍTULO 2: DISTRIBUCIÓN A PRIORI Y ROBUSTEZ BAYESIANA

$$W(\theta) = f(x | w(\theta))(w(\theta) - \theta) - \int_{\theta}^{w(\theta)} f(x | t) dt .$$

Entonces, la distribución a priori de máxima verosimilitud tipo II viene dada por $\hat{\pi}$, definida en cada caso como:

i) Si $V(\varepsilon) \geq 0$ y $\theta^* \in [\theta_0, \hat{\theta}]$ es la solución a $W(\theta) = 0$,

$$\hat{\pi}(\theta) = \begin{cases} (1 - \varepsilon)\pi_0(\theta^*) & \theta^* \leq \theta \leq w(\theta^*) \\ (1 - \varepsilon)\pi_0(\theta) & \text{en otro caso} \end{cases}$$

ii) Si $V(\varepsilon) < 0$ pero $V(-\varepsilon') \geq 0$ y sea $\rho^* \in [-\varepsilon', \varepsilon]$ tal que $V(\rho^*) = 0$,

$$\hat{\pi}(\theta) = \begin{cases} (1 - \rho^*)\pi_0(\theta_0) & \theta_0 \leq \theta \leq v(\rho^*) \\ (1 - \varepsilon)\pi_0(\theta) & \text{en otro caso} \end{cases}$$

iii) Si $V(-\varepsilon') < 0$ y $f(x|\theta_0) \leq f(x|v(-\varepsilon'))$, $\hat{\pi}$ es como en ii) con $\rho^* = -\varepsilon'$.

iv) Si $V(-\varepsilon') < 0$ y $f(x|\theta_0) > f(x|v(-\varepsilon'))$,

$$\hat{\pi}(\theta) = \begin{cases} (1 + \varepsilon')\pi_0(\theta_0) & \theta' \leq \theta \leq \theta'' \\ (1 - \varepsilon)\pi_0(\theta) & \text{en otro caso} \end{cases}$$

donde θ' y θ'' son las únicas soluciones a las ecuaciones:

$$\begin{aligned} f(x|\theta') &= f(x|\theta'') \\ (1 + \varepsilon')\pi_0(\theta_0)(\theta'' - \theta') - (1 - \varepsilon) \int_{\theta'}^{\theta''} \pi(\theta) d\theta &= \varepsilon \end{aligned}$$

Además, todas las cantidades que intervienen en la definición de $\hat{\pi}$ están bien definidas y son únicas.

Este teorema es directamente trasladable a nuestro caso, siempre que se cumplan las condiciones previamente indicadas, siendo la media μ de la variable aleatoria estudiada el parámetro desconocido θ , y $f^B(\mu)$ la función de verosimilitud dada en el teorema por $f(x|\theta)$.

Por tanto, si el estudio es suficientemente robusto, podemos utilizar la distribución a priori de máxima verosimilitud tipo II obtenida para obtener resultados a posteriori relativos al parámetro estudiado.

D) Distribuciones unimodales y simétricas.

Otra clase contaminante que puede resultar de interés es considerar la clase formada por contaminaciones simétricas y unimodales, con la misma moda θ_0 que π_0 . Esta clase es más restrictiva que las anteriores, puesto que consideramos la simetría de la distribución a priori, por lo que eliminamos de la clase algunas distribuciones consideradas en las clases anteriores.

Por tanto, consideramos la clase de ε -contaminación:

$$\Gamma_4 = \{\pi : \pi = (1 - \varepsilon)\pi_0 + \varepsilon \cdot q, q \in Z_4\},$$

donde $Z_4 = \{\text{distribuciones simétricas y unimodales, con la misma moda } \theta_0 \text{ que } \pi_0\}$, o lo que es lo mismo, $Z_4 = \{\text{densidades de la forma } q(|\theta - \theta_0|), \text{ donde } q \text{ es no creciente}\}$ es la clase contaminante.

Esta clase es particularmente razonable cuando la distribución a priori inicial π_0 es simétrica y unimodal sobre θ_0 , ya que bajo dichas circunstancias la distribución a priori resultante verifica las dos siguientes propiedades:

1. Los valores de θ lejanos a θ_0 no pueden dar un peso dominante.
2. Se consideran distribuciones a priori con colas mayores que la de la a priori inicial.

La elección del valor θ_0 por parte del investigador se debe basar en la información previa que dispone relativa al parámetro estudiado; o bien, si conoce la existencia de la unimodalidad, pero desconoce la moda, utilizará la moda de la distribución a priori base elegida para el estudio.

La considerable simplicidad de trabajar con esta clase contaminante viene del hecho de que en gran parte del análisis Z_4 puede ser sustituida por:

$$Z_4' = \{\text{Uniformes}(\theta_0 - a, \theta_0 + a), a \geq 0\},$$

donde la densidad cuando $a=0$ es la distribución degenerada en θ_0 . Por tanto, las optimizaciones afectarán solo a la variable a .

En este caso, es fundamental considerar el siguiente lema:

Lema 2.5. (Sivaganesan y Berger (1987))

Dada q una distribución de probabilidad perteneciente a la clase contaminante Z_4 y una función h tal que $\int_{\Theta} h(\theta) f(x|\theta) q(\theta) d\theta < \infty$, tenemos que:

CAPÍTULO 2: DISTRIBUCIÓN A PRIORI Y ROBUSTEZ BAYESIANA

$$\int_{\Theta} h(\theta) f(x | \theta) q(\theta) d\theta = \int_0^{\infty} H^h(z) dF(z),$$

donde F es alguna función de distribución, en concreto, la mixtura que produce q y

$$H^h(z) = \begin{cases} \frac{1}{2 \cdot z} \int_{\theta_0-z}^{\theta_0+z} h(\theta) f(x | \theta) d\theta & z \neq 0 \\ h(\theta_0) f(x | \theta_0) & z = 0 \end{cases}.$$

En el siguiente teorema vamos a ver como los extremos del rango de la media a posteriori de $h(\theta)$ sobre la clase de ε -contaminación se alcanzan sobre contaminaciones uniformes, sobre la clase:

$$\Gamma_4' = \{ \pi(\theta) = (1 - \varepsilon)\pi_0(\theta) + \varepsilon \cdot q(\theta) \mid q \approx U(\theta_0 - z, \theta_0 + z) \} \subset \Gamma_4$$

Teorema 2.7. (Sivaganesan (1988))

Sea $\rho(x, \pi)$ el valor esperado a posteriori de $h(\theta)$ con respecto a la distribución a priori π . Entonces, para la clase de ε -contaminación Γ_4 , la clase contaminante Z_4 , y siempre que $\int_{\Theta} h(\theta) f(x | \theta) q(\theta) d\theta < \infty$, tenemos que:

$$\sup_{\pi \in \Gamma_4} \left(\inf_{\pi \in \Gamma_4} \right) \rho(x, \pi) = \sup_{\pi \in \Gamma_4} \left(\inf_{\pi \in \Gamma_4} \right) \rho(x, \pi) = \sup_z \left(\inf_z \right) \frac{A_0 + H^h(z)}{A + H^0(z)},$$

donde $A = \frac{1 - \varepsilon}{\varepsilon} m(x | \pi_0)$, $A_0 = A \cdot \rho(x, \pi_0)$ y

$$H^0(z) = \begin{cases} \frac{1}{2 \cdot z} \int_{\theta_0-z}^{\theta_0+z} f(x | \theta) d\theta & z \neq 0 \\ f(x | \theta_0) & z = 0 \end{cases}$$

$$H^h(z) = \begin{cases} \frac{1}{2 \cdot z} \int_{\theta_0-z}^{\theta_0+z} h(\theta) f(x | \theta) d\theta & z \neq 0 \\ h(\theta_0) f(x | \theta_0) & z = 0 \end{cases}.$$

En nuestro caso, el parámetro θ de interés es la media μ de la variable aleatoria estudiada, la función $h(\mu)$ utilizada dependerá de la cantidad de interés que queramos estudiar y será función del parámetro de interés, y la función de verosimilitud $f(x | \theta)$ del

APROXIMACIÓN BAYESIANA A LA METODOLOGÍA DE LA VEROSIMILITUD EMPÍRICA

teorema corresponde a la función de verosimilitud $f^B(\mu)$ que hemos estudiado en el Capítulo 1 de esta memoria.

Hemos estudiado la robustez bayesiana de la clase de distribuciones a priori considerada. En el caso en el que tengamos un grado de robustez suficiente, podemos utilizar una función de distribución de la clase como distribución a priori para obtener conclusiones a posteriori. Esta función de distribución es la distribución a priori de máxima verosimilitud tipo II. El siguiente teorema nos permite encontrar la distribución a priori de máxima verosimilitud tipo II y la correspondiente distribución a posteriori para la clase contaminante considerada:

Teorema 2.8. (Berger y Berliner (1986))

Consideramos una muestra procedente de una variable aleatoria X con verosimilitud $f(x|\theta)$. Sea $\Gamma_4 = \{\pi : \pi = (1 - \varepsilon)\pi_0 + \varepsilon \cdot q, q \in Z_4\}$, donde:

$Z_4 = \{\text{distribuciones simétricas y unimodales, con la misma moda } \theta_0 \text{ que } \pi_0\}$ es la clase contaminante elegida.

Entonces:

La distribución a priori de máxima verosimilitud tipo II está dada por:

$$\hat{\pi}(\theta) = (1 - \varepsilon)\pi_0(\theta) + \varepsilon \cdot \hat{q}_x(\theta),$$

donde \hat{q}_x es una distribución Uniforme $(\theta_0 - \hat{a}, \theta_0 + \hat{a})$, donde \hat{a} es el valor de a que maximiza la función:

$$m(x|a) = \begin{cases} (2a)^{-1} \int_{\theta_0-a}^{\theta_0+a} f(x|\theta) d\theta & a > 0 \\ f(x|\theta_0) & a = 0 \end{cases}.$$

Este teorema es directamente trasladable a nuestro caso, siendo la media μ de la variable aleatoria estudiada el parámetro desconocido θ , y $f^B(\mu)$ la función de verosimilitud dada en el teorema por $f(x|\theta)$.

E) Contaminaciones con cuantiles fijos.

Una clase contaminante de interés es considerar la clase formada por distribuciones de probabilidad con algunos cuantiles especificados. Por tanto, consideramos la clase de ε -contaminación:

$$\Gamma_5 = \{ \pi : \pi = (1 - \varepsilon)\pi_0 + \varepsilon \cdot q, q \in Z_5 \}, \text{ donde}$$

$$Z_5 = \{ \text{distribuciones con } n \text{ cuantiles fijados} \} = \left\{ q : \int_{C_i} q(\theta) d\theta = \alpha_i, i = 1, \dots, n, \sum_{i=1}^n \alpha_i = 1 \right\}$$

es la clase contaminante y el conjunto de los C_i es una partición del espacio paramétrico.

Elegiremos como distribución a priori inicial π_0 una distribución perteneciente a la clase. La elección de los cuantiles por parte del investigador se debe basar en la información previa que dispone relativa al parámetro estudiado.

El supremo y el ínfimo de la media a posteriori de $h(\theta)$ para la distribución a priori dentro de la clase de contaminación dada se alcanza en una distribución a priori de la forma:

$$\pi^*(\theta) = (1 - \varepsilon)\pi_0(\theta) + \varepsilon \sum_{i=1}^n \alpha_i \delta_{\theta_i^*}(\theta),$$

donde θ_i^* es un punto en C_i o un límite de una serie de puntos en C_i . Este resultado se puede obtener a partir de este teorema, en el que calculamos el supremo y el ínfimo de una cantidad para q variando dentro de la clase contaminante:

Teorema 2.9. (Moreno y Cano (1991))

Para cualquier función $g_1(\theta)$ y $g_2(\theta)$ integrables con respecto a todas las a priori en Z_5 , tenemos que:

$$i) \sup_{q \in Z_5} (\inf) \int_{C_i} g_1(\theta) q(\theta) d\theta = \alpha_i \sup_{\theta \in C_i} (\inf) g_1(\theta).$$

$$ii) \sup_{q \in Z_5} (\inf) \frac{\int g_1(\theta) q(\theta) d\theta}{\int g_2(\theta) q(\theta) d\theta} = \sup_{\substack{\theta_i \in C_i \\ i=1, \dots, n}} (\inf) \frac{\sum_{i=1}^n \alpha_i g_1(\theta_i)}{\sum_{i=1}^n \alpha_i g_2(\theta_i)}.$$

Si elegimos $g_1(\theta) = h(\theta) f(x|\theta)$ y $g_2(\theta) = f(x|\theta)$, obtenemos el rango a posteriori de $\rho(x, q)$ para $h(\theta)$ la cantidad de interés a posteriori.

APROXIMACIÓN BAYESIANA A LA METODOLOGÍA DE LA VEROSIMILITUD EMPÍRICA

Para obtener el rango a posteriori de $\rho(x, \pi)$, sustituimos $g_1(\theta)$ por $(1 - \varepsilon) \int h(\theta) f(x | \theta) \pi_0(\theta) d\theta + \varepsilon \cdot h(\theta) f(x | \theta)$ y sustituimos $g_2(\theta)$ por $(1 - \varepsilon) \int f(x | \theta) \pi_0(\theta) d\theta + \varepsilon \cdot f(x | \theta)$, con lo que obtenemos el siguiente corolario:

Corolario 2.1. (Moreno y Cano (1991))

En las condiciones del teorema anterior, tenemos:

$$\sup_{\pi \in \Gamma_5} (\inf) \rho(x, \pi) = \sup_{\substack{\theta_i \in C_i \\ i=1, \dots, n}} (\inf) \frac{(1 - \varepsilon) \int h(\theta) f(x | \theta) \pi_0(\theta) d\theta + \varepsilon \sum_{i=1}^n \alpha_i h(\theta_i) f(x | \theta_i)}{(1 - \varepsilon) \int f(x | \theta) \pi_0(\theta) d\theta + \varepsilon \sum_{i=1}^n \alpha_i f(x | \theta_i)}$$

Podemos observar que el problema asociado con la clase contaminante Z_5 , se transforma en un problema de optimización con n restricciones, cuya solución no tiene que ser simple. La utilización de algoritmos de linealización nos ayuda. Por ejemplo, podemos ver que el supremo en el apartado ii) del teorema es la solución en λ de la

ecuación
$$0 = \sup_{\substack{\theta_i \in C_i \\ i=1, \dots, n}} (\inf) \left\{ \sum_{i=1}^n \alpha_i g_1(\theta_i) - \lambda \sum_{i=1}^n \alpha_i g_2(\theta_i) \right\}.$$

Si la cantidad de interés a posteriori considerada es la probabilidad de un determinado conjunto A , $\varphi(\theta) = I_A(\theta)$, obtenemos el siguiente corolario:

Corolario 2.2. (Moreno y Cano (1991))

$$\begin{aligned} \text{i) } \sup_{\pi \in \Gamma_5} P^\pi(A | x) &= \frac{(1 - \varepsilon) m(x | \pi_0) \beta_0 + \varepsilon \sum_{i \in J} \alpha_i \sup_{\theta \in A \cap C_i} f(x | \theta)}{(1 - \varepsilon) m(x | \pi_0) + \varepsilon \sum_{i \in K} \alpha_i \inf_{\theta \in C_i} f(x | \theta) + \varepsilon \sum_{i \in J} \alpha_i \sup_{\theta \in A \cap C_i} f(x | \theta)} \\ \text{ii) } \inf_{\pi \in \Gamma_5} P^\pi(A | x) &= \frac{(1 - \varepsilon) m(x | \pi_0) \beta_0 + \varepsilon \sum_{i \in I} \alpha_i \inf_{\theta \in C_i} f(x | \theta)}{(1 - \varepsilon) m(x | \pi_0) + \varepsilon \sum_{k \geq 1} \alpha_i \sup_{\theta \in A^c \cap C_k} f(x | \theta) + \varepsilon \sum_{i \in I} \alpha_i \inf_{\theta \in C_i} f(x | \theta)} \end{aligned}$$

donde $\beta_0 = P^{\pi_0}(A | x)$

$i \in I$ si y solo si $C_i \subset A$,

$i \in J$ si y solo si $A \cap C_i \neq \emptyset$.

$i \in K$ si y solo si $A \cap C_i = \emptyset$.

Una vez estudiada la robustez, vamos a obtener la distribución a priori de máxima verosimilitud tipo II de la clase de contaminación. Esta distribución la vamos a obtener a partir del siguiente teorema:

Teorema 2.10. (Moreno y González (1990))

Si consideramos la clase contaminante Z_5 , tenemos:

$$\sup_{q \in Z_5} m(x | q) = \sum_{i=1}^n \alpha_i \sup_{\theta \in C_i} f(x | \theta).$$

Como consecuencia del teorema, tenemos que la distribución de probabilidad que maximiza la densidad predictiva en la clase contaminante Z_5 es una distribución n -puntual dada por:

$$q^*(\theta) = \sum_{i=1}^n \alpha_i \delta_{\theta_i}^*(\theta).$$

Por tanto, la distribución a priori de máxima verosimilitud tipo II de la clase de contaminación está dada por:

$$\hat{\pi}(\theta) = (1 - \varepsilon)\pi_0(\theta) + \varepsilon \cdot \hat{q}(\theta),$$

donde \hat{q} cumple que $m(x | \hat{q}) = \sup_{q \in Z_5} m(x | q)$, y esta distribución no está necesariamente en la clase Z_5 . La condición que se debe de cumplir para que \hat{q} esté en la clase Z_5 la podemos ver en el siguiente lema:

Lema 2.6. (Moreno y González (1990))

Sea C_k^m una serie de conjuntos medibles no vacíos dados por:

$$C_k^m = C_k \cap \left\{ \theta : \sup_{\theta \in C_k} f(x | \theta) - f(x | \theta) \leq \frac{1}{m} \right\}, \quad k = 1, \dots, n,$$

donde m es un número real arbitrario.

APROXIMACIÓN BAYESIANA A LA METODOLOGÍA DE LA VEROSIMILITUD EMPÍRICA

Entonces, si $C_k^\infty = \bigcap_{m=1}^{\infty} C_k^m$, $k = 1, \dots, n$ son conjuntos no vacíos, la distribución de máxima verosimilitud tipo II está en Z_5 y está dada por $\hat{q}(\theta) = \sum_{i=1}^n \alpha_i \delta_{\theta_k}(\theta)$, donde θ_k es un punto arbitrario en C_k^∞ .

Entonces, si tenemos que $C_k^\infty = \{\theta_k\}$ para $k = 1, \dots, n$, podemos calcular la esperanza y varianza a posteriori de la distribución a priori de máxima verosimilitud tipo II de la clase Z_5 :

$$\delta^{\hat{q}}(x) = \frac{\sum_{k=1}^n \theta_k f(x | \theta_k) \alpha_k}{\sum_{k=1}^n f(x | \theta_k) \alpha_k} \text{ y}$$

$$V^{\hat{q}}(\theta | x) = \frac{\sum_{k=1}^n \theta_k^2 f(x | \theta_k) \alpha_k}{\sum_{k=1}^n f(x | \theta_k) \alpha_k} - (\delta^{\hat{q}}(x))^2$$

A partir de estas medidas podemos obtener la esperanza y varianza de la clase de contaminación Γ_5 , simplemente sustituyendo en las expresiones correspondientes ya estudiadas.

Una extensión de la clase de ε -contaminación estudiada es la clase de $\varepsilon(\theta)$ -contaminación estudiada por Moreno y otros (1996), que está dada por:

$$\Gamma_6 = \{ \pi : \pi = (1 - \varepsilon(\theta))\pi_0 + \varepsilon(\theta) \cdot q, q \in Z_5 \},$$

donde $\varepsilon(\theta) \in [0,1]$ refleja la incertidumbre de la distribución a priori base $\pi_0(\theta)$ en el punto θ . Para una función $\varepsilon(\theta)$ arbitraria no se garantiza el cumplimiento de

$$\left\{ q : \int_{C_i} q(\theta) d\theta = \alpha_i, i = 1, \dots, n, \sum_{i=1}^n \alpha_i = 1 \right\},$$

por lo que demuestran que una condición necesaria y suficiente para ello es que $\varepsilon(\theta) = \sum_{i=1}^n \varepsilon_i I_{C_i}(\theta)$, donde cada valor ε_i se puede interpretar como la incertidumbre que tenemos acerca de $\pi_0(\theta)$ en la clase C_i . Las

CAPÍTULO 2: DISTRIBUCIÓN A PRIORI Y ROBUSTEZ BAYESIANA

expresiones para la obtención de la robustez de la cantidad de interés a posteriori $h(\theta)$ son:

$$\sup_{\pi \in \Gamma_0}(\inf) \rho(x, \pi) = \sup_{\substack{\theta_i \in C_i \\ i=1, \dots, n}}(\inf) \frac{\sum_{i=1}^n (1 - \varepsilon_i) \int_{C_i} h(\theta) f(x | \theta) \pi_0(\theta) d\theta + \sum_{i=1}^n \varepsilon_i \alpha_i h(\theta_i) f(x | \theta_i)}{\sum_{i=1}^n (1 - \varepsilon_i) \int_{C_i} f(x | \theta) \pi_0(\theta) d\theta + \sum_{i=1}^n \varepsilon_i \alpha_i f(x | \theta_i)},$$

con $\alpha_i = \int_{C_i} \pi_0(\theta) d\theta, \quad i = 1, \dots, n.$

Para mejorar la robustez de estas clases de ε -contaminación podemos utilizar algoritmos iterativos. La idea es elegir una clase contaminante que contenga una serie de n cuantiles y estudiar la robustez de la clase. Si la robustez es suficiente para nuestro trabajo, utilizamos la clase elegida; pero si no lo es, utilizaremos un algoritmo iterativo, mediante el cual introducimos un nuevo cuantil en la clase C_i del espacio paramétrico en la que se alcanza menor robustez. La literatura nos ofrece al menos dos algoritmos para este caso, realizados por Liseo y otros (1996) y Moreno y otros (1996).

Algoritmo 2.1. (Liseo y otros (1996))

Paso 0: Si comenzamos con la clase de cuantiles dada por:

$$\Gamma_0(c^{(0)}, \alpha^{(0)}) = \left\{ \pi : \pi(C_j^{(0)}) = \alpha_j^{(0)}, j = 1, \dots, n, \sum_{j=1}^n \alpha_j^{(0)} = 1 \right\}, \quad \text{donde}$$

$$C^{(0)} = (C_1^{(0)}, \dots, C_n^{(0)}) \text{ es la partición inicial con } C_j^{(0)} = (c_{j-1}^{(0)}, c_j^{(0)}),$$

$$c^{(0)} = (c_0^{(0)}, \dots, c_n^{(0)}) \text{ y } \alpha^{(0)} = (\alpha_1^{(0)}, \dots, \alpha_n^{(0)}).$$

Paso 1: Calculamos el rango δ_θ de la cantidad de interés a posteriori $h(\theta)$.

Si δ_θ es menor o igual que una cantidad ε fijada, hemos terminado.

Si δ_θ es mayor que ε , pasamos al paso 2.

Paso 2: Calculamos el rango en cada una de las clases de la partición:

$$\delta_0^{(j)} = \delta(C_j^{(0)}, \Gamma_0) \text{ para } j=1, \dots, n, \text{ donde:}$$

$$E_A(h(\theta) | x) = E(h(\theta) | x, A) m(A) \text{ para } A \in B(\Theta), \text{ con:}$$

APROXIMACIÓN BAYESIANA A LA METODOLOGÍA DE LA VEROSIMILITUD EMPÍRICA

$$E(h(\theta) | x, A) = \frac{\int_A h(\theta) f(x | \theta) \pi(\theta) d\theta}{\int_A f(x | \theta) \pi(\theta) d\theta} \text{ y}$$

$$m(A) = \int_A f(x | \theta) \pi(\theta) d\theta.$$

Paso 3: Elegimos la clase $C_{j^*}^{(0)}$ como aquella que tiene máximo valor de $\delta_0^{(j)}$, puesto que es la menos robusta de la partición.

Paso 4: Dentro de $C_{j^*}^{(0)}$ incluimos otro cuantil, eligiendo un punto b tal que nuestra probabilidad subjetiva de los conjuntos $\left(a_{j^*-1}^{(0)}, b\right]$ y $\left(b, a_{j^*}^{(0)}\right]$ son ambas iguales a $\alpha_{j^*}^{(0)} / 2$.

Paso 5: Repetimos el proceso con la nueva clase de cuantiles formada.

Los autores demostraron que el algoritmo converge si $h(\theta)$ es continua y Γ_0^* -integrable.

En este algoritmo podemos observar como, en principio, no se utiliza una clase de ε -contaminación habitual, sino una clase de cuantiles directamente, es decir, se considera que ε vale uno en la clase de ε -contaminación.

Algoritmo 2.2. (Moreno y otros (1996))

Paso 0: Partimos de la clase de ε -contaminación:

$$\Gamma_6 = \{\pi : \pi = (1 - \varepsilon(\theta))\pi_0 + \varepsilon(\theta) \cdot q, q \in Z_5\}, \text{ donde}$$

$$Z_5 = \left\{ q : \int_{C_i} q(\theta) d\theta = \alpha_i, i = 1, \dots, n, \sum_{i=1}^n \alpha_i = 1 \right\}.$$

Paso 1: Calculamos el rango de la cantidad a posteriori $h(\theta)$.

Si es menor o igual que una cantidad ε fijada, hemos terminado.

Si es mayor que ε , pasamos al paso 2.

Paso 2: Calculamos el rango en cada clase de la partición:

$$\Gamma_6^i = \{\pi(\theta) : \pi(\theta) = \pi_0(\theta) + \varepsilon_i [q(\theta) - \pi_0(\theta)] I_{C_i}(\theta), q(\theta) \in Z_5\}.$$

Elegimos como clase menos robusta $C_k = (c_{k-1}, c_k)$ la que tiene mayor rango.

CAPÍTULO 2: DISTRIBUCIÓN A PRIORI Y ROBUSTEZ BAYESIANA

Paso 3: Si nuestra confianza en $\pi_0(\theta)I_{C_k}$ no es pequeña, para introducir un

nuevo cuantil en dicha clase, imponemos la condición $\alpha^* = \int_{c_{k-1}}^{c^*} \pi_0(\theta) d\theta$.

Si $R^*(x, c^*)$ es el rango a posteriori de nuestra cantidad de interés si la distribución a priori varía en la clase:

$$\Gamma_6^* = \left\{ \pi(\theta) : \pi(\theta) \in \Gamma_6, \int_{c_{k-1}}^{c^*} \pi(\theta) d\theta = \int_{c_{k-1}}^{c^*} \pi_0(\theta) d\theta \right\},$$

el punto c^* es el más favorable en la clase C_k , es decir:

$$R^*(x, c^*) = \inf_{b \in C_k} R^*(x, b)$$

y volvemos al paso 1.

Paso 4: Si nuestra confianza en $\pi_0(\theta)I_{C_k}$ es pequeña, $\pi_0(\theta)I_{C_k}$ lo sustituimos por otra distribución a priori base $\pi_1(\theta)I_{C_k}$, con lo que estamos comenzando de nuevo el problema de modelizar la incertidumbre a priori, pero solo en la clase C_k y trabajamos igual que en el paso 3.

Se puede observar que la clase en la que se introduce el nuevo cuantil depende de la información muestral utilizada.

F) Distribuciones unimodales con cuantiles conocidos.

Introducimos a la clase de ε -contaminación estudiada en el apartado anterior la restricción de unimodalidad. Esta clase, o aproximaciones a ella, ha sido estudiada por diversos autores, los cuales han obtenido complejos métodos para el estudio de su robustez.

Berger y O'Hagan (1988) estudian los rangos de probabilidades a posteriori para distribuciones a priori unimodales con cuantiles especificados, es decir, considerando la clase de ε -contaminación con ε igual a 1. El algoritmo utilizado era bastante complejo, por lo que los mismos autores, O'Hagan y Berger (1988) proponen un nuevo algoritmo para el mismo estudio, pero suavizando la hipótesis de unimodalidad a quasi-unimodalidad, definida en dicho trabajo. El trabajo que desarrolla esta clase de ε -contaminación es el de Sivaganesan (1991).

Consideramos la clase de ε -contaminación:

APROXIMACIÓN BAYESIANA A LA METODOLOGÍA DE LA VEROSIMILITUD EMPÍRICA

$\Gamma_7 = \{\pi : \pi = (1 - \varepsilon)\pi_0 + \varepsilon \cdot q, q \in Z_7\}$, donde

$Z_7 = \left\{ q : q \text{ es unimodal con moda } \theta_k \text{ y verifica } \int_{C_i} q(\theta) d\theta = \alpha_i, i = 1, \dots, n, \sum_{i=1}^n \alpha_i = 1 \right\}$ es la

clase contaminante y el conjunto de los C_i es una partición del espacio paramétrico.

Elegiremos como distribución a priori inicial π_0 una distribución perteneciente a la clase. La elección de los cuantiles y de la moda por parte del investigador se debe basar en la información previa que dispone relativa al parámetro estudiado.

El supremo y el ínfimo de la media a posteriori de $h(\theta)$ para la distribución a priori dentro de la clase de contaminación dada se obtiene a partir del siguiente teorema:

Teorema 2.11. (Sivaganesan (1991))

Sean $f(x|\theta)$ y $h(\theta)$ dos funciones continuas que satisfacen $\int f(x|\theta) d\theta < \infty$ y $\int f(x|\theta) h(\theta) d\theta < \infty$. Además, supongamos que, para cualesquiera valores reales a, c y λ tenemos que $\{h(a+z) - \lambda\} f(x|a+z) = c$ tiene al menos dos soluciones. Entonces,

$$\sup_{\pi \in \Gamma_7} (\inf) \rho(x, \pi) = \sup_{\pi \in \Gamma_7} (\inf) \int h(\theta) \pi(\theta | x) d\theta$$

se alcanza cuando la contaminación es una función a saltos con al menos un salto en el interior de cada C_i y, posiblemente otro en el extremo superior (inferior) si estamos por encima (debajo) de la moda.

La idea del teorema es aplicar el caso de distribuciones unimodales en cada C_i , puesto que en cada C_i la moda es uno de los extremos (el superior si estamos por debajo de la moda y el inferior si estamos por encima de ella). Utilizando las restricciones de los cuantiles, el punto z_i en el interior de C_i en el que q tiene un salto, se puede escribir en términos de la altura h_i en z_i y la altura h_i' en el extremo. Por tanto, q y la esperanza a posteriori de una cantidad de interés para $\pi \in \Gamma_7$ se puede escribir en términos de las alturas de los saltos de q .

Si la cantidad de interés estudiada es la media a posteriori $h(\theta) = \theta$, si $1/f(x|\theta)$ es convexo, se cumple la condición del teorema; y si queremos estudiar la probabilidad a posteriori de un conjunto C , $h(\theta) = I_C(\theta)$, si $f(x|\theta)$ es unimodal, se cumple la condición

CAPÍTULO 2: DISTRIBUCIÓN A PRIORI Y ROBUSTEZ BAYESIANA

del teorema. Para poder utilizar el teorema tenemos que comprobar que la verosimilitud utilizada cumple las condiciones indicadas, lo cual podemos comprobar gráficamente, al representar $1/f(x|\theta)$ y $f(x|\theta)$.

En cuanto a la distribución a priori de máxima verosimilitud de tipo II de la clase, Moreno y González (1990) estudian dos casos particulares de la clase considerada, las unimodales en las que la moda es un cuantil concreto y las unimodales a trozos con cuantiles conocidos.

Teorema 2.12. (Moreno y González (1990))

Sea $\pi_0(\theta)$ una distribución unimodal con moda θ_0 y consideramos la clase de ε -contaminación:

$$\Gamma_7' = \{ \pi : \pi = (1 - \varepsilon)\pi_0 + \varepsilon \cdot q, q \in Z_7' \},$$

donde $Z_7' = \left\{ q : \int_{-\infty}^{\theta_0} q(\theta) d\theta = \alpha, 0 < \alpha < 1 \right\}$.

Entonces:

i) Para cualquier $\pi \in \Gamma_7'$, tenemos que:

$$m(x | \pi) = (1 - \varepsilon)m(x | \pi_0) + \varepsilon \int_{-\infty}^{+\infty} H(a) dF(a), \text{ donde}$$

$$H(a) = \begin{cases} \frac{1}{a} \int_{\theta_0}^{\theta_0+a} f(x | \theta) d\theta & a \neq 0 \\ f(x | \theta_0) & a = 0 \end{cases} \quad \text{y}$$

F es alguna función de distribución en $(-\infty, \infty)$ que cumple:

$$\int_{-\infty}^{\theta_0} dF(a) = \alpha \quad \text{y} \quad \int_{\theta_0}^{+\infty} dF(a) = 1 - \alpha$$

ii) La distribución a priori de máxima verosimilitud tipo II de la clase Γ_7' está dada por $\hat{\pi}(\theta) = (1 - \varepsilon)\pi_0 + \varepsilon \hat{q}$, donde:

$$\hat{q}(\theta) = \frac{-\alpha}{\hat{a}_1(x)} 1_{(\theta_0+\hat{a}_1(x), \theta_0)}(\theta) + \frac{1-\alpha}{\hat{a}_2(x)} 1_{(\theta_0, \hat{a}_2(x)+\theta_0)}(\theta)$$

con $\hat{a}_1(x)$ y $\hat{a}_2(x)$ los valores definidos por:

$$H(\hat{a}_1(x)) = \sup_{a \leq 0} H(a) \quad \text{y} \quad H(\hat{a}_2(x)) = \sup_{a > 0} H(a).$$

Teorema 2.13. (Moreno y González (1991))

La distribución a priori de máxima verosimilitud de tipo II para la clase de ε -contaminación dada por:

$$\Gamma_7'' = \{ \pi : \pi = (1 - \varepsilon)\pi_0 + \varepsilon q, q \in Z_7'' \},$$

$$\text{donde } Z_7'' = \left\{ q : q \text{ es unimodal a trozos, } \int_{\theta_i}^{\theta_{i+1}} q(\theta) d\theta = \alpha_i, i = -n-1, \dots, -1 \right\}, \\ \left. \int_{\theta_i}^{\theta_{i+1}} q(\theta) d\theta = \alpha_{i+1}, i = 0, 1, \dots, m \right\},$$

considerando que una densidad unimodal a trozos significa que la función $q(|\theta - \theta_0|)$ es no creciente sobre cada intervalo de la partición, viene dada por la expresión $\hat{\pi}(\theta) = (1 - \varepsilon)\pi_0 + \varepsilon \hat{q}$, donde:

$$\hat{q}(\theta) = \sum_{i=-n}^0 \left\{ \hat{h}_{i-1} 1_{(\theta_{i-1}, \theta_i)}(\theta) + \frac{1}{\hat{a}_i + \theta_0 - \theta_i} 1_{(a_i + \theta_0, \theta_i)}(\theta) \right\} + \\ + \sum_{i=1}^{m+1} \left\{ \hat{h}_i 1_{(\theta_{i-1}, \theta_i)}(\theta) + \frac{1}{\hat{a}_i + \theta_0 - \theta_{i-1}} 1_{(\theta_{i-1}, \hat{a}_i + \theta_0)}(\theta) \right\},$$

con los $\hat{a}_i(x)$ que verifican:

$$\left\{ \begin{array}{l} \frac{1}{\theta_i - \theta_0 - \hat{a}_i} \int_{\theta_0 + \hat{a}_i}^{\theta_i} f(x | \theta) d\theta = \sup_{\theta_{i-1} - \theta_0 \leq a \leq \theta_i - \theta_0} \frac{1}{\theta_i - \theta_0 - a} \int_{\theta_0 + a}^{\theta_i} f(x | \theta) d\theta \quad i = -n, \dots, 0 \\ \frac{1}{-\theta_{i-1} + \theta_0 + \hat{a}_i} \int_{\theta_{i-1}}^{\theta_0 + \hat{a}_i} f(x | \theta) d\theta = \sup_{\theta_{i-1} - \theta_0 \leq a \leq \theta_i - \theta_0} \frac{1}{-\theta_{-1i} + \theta_0 + a} \int_{\theta_{i-1}}^{\theta_0 + a} f(x | \theta) d\theta \quad i = 1, \dots, m+1 \end{array} \right.$$

y los \hat{h}_i tales que:

$$\sup_{0 \leq h_i \leq \frac{\alpha_i}{\theta_{i+1} - \theta_i}} \left\{ \begin{array}{l} \sum_{i=-n}^0 (\alpha_{i-1} - h_{i-1}(\theta_i - \theta_{i-1})) \sup_{\theta_{i-1} - \theta_0 \leq a \leq \theta_i - \theta_0} \frac{1}{\theta_i - \theta_0 - a} \int_{\theta_0 + a}^{\theta_i} f(x | \theta) d\theta + \\ + \sum_{i=1}^{m+1} (\alpha_i - h_i(\theta_i - \theta_{i-1})) \sup_{\theta_{i-1} - \theta_0 \leq a \leq \theta_i - \theta_0} \frac{1}{a + \theta_0 - \theta_{i-1}} \int_{\theta_{i-1}}^{\theta_0 + a} f(x | \theta) d\theta + \\ + \sum_{i=-n}^0 h_{i-1} \int_{\theta_{i-1}}^{\theta_i} f(x | \theta) d\theta + \sum_{i=1}^{m+1} h_i \int_{\theta_{i-1}}^{\theta_i} f(x | \theta) d\theta \end{array} \right.$$

2.5.2. Clases de Bandas de Medidas de Probabilidad.

Continuamos estudiando otro tipo de clases de distribuciones a priori relativas al parámetro estudiado (media de la variable aleatoria estudiada). Esta clase de distribuciones de probabilidad $\Gamma(L,U)$ viene dada por dos funciones, $L(\theta)$ y $U(\theta)$, tales que $L(\theta) \leq U(\theta)$ para cualquier θ y $\int L(\theta)d\theta \leq 1 \leq \int U(\theta)d\theta$, imponiendo la condición de que cualquier distribución π de la clase verifique que $\int \pi(\theta)d\theta = 1$; es decir, acotamos la distribución a priori mediante dos funciones. Por tanto, la clase viene dada, para cualquier conjunto C por:

$$\Gamma(L,U) = \left\{ \pi(\theta) : L(C) \leq \pi(C) \leq U(C), \int_{\Theta} \pi(\theta)d\theta = 1 \right\}.$$

Las elecciones de las funciones L y U se eligen a partir de la incertidumbre del investigador respecto a la distribución a priori del parámetro estudiado.

De la misma forma que en las clases de contaminación imponíamos restricciones a la distribución a priori, en este tipo de clases vamos a hacer lo mismo, considerando distintas clases de bandas de medidas de probabilidad.

A) Clases con probabilidad de conjuntos acotadas.

Comenzamos estudiando el caso más simple, que está basado en la flexibilización del caso de la clase de cuantiles dada por:

$$\Gamma_{\alpha} = \left\{ \pi(\theta) : \int_{C_i} \pi(\theta)d\theta = \alpha_i, i = 1, \dots, n, \sum_{i=1}^n \alpha_i = 1 \right\},$$

donde θ es el parámetro desconocido y la clase de conjuntos de los C_i es una partición del espacio paramétrico. Esta clase es un caso particular de la clase de ε -contaminación de contaminaciones con cuantiles fijos, Γ_5 , en la que consideramos $\varepsilon = 1$.

La clase que vamos a estudiar está dada por:

$$\Gamma_{\varepsilon} = \left\{ \pi(\theta) : \alpha_i \leq \int_{C_i} \pi(\theta)d\theta \leq \beta_i, i = 1, \dots, n, \sum_{i=1}^n \alpha_i \leq \int_{\Theta} \pi(\theta)d\theta = 1 \leq \sum_{i=1}^n \beta_i \right\},$$

es decir, la probabilidad de cada conjunto C_i de la partición está acotada superior e inferiormente, de acuerdo a la información previa del investigador.

APROXIMACIÓN BAYESIANA A LA METODOLOGÍA DE LA VEROSIMILITUD EMPÍRICA

De esta clase podemos obtener con facilidad el rango de probabilidades a posteriori de cualquier conjunto, utilizando el siguiente teorema:

Teorema 2.14. (Moreno y Pericchi (1992))

$$i) \inf_{\pi \in \Gamma_8} P^\pi(C | x) = \inf_{\gamma \in D} \frac{\sum_{i \in y} \inf_{\theta \in C_i} f(x | \theta) \gamma_i}{\sum_{i \geq 1} \sup_{\theta \in C_i \cap C^c} f(x | \theta) \gamma_i + \sum_{i \in y} \inf_{\theta \in C_i} f(x | \theta) \gamma_i},$$

donde y es un subconjunto de los índices $\{1, \dots, n\}$ definido como $i \in y$ si y solo si $C_i \subset C$ y $D = \left\{ \gamma = (\gamma_1, \dots, \gamma_n) : \alpha_i \leq \gamma_i \leq \beta_i, \sum_{i=1}^n \gamma_i = 1 \right\}$.

ii) $\sup_{\pi \in \Gamma_8} P^\pi(C | x)$ se obtiene utilizando la relación general dada por

$$\sup_{\pi \in \Gamma_8} P^\pi(C | x) = 1 - \inf_{\pi \in \Gamma_8} P^\pi(C^c | x) \quad \text{y aplicándole el apartado i).}$$

B) Clase de bandas de medidas de probabilidad.

Vamos a utilizar una clase que generaliza a la clase anterior, puesto que vamos a considerar una clase que acota superior e inferiormente las probabilidades a priori de cualquier conjunto C , es decir, consideramos la clase:

$$\Gamma_9(L, U) = \left\{ \pi(\theta) : L(C) \leq \Pi(C) \leq U(C), \int_{\Theta} \pi(\theta) d\theta = 1 \right\},$$

donde $U(\cdot)$ y $L(\cdot)$ son medidas fijas predeterminadas que cumplen que $L(\Theta) < 1 < U(\Theta)$, con densidades dadas por $u(\cdot)$ y $l(\cdot)$.

Nuestro objetivo vuelve a ser encontrar el rango para una cantidad de interés a posteriori, para lo cual es necesario el siguiente lema:

Lema 2.7. (Moreno y Pericchi (1993))

Si la verosimilitud $f(x|\theta)$ y las medidas $U(\cdot)$ y $L(\cdot)$ cumplen que

$$\int_{\{\theta: f(x|\theta)=k\}} u(\theta) d\theta = 0 \quad \text{para todo valor de } k \geq 0,$$

Entonces, existe un número $z \geq 0$ (no necesariamente único), tal que:

$$\int_{\{\theta: f(x|\theta) \geq z\}} u(\theta) d\theta + \int_{\{\theta: f(x|\theta) < z\}} l(\theta) d\theta = 1.$$

CAPÍTULO 2: DISTRIBUCIÓN A PRIORI Y ROBUSTEZ BAYESIANA

Ahora ya podemos obtener los rangos de cualquier cantidad de interés a posteriori:

Teorema 2.15. (Moreno y Pericchi (1993))

En las mismas condiciones del lema 2.7., para la clase de distribuciones a priori $\Gamma_9(L, U)$ y para cualquier par de funciones $f(\theta)$ y $g(\theta)$ integrables respecto de $\pi \in \Gamma_9(L, U)$, tenemos:

$$i) \inf_{\pi \in \Gamma_9(L, U)} \frac{\int f(\theta)\pi(\theta)d\theta}{\int g(\theta)\pi(\theta)d\theta} = \frac{\int f(\theta)\pi_0(\theta)d\theta}{\int g(\theta)\pi_0(\theta)d\theta}, \text{ con}$$

$\pi_0(\theta) = u(\theta)I_A(\theta) + l(\theta)I_{A^c}(\theta)$, donde $A = \{\theta : f(\theta) - \lambda g(\theta) \leq z\}$ y los escalares λ y z están implícitamente definidos por las ecuaciones:

$$\begin{cases} \int_A \{f(\theta) - \lambda g(\theta)\}u(\theta)d\theta + \int_{A^c} \{f(\theta) - \lambda g(\theta)\}l(\theta)d\theta = 0 \\ \int_A u(\theta)d\theta + \int_{A^c} l(\theta)d\theta = 1 \end{cases}$$

ii) La solución para $\sup_{\pi \in \Gamma_9(L, U)}$ es análoga a la de i), sustituyendo el conjunto A por

$$B = \{\theta : f(\theta) - \lambda g(\theta) \geq z\}.$$

Si sustituimos $g(\theta) = f(x|\theta)$ y $f(\theta) = h(\theta)f(x|\theta)$, tenemos el rango de la cantidad de interés a posteriori, la cual dependerá de la elección de $h(\theta)$.

Si utilizamos nuestra verosimilitud, deberemos comprobar que se cumplen las condiciones del lema 2.7.

Como caso particular de este teorema, vamos a ver cómo podemos obtener el rango de la probabilidad a posteriori de cualquier conjunto C :

Corolario 2.3. (Moreno y Pericchi (1992))

Bajo las condiciones del lema 2.7., para un conjunto arbitrario C :

i) Si $u(C) + l(C^c) > 1$, entonces $\sup_{\pi \in \Gamma_9(L, U)} P^\pi(C|x) = P^{\pi_0}(C|x)$, donde

$$\pi_0(\theta) = u(\theta)I_{C \cap \{\theta: f(x|\theta) \geq z_C\}}(\theta) + l(\theta)I_{C^c \cap \{\theta: f(x|\theta) < z_C\}}(\theta), \text{ con } z_C \text{ tal que}$$

$$I_{\Theta}(\Theta) = 1.$$

APROXIMACIÓN BAYESIANA A LA METODOLOGÍA DE LA VEROSIMILITUD EMPÍRICA

ii) Si $U(C) + L(C^c) = 1$, entonces $\sup_{\pi \in \Gamma_9(L,U)} P^\pi(C|x) = P^{\pi_0}(C|x)$, donde

$$\pi_0(\theta) = u(\theta)I_C(\theta) + l(\theta)I_{C^c}(\theta).$$

iii) Si $U(C) + L(C^c) < 1$, entonces $\sup_{\pi \in \Gamma_9(L,U)} P^\pi(C|x) = P^{\pi_0}(C|x)$, donde,

$$\pi_0(\theta) = u(\theta)I_{C \cup (C^c \cap \{\theta: f(x|\theta) \geq z_C\})}(\theta) + l(\theta)I_{C^c \cap \{\theta: f(x|\theta) \geq z_C\}}(\theta) \quad \text{con } z_C \text{ tal que } \pi_0(\Theta) = 1.$$

Una vez estudiados los rangos de la cantidad de interés a posteriori, vamos a obtener la distribución a priori de máxima verosimilitud tipo II de la clase estudiada:

Corolario 2.4. (Moreno y Pericchi (1993))

Bajo las condiciones del lema 2.7., la distribución a priori de máxima verosimilitud tipo II de la clase $\Gamma_9(L,U)$ está dada por:

$$\hat{\pi}(\theta) = u(\theta)I_{\{\theta: f(x|\theta) \geq z\}}(\theta) + l(\theta)I_{\{\theta: f(x|\theta) < z\}}(\theta),$$

donde $z = z(x)$ está determinada por la condición $\int_{\Theta} \hat{\pi}(\theta) d\theta = 1$.

C) Restricciones lineales en la clase de bandas de medidas de probabilidad.

Vamos a utilizar una clase más restrictiva que la anterior, puesto que le vamos a añadir restricciones lineales, es decir, para cualquier conjunto C , consideramos la clase:

$$\Gamma_{10}(L,U) = \left\{ \pi(\theta) : L(C) \leq \Pi(C) \leq U(C), \int h_i(\theta)\pi(\theta)d\theta = \alpha_i, i = 1, \dots, n \right\},$$

donde $U(\cdot)$ y $L(\cdot)$ son medidas fijas predeterminadas que cumplen que $L(\Theta) < 1 < U(\Theta)$, $h_i(\cdot)$ son funciones conocidas y α_i son constantes conocidas.

Nuestro objetivo vuelve a ser encontrar el rango para una cantidad de interés a posteriori. Para ello, consideramos:

$$H_{max} = \sup_{\pi \in \Gamma_{10}(L,U)} \int h_0(\theta)\pi(\theta)d\theta \quad \text{y} \quad H_{min} = \inf_{\pi \in \Gamma_{10}(L,U)} \int h_0(\theta)\pi(\theta)d\theta,$$

donde las funciones h_i para $i=0, \dots, n$ son U -integrables y Γ es no vacío. Calcularemos H_{max} , puesto que H_{min} se calcula utilizando un argumento análogo.

Proposición 2.1. (Perone-Pacífico y otros (1996))

Dado $\alpha = (\alpha_1, \dots, \alpha_n)$, existe $d^* = (d_1^*, \dots, d_n^*)$ y $\pi^* \in \Gamma$ con densidad:

$$\pi^*(\theta) = \begin{cases} l(\theta) & \text{si } h_0(\theta) - \sum_{i=1}^n d_i^* h_i(\theta) < 0 \\ u(\theta) & \text{si } h_0(\theta) - \sum_{i=1}^n d_i^* h_i(\theta) > 0 \end{cases},$$

donde $l(\cdot)$ y $u(\cdot)$ son las densidades de $L(\cdot)$ y $U(\cdot)$, respectivamente y, además, tenemos que:

$$H_{max} = \int h_0(\theta) \pi^*(\theta) d\theta$$

Por tanto, el cálculo de H_{max} se reduce a encontrar d^* de manera que $\pi^* \in \Gamma_{10}(L, U)$. Entonces, si llamamos $\rho(x, \pi) = \frac{\int h(\theta) f(x | \theta) \pi(\theta) d\theta}{\int f(x | \theta) \pi(\theta) d\theta}$ a la cantidad de

interés a posteriori que queremos estudiar y si $D(\pi) = \int f(x | \theta) \pi(\theta) d\theta$, es decir, el denominador de dicha cantidad, notaremos $D_{max} = \sup_{\pi \in \Gamma_{10}(L, U)} D(\pi)$ y

$D_{min} = \inf_{\pi \in \Gamma_{10}(L, U)} D(\pi)$, que podemos obtenerlos a partir de la proposición 2.1. A partir de aquí, obtenemos:

Proposición 2.2. (Perone-Pacífico y otros (1996))

Bajo las condiciones:

- i) La verosimilitud $f(x | \theta)$ y el producto $h(\theta) f(x | \theta)$ son U -integrables.
- ii) $L(\{\theta: f(x | \theta) > 0\}) > 0$ ó $U(\{\theta: f(x | \theta) = 0\}) < 1$

tenemos que:

$$\sup_{\pi \in \Gamma_{10}(L, U)} \rho(x, \pi) = \sup_{y \in [D_{min}, D_{max}]} \frac{1}{y} \int h(\theta) f(x | \theta) \pi_y(\theta) d\theta,$$

donde π_y tiene densidad:

$$\pi_y(\theta) = \begin{cases} l(\theta) & \text{si } [h(\theta) - d_0^y] f(x | \theta) - \sum_{i=1}^n d_i^y h_i(\theta) < 0 \\ u(\theta) & \text{si } [h(\theta) - d_0^y] f(x | \theta) - \sum_{i=1}^n d_i^y h_i(\theta) > 0 \end{cases}$$

y (d_0^y, \dots, d_n^y) es tal que π_y verifica que:

$$\begin{cases} \int f(x|\theta)\pi_y(\theta)d\theta = y \\ \int h_i(\theta)\pi_y(\theta)d\theta = \alpha_i \quad i = 1, \dots, n \end{cases}$$

D) Bandas con amplitud constante.

Sivaganesan (1994) introduce la clase en la que la función de densidad se encuentra en una banda con amplitud constante, es decir:

$$\Gamma_{11} = \{\pi. l \leq \pi \leq u = l + \delta\}, \text{ para } \delta > 0 \text{ un valor fijo.}$$

Nuestro objetivo es calcular el rango de una cantidad de interés a posteriori

$$\rho(x, \pi) = \frac{\int h(\theta)f(x|\theta)\pi(\theta)d\theta}{\int f(x|\theta)\pi(\theta)d\theta}. \text{ Sea } \bar{\rho} = \sup_{\pi \in \Gamma_{11}} \rho(x, \pi), \text{ que se alcanza para la densidad}$$

$\bar{\pi}$. Entonces, tenemos que $\int f(x|\theta)\pi(\theta)(h(\theta) - \bar{\rho}) \leq 0$ para todo $\pi \in \Gamma_{11}$ y vale cero cuando $\pi = \bar{\pi}$. Por tanto, podemos obtener $\bar{\rho}$ eligiendo un λ para el cual:

$$B_h(\lambda) = \sup_{\pi \in \Gamma_{11}} \int f(x|\theta)\pi(\theta)(h(\theta) - \lambda)d\theta = 0.$$

Si definimos $L_h(\lambda) = \int (h(\theta) - \lambda)l(\theta)f(x|\theta)d\theta$, tenemos:

Teorema 2.16. (Sivaganesan (1994))

$$B_h(\lambda) = L_h(\lambda) + \delta \int_{A_s} f(x|\theta)(h(\theta) - \lambda)d\theta,$$

donde A_s es tal que tiene medida de Lebesgue $\mu(A_s) = \frac{(1 - L(\Theta))}{\delta} = s$ y

$$A_s = \{\theta : f(x|\theta)(h(\theta) - \lambda) > c\}, \text{ para un valor de } c \text{ constante conveniente.}$$

Para obtener el supremo y el ínfimo de la cantidad a posteriori, necesitamos que $f(x|\theta)$ y $h(\theta)f(x|\theta)$ sean integrables, lo cual es cierto en nuestro caso. Entonces, tenemos el siguiente teorema:

Teorema 2.17. (Sivaganesan (1994))

Consideramos la situación en la que $f(x|\theta)(h(\theta) - \lambda)$ tiene al menos dos extremos para cada valor de λ , por lo que A_s tiene amplitud s .

CAPÍTULO 2: DISTRIBUCIÓN A PRIORI Y ROBUSTEZ BAYESIANA

Si definimos
$$\rho(z) = \frac{L_h(0) + \delta \int_z^{z+s} h(\theta) f(x|\theta) d\theta}{L_1(0) + \delta \int_z^{z+s} f(x|\theta) d\theta},$$
 con $L_1(0) = \int f(x|\theta) l(\theta) d\theta,$

entonces:

$$\bar{\rho} = \sup_z(\inf_z) \rho(z).$$

E) Bandas con amplitud constante para distribuciones a priori unimodales.

Introducimos la restricción de unimodalidad a la clase anterior:

$$\Gamma_{12} = \{ \pi. l \leq \pi \leq l + \delta \text{ y } \pi \text{ unimodal con moda } \theta_0 \}, \text{ para } \delta > 0 \text{ un valor fijo.}$$

Para poder obtener el rango de la cantidad a posteriori $\rho(x, \pi)$, tenemos que definir una subclase de Γ_{12} que contenga todas las distribuciones a priori $\pi \in \Gamma_{12}'$ de la forma:

$$\pi(\theta) = \begin{cases} \min(c, u(\theta)) & \theta \in B \\ l(\theta) & \text{en otro caso} \end{cases},$$

donde $B = (a, b)$, $c = \max(l(a), l(b))$ cuando $\theta_0 \notin \bar{B}$, la clausura de B y $l(\theta_0) \leq c \leq u(\theta_0)$ cuando $\theta_0 \in \bar{B}$. La longitud de B y el valor de c cuando $\theta_0 \in \bar{B}$ se determinan imponiendo la condición de que $\pi \in \Gamma_{12}$.

Podemos ver que la distribución a priori de máxima verosimilitud de tipo II de la clase Γ_{12} se puede obtener a partir de la clase Γ_{12}' :

Lema 2.8. (Sivaganesan (1994))

Si $f(x|\theta)$ es unimodal, entonces la distribución a priori de máxima verosimilitud tipo II de la clase Γ_{12} pertenece a la clase Γ_{12}' .

Veamos cómo obtenemos el rango de la cantidad a posteriori de interés:

Teorema 2.18. (Sivaganesan (1994))

Sea $g(\theta) = f(x|\theta)(h(\theta) - \lambda)$ tal que tiene al menos dos extremos para cualquier λ y es integrable.

Entonces, $\sup_{\Gamma_{12}}(\inf) \int g \pi = \sup_{\Gamma_{12}'}(\inf) \int g \pi$, por lo que:

APROXIMACIÓN BAYESIANA A LA METODOLOGÍA DE LA VEROSIMILITUD EMPÍRICA

$$\sup_{\Gamma_{12}}(\inf) \rho(x, \pi) = \sup(\inf)_{\Gamma_{12}'} \rho(x, \pi).$$

F) Bandas de funciones de distribución.

Basu (1994) estudia las bandas de funciones de distribución:

$$\Gamma_{13} = \{F : F \text{ es función de distribución y } F_L(\theta) \leq F(\theta) \leq F_U(\theta), \forall \theta\}$$

y la desarrolla en posteriores trabajos. Nosotros no podemos utilizar esta clase porque en su desarrollo se exige la derivabilidad y continuidad de la derivada de la verosimilitud, lo cual no sabemos si es cierto para nuestra verosimilitud.

Capítulo 3:

Inferencia estadística y análisis de robustez con una función de verosimilitud empírica.

3.1. Introducción.

En este capítulo de la memoria vamos a estudiar diversos aspectos relativos a la inferencia estadística bayesiana, que aplicaremos a nuestro caso particular. Hasta ahora hemos estudiado la elección de la verosimilitud del problema (capítulo 1) y la utilización de clases de distribuciones a priori relativas al parámetro estudiado, junto con el estudio de su robustez (capítulo 2). Lo que vamos a hacer es, tras obtener la verosimilitud del problema, utilizar como distribución a priori una representativa de la clase de distribuciones a priori que hemos considerado y que cumple la exigencia de robustez, es decir, la sensibilidad relativa (R.S.) es menor que una cantidad fijada por el investigador; y dicha distribución a priori representativa será, si la conocemos, la de máxima verosimilitud tipo II de la clase. Además, compararemos los resultados obtenidos cuando la distribución a priori utilizada es la uniforme con el caso de la verosimilitud empírica generalizada que, como vimos en el capítulo 1 de la memoria, son casos equivalentes. Los aspectos que trataremos son la obtención estimadores puntuales bayesianos (moda, mediana y media a posteriori) para el parámetro estudiado,

la obtención de intervalos de confianza para la media a posteriori del parámetro estudiado y el contraste de hipótesis relativas al parámetro objeto de estudio.

Seguidamente, se realizará el desarrollo completo de un ejemplo ilustrativo del método presentado en la memoria.

3.2. Estimación puntual.

En este apartado queremos estimar puntualmente el parámetro de interés, la media, de la variable aleatoria que estamos estudiando.

Para ello, si utilizáramos un método clásico, la técnica más usual nos diría que el estimador de máxima verosimilitud para el parámetro θ es el valor de θ que maximiza la función de verosimilitud escogida para una muestra fija, $f(x|\theta)$. Pero, como estamos utilizando un método bayesiano, utilizaremos el estimador de máxima verosimilitud generalizado para el parámetro θ , que se define de la siguiente forma:

Definición 3.1.

El estimador de máxima verosimilitud generalizado del parámetro θ es la moda de la distribución a posteriori $\pi(\theta|x)$ obtenida, es decir, el valor de θ que maximiza la distribución a posteriori considerada, para una muestra fija, como una función de θ .

Entonces, la relación existente entre un estimador de máxima verosimilitud para el parámetro media de la variable aleatoria estudiada utilizando la función de verosimilitud elegida, es decir, el estimador de máxima verosimilitud clásico y un estimador de máxima verosimilitud generalizado va a depender siempre de la distribución a priori elegida. Además, si la moda de la distribución a priori y de la verosimilitud coinciden, el estimador de máxima verosimilitud generalizado va a coincidir también con el de las distribuciones anteriormente citadas.

Lema 3.1.

Si la moda de la verosimilitud y de la distribución a priori coinciden, entonces el estimador de máxima verosimilitud de la distribución a posteriori, es decir, la

APROXIMACIÓN BAYESIANA A LA METODOLOGÍA DE LA VEROSIMILITUD EMPÍRICA

moda a posteriori del parámetro sujeto a estudio, va a coincidir con la moda de las distribuciones anteriormente citadas.

Demostración:

Es consecuencia inmediata de la construcción de la distribución a posteriori.

Otros estimadores para el parámetro θ utilizados de manera común en inferencia bayesiana son la media y la mediana de la distribución a posteriori, que son frecuentemente mejores estimadores del parámetro que la moda.

3.3. Obtención de regiones de confianza.

Vamos a construir regiones de confianza con nivel de confianza $(1-\alpha)$ para el parámetro de interés, la media, de la variable aleatoria estudiada.

Desde el punto de vista bayesiano, se definen las regiones de confianza como:

Definición 3.2.

Una región de confianza con nivel de confianza $(1-\alpha)$ para el parámetro θ es un subconjunto C del espacio paramétrico tal que:

$$1-\alpha \leq P(C|x),$$

es decir, la probabilidad a posteriori de que θ esté en el conjunto C sea mayor o igual que $(1-\alpha)$.

Para elegir una región de confianza debemos minimizar su tamaño. Para ello, se deben incluir en el conjunto solo aquellos puntos con mayor densidad a posteriori, es decir, los valores más probables del parámetro. Entonces, definimos las regiones de confianza con densidad a posteriori más grande (HPD) como:

Definición 3.3.

La región de confianza HPD con nivel de confianza $(1-\alpha)$, es el subconjunto C del espacio paramétrico de la forma:

$$C = \{\theta \in \Theta : \pi(\theta | x) \geq k(\alpha)\},$$

donde:

$\pi(\theta | x)$ es la distribución a posteriori; y

$k(\alpha)$ es la mayor constante tal que:

$$P(C | x) \geq 1 - \alpha .$$

Las regiones de confianza para las distintas distribuciones a posteriori van a depender de la distribución a priori utilizada.

Como hemos comentado con anterioridad, podemos aplicar el teorema de Wilks para obtener la región de confianza en el caso en el que la distribución a priori considerada sea la distribución uniforme, lo cual va a ser una aproximación a la región de confianza HPD. En este caso.

Sea $L(\mu)$ la verosimilitud empírica.

Sea $L(\hat{\mu})$ el máximo de la verosimilitud empírica.

Sea $R(\mu) = L(\mu)/L(\hat{\mu})$ la razón de verosimilitud empírica.

Sea $l(\mu) = -2 \cdot \log(R(\mu))$ el logaritmo de la razón de verosimilitud empírica.

Aplicando el teorema 1.1. (análogo al de Wilks) del apartado 1.3., la región de confianza es:

$$R_c = \{\mu: l(\mu) \leq \chi^2_{1,1-\alpha}\}$$

Entonces, tenemos que las regiones de confianza vienen dadas por los valores de μ tales que:

$$l(\mu) \leq \chi^2_{1,0.90} = 2.71 \text{ para un nivel de confianza del 90\%.}$$

$$l(\mu) \leq \chi^2_{1,0.95} = 3.84 \text{ para un nivel de confianza del 95\%.}$$

$$l(\mu) \leq \chi^2_{1,0.99} = 6.64 \text{ para un nivel de confianza del 99\%.}$$

3.4. Contrastes de hipótesis.

Veamos cómo podemos realizar contrastes de hipótesis para el parámetro estudiado.

Para ello, vamos a comparar la relación existente entre los contrastes de hipótesis para la media realizados mediante la verosimilitud empírica y los realizados utilizando la distribución a posteriori:

APROXIMACIÓN BAYESIANA A LA METODOLOGÍA DE LA VEROSIMILITUD EMPÍRICA

1. El caso de los contrastes de hipótesis para un parámetro utilizando la Verosimilitud Empírica se resuelve de manera análoga a un contraste de hipótesis clásico, utilizando el teorema de Wilks:

1.1. En el caso de un contraste bilateral, es decir,

$$\begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta \neq \theta_0 \end{cases},$$

rechazaremos la hipótesis nula con un nivel de significación α si el valor θ_0 no está dentro del intervalo de confianza para dicho parámetro con un nivel de confianza $(1-\alpha)$.

1.2. En el caso de un contraste unilateral, es decir,

$$\begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta \leq \theta_0 \end{cases} \quad \text{o} \quad \begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta \geq \theta_0 \end{cases}$$

rechazaremos la hipótesis nula con un nivel de significación α si el valor θ_0 no está dentro del intervalo de confianza a un lado correspondiente para dicho parámetro con un nivel de confianza $(1-\alpha)$.

2. El caso de los contrastes de hipótesis para un parámetro utilizando la técnica introducida en la memoria, se resuelve de manera análoga a un contraste de hipótesis bayesiano, aunque, como ya hemos visto en los puntos anteriores, si la distribución a priori es la distribución uniforme, como estamos en el mismo caso de la verosimilitud empírica, se podría resolver mediante la técnica de la Verosimilitud Empírica.

Consideremos el caso general de un contraste de hipótesis para un parámetro:

$$\begin{cases} H_0 : \theta \in \Theta_0 \\ H_1 : \theta \in \Theta_1 \end{cases}.$$

En este caso, debemos calcular la probabilidad a posteriori de ambas hipótesis $\alpha_0 = P(\Theta_0 | x)$ y $\alpha_1 = P(\Theta_1 | x)$, que dependen tanto de los datos como de la opinión a priori, para decidir, en consecuencia, entre las hipótesis; también van a ser de interés las probabilidades a priori de ambas hipótesis (π_0 y π_1); definimos α_0/α_1 como la razón de probabilidades a posteriori de la hipótesis nula frente a la alternativa y π_0/π_1 como la razón de probabilidades a priori; y definimos la

cantidad $B = \frac{\text{razón de probabilidades a posteriori}}{\text{razón de probabilidades a priori}}$ como el factor Bayes respecto a la hipótesis nula.

Entonces, para resolver el contraste, rechazaremos la hipótesis nula si la razón de probabilidades a posteriori es menor que uno, es decir, la probabilidad a posteriori de la hipótesis alternativa es mayor que la de la hipótesis nula; a partir de la razón de probabilidades a priori, podemos observar la relación que existe a priori entre las hipótesis; mediante el factor Bayes respecto de la hipótesis nula podemos observar la influencia de los datos muestrales en la resolución del contraste después de tener la información muestral (a posteriori) y antes de tenerla (a priori).

La única situación en la que se debe de modificar esta técnica es en el caso de que la hipótesis nula sea simple, es decir, $\theta = \theta_0$ y la distribución a priori sea continua, ya que en este caso, la probabilidad a priori de la hipótesis nula sería cero y, por tanto, la probabilidad a posteriori de la hipótesis nula también sería cero.

Entonces, tenemos dos opciones:

- a) Considerar la hipótesis nula como un intervalo centrado en θ_0 , de manera que contenga todos los valores que se puedan considerar indistinguibles de θ_0 , y trabajar como en el caso general.
- b) Asignar una probabilidad a priori no cero π_0 a la hipótesis nula y $\pi_1 = 1 - \pi_0$ al resto del espacio muestral. Por tanto, la distribución a priori quedaría:

$$\begin{cases} \pi_0 & ; \theta = \theta_0 \\ \pi_1 \pi(\theta) & ; \theta \neq \theta_0 \end{cases},$$

donde $\pi(\theta)$ es la distribución a priori original.

3.5. Análisis de Robustez con una función de verosimilitud empírica.

Vamos a analizar el desarrollo de la metodología de trabajo propuesta en la memoria mediante el estudio de un caso particular, que nos puede servir como ilustración, utilizando un programa informático realizado en Mathematica, que se puede ver en el Apéndice. Vamos a estudiar los ingresos monetarios familiares durante el primer trimestre de 1997, medidos en miles de pesetas, datos procedentes de la Encuesta

APROXIMACIÓN BAYESIANA A LA METODOLOGÍA DE LA VEROSIMILITUD EMPÍRICA

continua de Presupuestos familiares correspondientes al primer trimestre de 1997 (INE), en la que se dispone de una muestra con 3114 datos, cuyo valor mínimo es de cero pesetas y el máximo de 9994000 pesetas.

Queremos estudiar la distribución de probabilidad de la media de la variable aleatoria ingresos monetarios familiares durante el primer trimestre de 1997, medidos en miles de pesetas. Para ello, vamos a realizar un estudio bayesiano sobre ella, por lo que calcularemos, en primer lugar, la función de verosimilitud para el problema, continuaremos con la elección de la distribución a priori para el parámetro media de interés utilizando la robustez bayesiana y, finalmente, obtendremos conclusiones a posteriori.

3.5.1. Elección de la función de verosimilitud del problema.

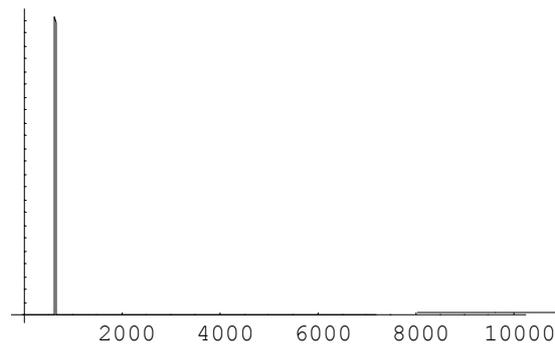
Lo primero que vamos a hacer es calcular la función de verosimilitud para este caso. La función de verosimilitud que vamos a utilizar es la propuesta en el Capítulo 1 de la memoria, por las razones ya comentadas con anterioridad. Por tanto, la función de verosimilitud considerada es:

$$f(x | \mu) = f^B(\mu) = \begin{cases} \underset{p=(p_1, \dots, p_k)}{\text{Max}} \prod_{i=1}^k p_i^{n_i} \\ \text{s.a.} \left\{ \begin{array}{l} \sum_{i=1}^k p_i = 1 \\ \sum_{i=1}^k p_i \cdot x_i = \mu \\ p_i > 0 \quad i = 1, \dots, k \end{array} \right. \end{cases}$$

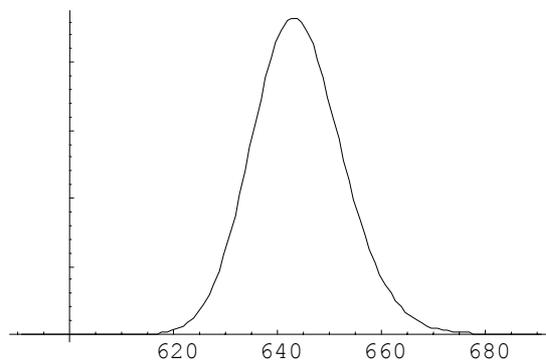
En la práctica, lo que realiza el programa informático, es calcular la verosimilitud puntualmente para valores de la media (μ) de la variable aleatoria en un número de puntos suficientemente grande a lo largo del intervalo de valores entre el que se mueve la muestra obtenida, puesto que fuera de él la verosimilitud vale cero. Tenemos que resolverlo de esta forma, debido a que la verosimilitud no tiene una forma funcional concreta.

Entonces, la gráfica de la verosimilitud obtenida para la media es:

CAPÍTULO 3: INFERENCIA ESTADÍSTICA



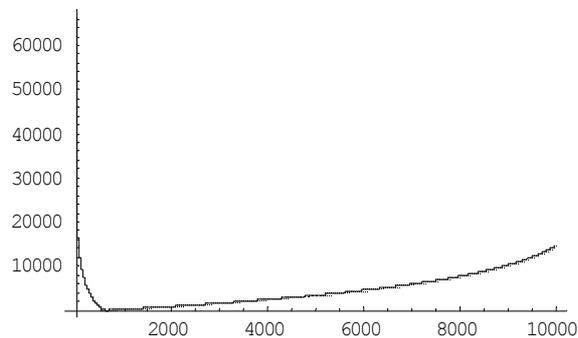
Si ampliamos la gráfica entre 600 y 700 mil pesetas de ingresos familiares trimestrales, tenemos:



Veamos las distintas aplicaciones de esta función:

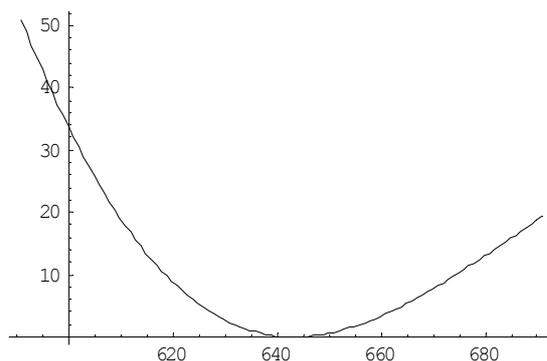
- En la teoría de la Verosimilitud Empírica se realizan todos los cálculos sobre el parámetro estudiado (la media de la variable aleatoria, en nuestro caso) a partir de esta función, a la que llama función de Verosimilitud Empírica.

A partir de ella, se obtiene la función del logaritmo de la Razón de Verosimilitud Empírica, que gráficamente es:



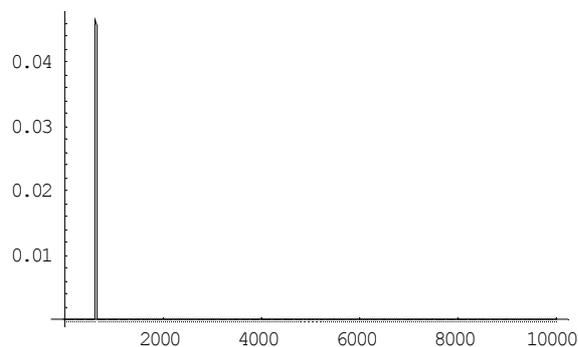
Si ampliamos la gráfica entre 600 y 700 mil pesetas de ingresos familiares trimestrales, tenemos:

APROXIMACIÓN BAYESIANA A LA METODOLOGÍA DE LA VEROSIMILITUD EMPÍRICA

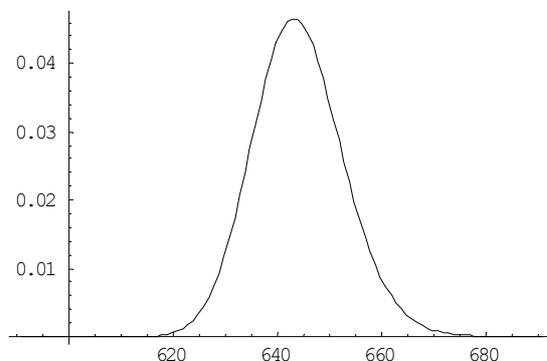


A esta función, se aplica el Teorema de Wilks para obtener resultados relativos al parámetro media de la variable aleatoria estudiada.

- En la teoría referente a la Curva de Verosimilitud más Probable, la Curva de Verosimilitud más Probable es esta función normalizada respecto del parámetro, la media, que es una verdadera función de probabilidad. A partir de ella, se obtienen probabilidades de intervalos, regiones de confianza y contrastes de hipótesis. Entonces, la curva de verosimilitud más probable en nuestro ejemplo viene dada por la siguiente gráfica:



Si ampliamos la gráfica entre 600 y 700 mil pesetas de ingresos familiares trimestrales, tenemos:



- Nosotros vamos a utilizar esta función como función de verosimilitud del problema para, tras combinarla con la distribución a priori del parámetro estudiado (la media),

obtener la distribución a posteriori del parámetro respecto de la muestra considerada, a partir de la cual vamos a obtener resultados relativos a la media de la variable aleatoria estudiada.

3.5.2. Elección de la distribución a priori del parámetro media.

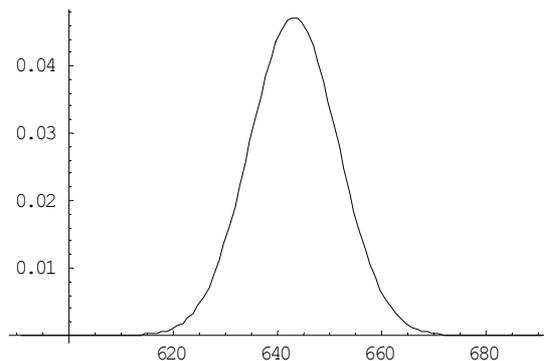
El siguiente paso va a ser la elección de la distribución a priori del parámetro media, para poder realizar el estudio bayesiano de nuestro caso, combinándola con la verosimilitud seleccionada.

A) Utilización del método de máxima verosimilitud tipo II considerando clases paramétricas usuales.

Vamos a elegir una distribución a priori procedente de cada clase paramétrica usual y, de entre todas ellas, nos quedaremos con una, según el criterio de máxima verosimilitud tipo II, estudiado en el apartado 2.3. de la memoria.

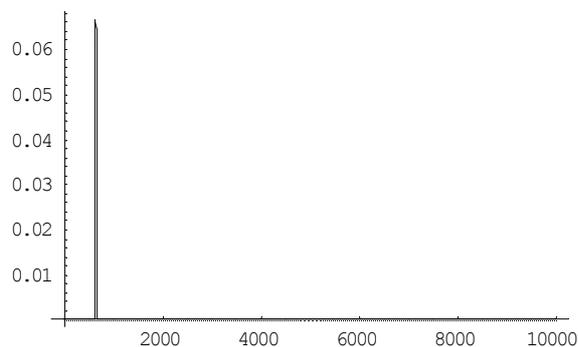
- Si elegimos como distribución a priori la distribución Uniforme en el intervalo dado por el menor y el mayor valor muestral obtenido, puesto que la media muestral está entre esos dos valores, es decir, el intervalo $[0, 9994]$ o en cualquier intervalo que lo contenga, la distribución a posteriori que obtenemos es la normalización de la función de verosimilitud utilizada respecto de la media, es decir, la Curva de Verosimilitud más Probable, ya vista con anterioridad.
- Si consideramos como clase de distribuciones a priori la formada por todas las distribuciones Normales, tenemos:

Los parámetros de la distribución, como vimos en el apartado 2.3.1., vienen dados por la media muestral y por la desviación típica muestral dividida entre la raíz cuadrada del tamaño muestral, por lo que la distribución a priori es una distribución Normal (643.213, 8.4380):

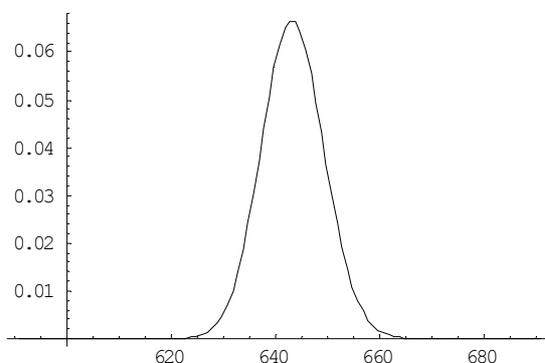


APROXIMACIÓN BAYESIANA A LA METODOLOGÍA DE LA VEROSIMILITUD EMPÍRICA

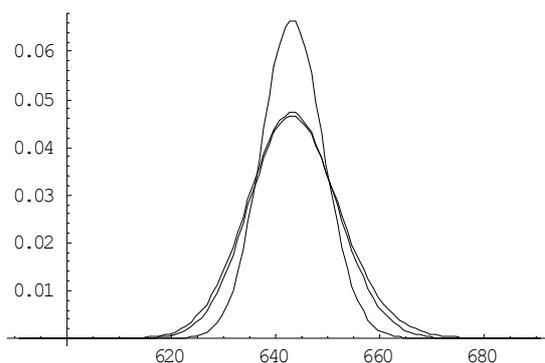
Por consiguiente, la distribución a posteriori obtenida es:



Si ampliamos la gráfica entre 600 y 700 mil pesetas de ingresos familiares trimestrales, tenemos:



En la siguiente gráfica podemos observar la verosimilitud, la distribución a priori y la distribución a posteriori entre 600 y 700 mil pesetas de ingresos familiares trimestrales, de manera que podemos comprobar cómo influye la distribución a priori utilizada respecto de la verosimilitud para la obtención de la distribución a posteriori:

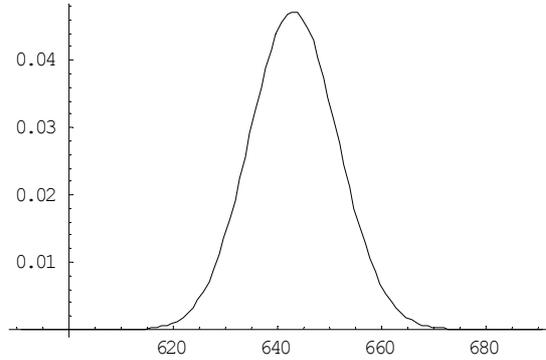


- Si consideramos como clase de distribuciones a priori la formada por todas las distribuciones Gamma, tenemos:

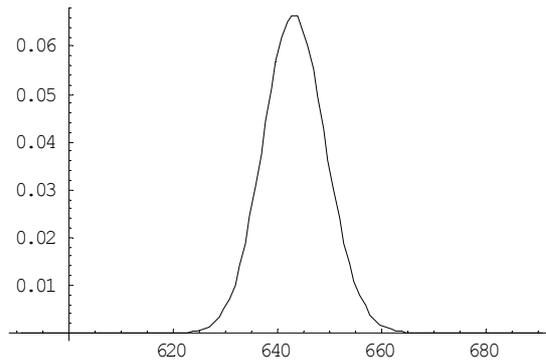
Los parámetros de la distribución, como vimos en el apartado 2.3.1., vienen dados por:

$$\left\{ \begin{array}{l} \hat{\alpha} = \frac{\bar{x}^2 \cdot n}{S_x^2} = \frac{643.213^2 \cdot 3114}{470.87^2} = 5810.66973 \\ \hat{\beta} = \frac{S_x^2}{\bar{x} \cdot n} = \frac{470.87^2}{643.213 \cdot 3114} = 0.11069 \end{array} \right.$$

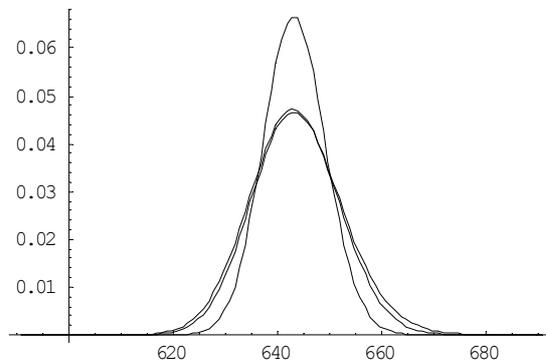
Por tanto, la distribución a priori es una Gamma (5810.66973, 0.11069).



Por consiguiente, la distribución a posteriori obtenida es:



En la siguiente gráfica podemos observar la verosimilitud, la distribución a priori y la distribución a posteriori entre 600 y 700 mil pesetas de ingresos familiares trimestrales, de manera que podemos comprobar cómo influye la distribución a priori utilizada respecto de la verosimilitud para la obtención de la distribución a posteriori:



- Si consideramos como clase de distribuciones a priori la formada por todas las distribuciones Beta, lo primero que tenemos que hacer, puesto que la distribución

APROXIMACIÓN BAYESIANA A LA METODOLOGÍA DE LA VEROSIMILITUD EMPÍRICA

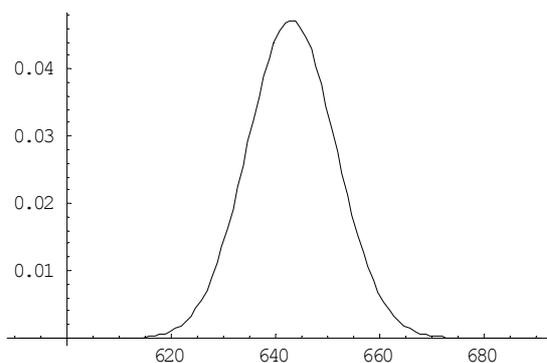
Beta varía en el intervalo [0,1], es considerar el cambio de variable $X' = X/10000$ para obtener los parámetros de la distribución.

Entonces los parámetros de la distribución Beta, como vimos en el apartado 2.3.1., vienen dados por:

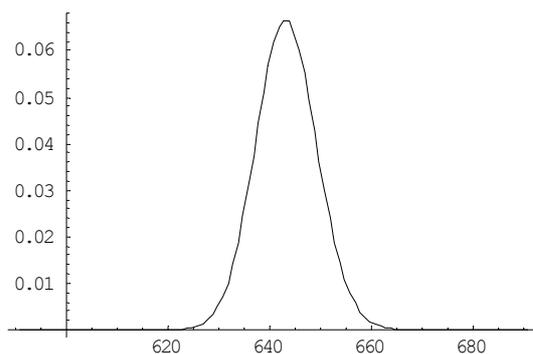
$$\left\{ \begin{aligned} \hat{\alpha} &= \frac{n \cdot \bar{x} \left(\bar{x} - \bar{x}^2 - \frac{S_x^2}{n} \right)}{S_x^2} = \\ &= \frac{3114 \cdot 0.0643213 \left(0.0643213 - 0.0643213^2 - \frac{0.047087^2}{3114} \right)}{0.047087^2} = 5436.855 \\ \hat{\beta} &= \frac{n(1 - \bar{x}) \left(\bar{x} - \bar{x}^2 - \frac{S_x^2}{n} \right)}{S_x^2} = \\ &= \frac{3114(1 - 0.0643213) \left(0.0643213 - 0.0643213^2 - \frac{0.047087^2}{3114} \right)}{0.047087^2} = 79089.66 \end{aligned} \right.$$

donde los valores muestrales se han obtenido tras realizar el cambio de variable.

Por tanto, la distribución a priori es una Beta (5436.855, 79089.66) aplicada en la variable aleatoria X' .

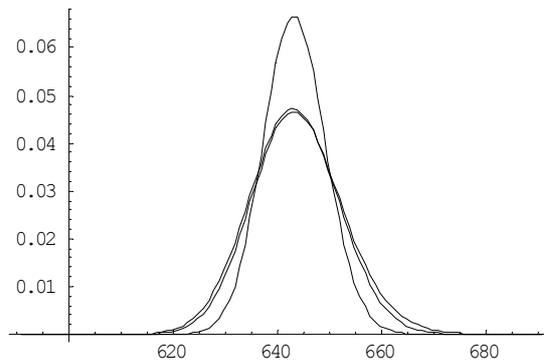


Por consiguiente, la distribución a posteriori obtenida es:



CAPÍTULO 3: INFERENCIA ESTADÍSTICA

En la siguiente gráfica podemos observar la verosimilitud, la distribución a priori y la distribución a posteriori entre 600 y 700 mil pesetas de ingresos familiares trimestrales, de manera que podemos comprobar cómo influye la distribución a priori utilizada respecto de la verosimilitud para la obtención de la distribución a posteriori:



Por tanto, tenemos 4 posibles elecciones de la distribución a priori para nuestro ejemplo. Para, de entre ellas elegir una, vamos a utilizar el criterio de máxima verosimilitud tipo II ya estudiado. Para ello, calculamos la densidad predictiva para cada distribución a priori considerada:

- Si la a priori es la Uniforme $[0, 9994]$, su densidad predictiva vale $6.2635 \cdot 10^{-8930}$.
- Si la a priori es la Normal $(643.213, 8.4380)$, su densidad predictiva vale $2.0691871 \cdot 10^{-8927}$.
- Si la a priori es la Gamma $(5810.669, 0.11069)$, su densidad predictiva vale $2.0691128 \cdot 10^{-8927}$.
- Si la a priori es la Beta $(5436.855, 79089.66)$, su densidad predictiva vale $2.0691112 \cdot 10^{-8927}$.

Como la distribución a priori con densidad predictiva mayor de entre las consideradas es la Normal $(643.213, 8.4380)$, la elegimos como distribución a priori para el estudio de a posteriori de la media de la variable aleatoria considerada.

B) Estudio de la Robustez Bayesiana del problema.

Pero, como se vio en el Capítulo 2 de la memoria, lo más adecuado es obtener la distribución a priori a partir del estudio de la Robustez Bayesiana del problema. Vamos a estudiar la robustez de la media a posteriori del parámetro que estamos estudiando, es decir, de la media de la variable aleatoria ingresos monetarios familiares en un trimestre, para lo que vamos a utilizar distintas clases de contaminación y

APROXIMACIÓN BAYESIANA A LA METODOLOGÍA DE LA VEROSIMILITUD EMPÍRICA

consideraremos varios grados de contaminación ε (0, 0.05, 0.10, 0.15, 0.20, 0.25, 0.30), es decir, vamos a modificar nuestra confianza en la distribución a priori base elegida.

- Si elegimos como clase contaminante la clase formada por todas las distribuciones de probabilidad (Γ_1), la distribución de probabilidad que vamos a elegir como distribución a priori inicial ($\pi_0(\mu)$) va a ser la distribución Normal (643.213, 8.4380), debido a que es la de máxima verosimilitud tipo II entre todas las que hemos considerado hasta ahora. Vamos a estudiar el rango de valores para la media a posteriori de la media de la variable aleatoria estudiada para distintos grados de contaminación ε , para lo cual aplicamos el teorema 2.1. del apartado 2.5.1.A:

$$\sup_{\pi \in \Gamma_1} \left(\inf_{\pi \in \Gamma_1} \right) \rho(x, \pi) = \sup_{\mu} \left(\inf_{\mu} \right) \frac{A_0 + \mu f^B(\mu)}{A + f^B(\mu)},$$

donde $A = \frac{1-\varepsilon}{\varepsilon} m(x | \pi_0) = \frac{1-\varepsilon}{\varepsilon} 2.069 \cdot 10^{-8927}$ y $A_0 = A \cdot \rho(x, \pi_0) = A \cdot 643.424$;

y además, vamos a estudiar la sensibilidad relativa:

$$R.S. = \frac{\sup_{\pi \in \Gamma} (\rho(x, \pi)) - \inf_{\pi \in \Gamma} (\rho(x, \pi))}{2 \cdot \rho(x, \pi_0)} \cdot 100,$$

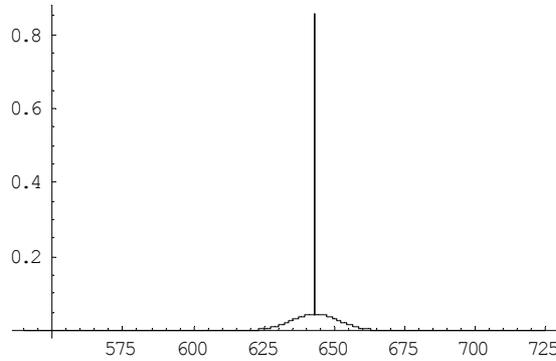
que mide la variación de la media a posteriori cuando la distribución a priori varía dentro de la clase de contaminación considerada, para los distintos grados de contaminación considerados. Estos resultados los podemos ver en la siguiente tabla:

Grado de contaminación (ε)	Rango	Sensibilidad relativa (en %)
0	643.424	0
0.05	(643.062, 643.793)	0.0567
0.10	(642.694, 644.170)	0.1147
0.15	(642.320, 644.557)	0.1738
0.20	(641.939, 644.952)	0.2341
0.25	(641.549, 645.359)	0.2960
0.30	(641.141, 645.786)	0.3609

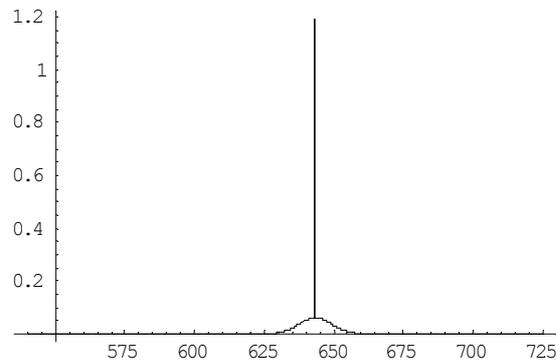
Si consideramos un grado de contaminación $\varepsilon = 0.05$, la distribución a priori de máxima verosimilitud tipo II de la clase está dada, según el teorema 2.2. por:

$$\hat{\pi}(\mu) = (1 - \varepsilon)\pi_0(\mu) + \varepsilon\hat{q}(\mu) = 0.95\pi_0(\mu) + 0.05\cdot I_{\hat{\mu}} = \begin{cases} 0.95\pi_0(\mu) & \mu \neq 643.2 \\ 0.95\pi_0(\mu) + 0.05 & \mu = 643.2 \end{cases}$$

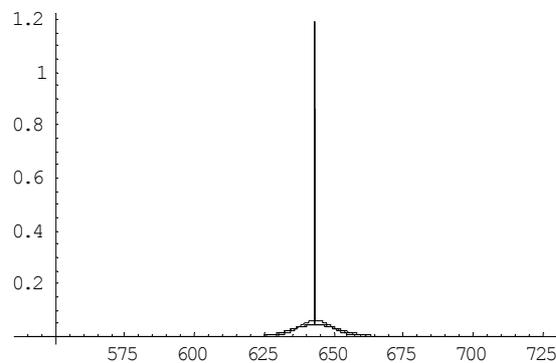
puesto que $\hat{\mu} = 643.2$ es el estimador de máxima verosimilitud de la media poblacional.



Y la distribución a posteriori obtenida es:



En la siguiente gráfica podemos observar la verosimilitud, la distribución a priori y la distribución a posteriori, de manera que podemos comprobar cómo influye la distribución a priori utilizada respecto de la verosimilitud para la obtención de la distribución a posteriori:



- Como pretendemos mejorar la sensibilidad relativa, y sabemos por el estudio en años anteriores que la media de la variable ingresos monetarios trimestrales de las familias es unimodal con moda próxima a la media muestral, consideramos que la

APROXIMACIÓN BAYESIANA A LA METODOLOGÍA DE LA VEROSIMILITUD EMPÍRICA

distribución a priori es unimodal, con moda próxima al valor 643.213, por lo que elegimos una clase contaminante un poco más restringida, la clase formada por todas las distribuciones unimodales con la misma moda θ_0 que π_0 (Γ_2), donde la distribución de probabilidad que vamos a elegir como distribución a priori inicial ($\pi_0(\mu)$) va a ser la distribución Normal de media 643.213 y desviación típica dada por la desviación típica muestral de la media, es decir, la distribución Normal (643.213, 8.4380), puesto que pertenece a la clase y es una distribución de probabilidad bastante representativa de la clase que estamos utilizando, como ya hemos estudiado. Como la moda de π_0 (μ) es 643.213, la clase incluye todas las distribuciones de probabilidad con moda 643.213. En la siguiente tabla vamos a observar el rango de valores para la media a posteriori de la media de la variable aleatoria estudiada para distintos grados de contaminación considerados, que viene dado por el teorema 2.3. del apartado 2.5.1.B:

$$\sup_{\pi \in \Gamma_2} \left(\inf_{\pi \in \Gamma_2} \right) \rho(x, \pi) = \sup_z \left(\inf_z \right) \frac{A_0 + H^h(z)}{A + H^0(z)},$$

donde $A = \frac{1-\varepsilon}{\varepsilon} m(x | \pi_0) = \frac{1-\varepsilon}{\varepsilon} 2.069 \cdot 10^{-8927}$, $A_0 = A \cdot \rho(x, \pi_0) = A \cdot 643.424$ y

$$H^0(z) = \begin{cases} \frac{1}{z} \int_{643.213}^{643.213+z} f^B(\mu) d\mu & z \neq 0 \\ f^B(643.213) & z = 0 \end{cases},$$

$$H^h(z) = \begin{cases} \frac{1}{z} \int_{643.213}^{643.213+z} \mu f^B(\mu) d\mu & z \neq 0; \\ 643.213 f^B(643.213) & z = 0 \end{cases};$$

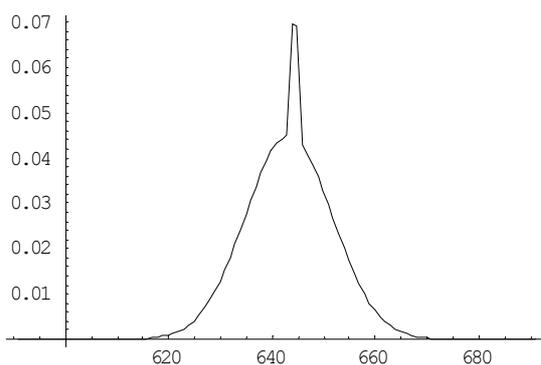
y también observaremos su correspondiente sensibilidad relativa:

Grado de contaminación (ε)	Rango	Sensibilidad relativa (en %)
0	643.424	0
0.05	(643.135, 643.695)	0.0435
0.10	(642.849, 643.965)	0.0867
0.15	(642.565, 644.234)	0.1297
0.20	(642.282, 644.504)	0.1726
0.25	(642.001, 644.775)	0.2155
0.30	(641.719, 645.049)	0.2587

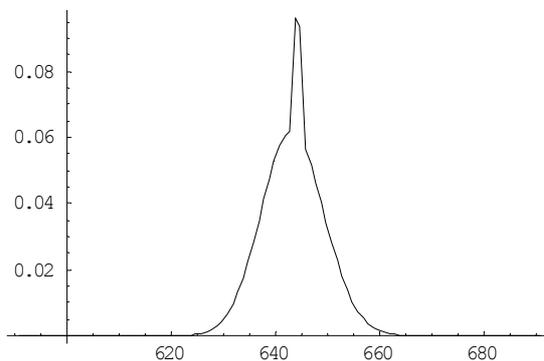
CAPÍTULO 3: INFERENCIA ESTADÍSTICA

Si consideramos un grado de contaminación $\varepsilon=0.05$, utilizando el teorema 2.4., la distribución a priori de máxima verosimilitud tipo II de la clase, como \hat{q} sigue una distribución Uniforme en el intervalo (643.213, 644.213), puesto que $a=1$ es el valor que maximiza la función $m(x|a)$, está dada por:

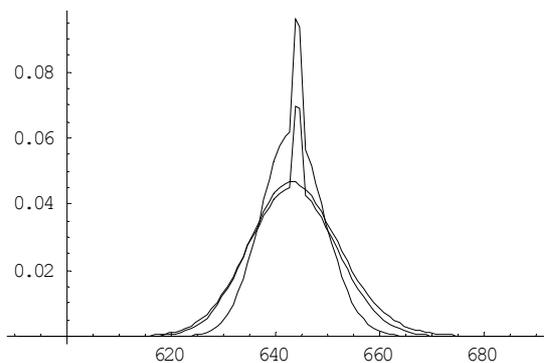
$$\hat{\pi}(\mu) = 0.95\pi_0(\mu) + 0.05\hat{q}(\mu) = \begin{cases} 0.95\pi_0(\mu) & \mu \notin (643.213, 644.213) \\ 0.95\pi_0(\mu) + 0.05\frac{1}{1} & \mu \in (643.213, 644.213) \end{cases}$$



Y la distribución a posteriori obtenida es:



En la siguiente gráfica podemos observar la verosimilitud, la distribución a priori y la distribución a posteriori, de manera que podemos comprobar cómo influye la distribución a priori utilizada respecto de la verosimilitud para la obtención de la distribución a posteriori:



APROXIMACIÓN BAYESIANA A LA METODOLOGÍA DE LA VEROSIMILITUD EMPÍRICA

- De nuevo, pretendemos mejorar la sensibilidad relativa y, como sabemos por el estudio en años anteriores que la media de la variable ingresos monetarios trimestrales de las familias es, además de unimodal, simétrica respecto de su media, introducimos esta nueva restricción en la clase, por lo que tenemos la clase formada por todas las distribuciones simétricas y unimodales con la misma moda θ_0 que π_0 (Γ_4). La distribución de probabilidad que vamos a elegir como distribución a priori inicial ($\pi_0(\mu)$) va a ser, de nuevo, la distribución Normal (643.213, 8.4380), puesto que está en la clase y es representativa de ella. Como la moda de $\pi_0(\mu)$ es 643.213, la clase incluye todas las distribuciones de probabilidad simétricas con moda 643.213. En la siguiente tabla vamos a observar el rango de valores para la media a posteriori de la media de la variable aleatoria estudiada para distintos grados de contaminación considerados, que viene dado por el teorema 2.7. del apartado 2.5.1.D:

$$\sup_{\pi \in \Gamma_4} \left(\inf_{\pi \in \Gamma_4} \right) \rho(x, \pi) = \sup_z \left(\inf_z \right) \frac{A_0 + H^h(z)}{A + H^0(z)},$$

donde $A = \frac{1-\varepsilon}{\varepsilon} m(x | \pi_0) = \frac{1-\varepsilon}{\varepsilon} 2.069 \cdot 10^{-8927}$, $A_0 = A \cdot \rho(x, \pi_0) = A \cdot 643.424$ y

$$H^0(z) = \begin{cases} \frac{1}{2 \cdot z} \int_{643.213-z}^{643.213+z} f^B(\mu) d\mu & z \neq 0 \\ f^B(643.213) & z = 0 \end{cases}$$

$$H^h(z) = \begin{cases} \frac{1}{2 \cdot z} \int_{643.213-z}^{643.213+z} \mu f^B(\mu) d\mu & z \neq 0 \\ 643.213 f^B(643.213) & z = 0 \end{cases};$$

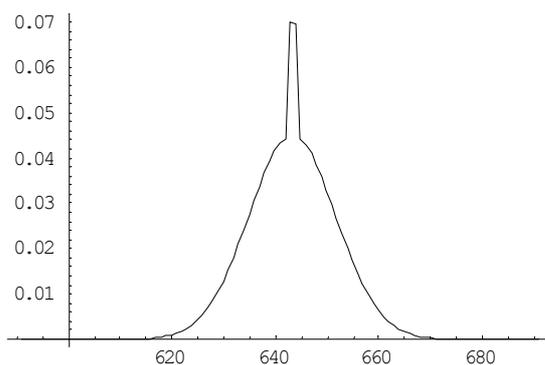
y también observaremos su correspondiente sensibilidad relativa:

Grado de contaminación (ε)	Rango	Sensibilidad relativa (en %)
0	643.424	0
0.05	(643.367,643.442)	0.0058
0.10	(643.315,643.460)	0.0112
0.15	(643.269,643.479)	0.0163
0.20	(643.226,643.499)	0.0211
0.25	(643.187,643.520)	0.0258
0.30	(643.152,643.542)	0.0303

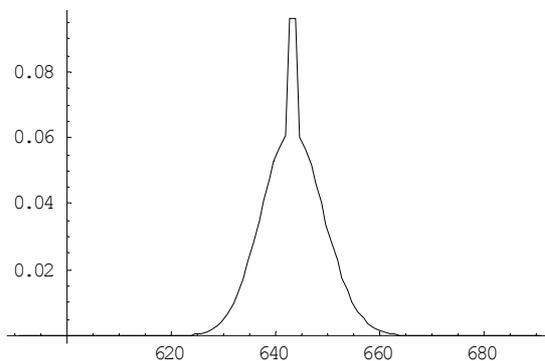
CAPÍTULO 3: INFERENCIA ESTADÍSTICA

Si consideramos un grado de contaminación $\varepsilon=0.05$, utilizando el teorema 2.8., la distribución a priori de máxima verosimilitud tipo II de la clase, como \hat{q} sigue una distribución Uniforme en el intervalo (642.13, 644.13), puesto que $a=1$ es el valor que maximiza la función $m(x|a)$, está dada por:

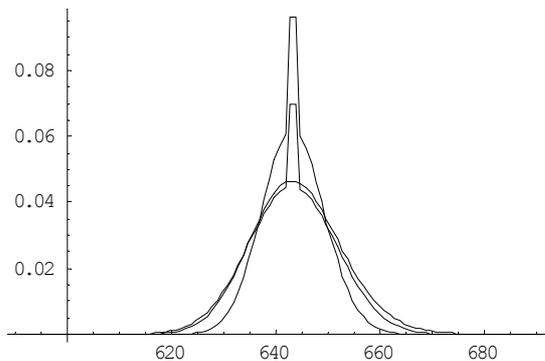
$$\hat{\pi}(\mu) = 0.95\pi_0(\mu) + 0.05\cdot\hat{q}(\mu) = \begin{cases} 0.95\pi_0(\mu) & \mu \notin (642.213,644.213) \\ 0.95\pi_0(\mu) + 0.05\frac{1}{2} & \mu \in (642.213,644.213) \end{cases}$$



Y la distribución a posteriori obtenida es:



En la siguiente gráfica podemos observar la verosimilitud, la distribución a priori y la distribución a posteriori, de manera que podemos comprobar cómo influye la distribución a priori utilizada respecto de la verosimilitud para la obtención de la distribución a posteriori:



APROXIMACIÓN BAYESIANA A LA METODOLOGÍA DE LA VEROSIMILITUD EMPÍRICA

Como la sensibilidad relativa obtenida en esta última clase de contaminación es suficientemente pequeña, podemos considerar adecuada esta clase contaminante para la distribución a priori, por lo que utilizaremos la distribución a priori de máxima verosimilitud de tipo II de dicha clase para, combinada con la verosimilitud, obtener las conclusiones a posteriori del caso estudiado.

3.5.3. Obtención de conclusiones a posteriori.

Pasamos a obtener conclusiones a posteriori sobre el caso estudiado. Utilizamos, por supuesto, la verosimilitud elegida anteriormente, combinada con la distribución a priori seleccionada. Aunque las conclusiones las vamos a obtener a partir de la clase contaminante elegida, vamos a comparar los resultados con el resto de distribuciones a priori consideradas (en las distintas clases contaminantes los resultados se obtienen a partir de la distribución a priori de máxima verosimilitud de tipo II de la clase), para comparar los resultados.

A) Estimación puntual.

Veamos en el siguiente cuadro distintas estimaciones puntuales para la media en nuestro ejemplo, utilizando las distintas distribuciones a priori ya estudiadas con anterioridad. Indicar que los estimadores puntuales de la función de verosimilitud coinciden con los estimadores puntuales de la distribución a posteriori cuando la distribución a priori es la Uniforme, puesto que, salvo constantes normalizadoras, son la misma distribución.

	<i>Algunas características de la distribución a priori</i>			<i>Algunas características de la distribución a posteriori</i>		
<i>Distribución a priori</i>	<i>Moda</i>	<i>Mediana</i>	<i>Media</i>	<i>Moda</i>	<i>Mediana</i>	<i>Media</i>
<i>Uniforme</i>	-	4997.5	4997.5	643.2	643.3	644.13
<i>Normal</i>	643.213	643.213	643.213	643.2	642.9	643.424
<i>Gamma</i>	643.3	642.7	643.213	643.2	642.8	643.397
<i>Beta</i>	643.3	642.7	643.213	643.2	642.8	643.399
<i>(1)</i>	643.213	643.213	643.213	643.2	643.2	643.411
<i>(2)</i>	643.213	643.623	643.223	643.2	643.4	643.455
<i>(3)</i>	643.213	643.213	643.213	643.2	642.9	643.419

CAPÍTULO 3: INFERENCIA ESTADÍSTICA

- (1) es la distribución a priori de máxima verosimilitud tipo II de la clase de contaminación con clase contaminante formada por todas las distribuciones de probabilidad.
- (2) es la distribución a priori de máxima verosimilitud tipo II de la clase de contaminación con clase contaminante formada por todas las distribuciones simétricas y unimodales con la misma moda que π_0 .
- (3) es la distribución a priori de máxima verosimilitud tipo II de la clase de contaminación con clase contaminante formada por todas las distribuciones unimodales con la misma moda que π_0 .

Como se puede ver, las principales características de la distribución a posteriori dependen de la distribución a priori utilizada, puesto que sus valores varían según la distribución a priori considerada.

B) Regiones de confianza.

Veamos en el siguiente cuadro distintas regiones de confianza HPD obtenidas para la media en nuestro ejemplo, utilizando las distintas distribuciones a priori consideradas con distintos niveles de confianza:

<i>Distribución a priori</i>	<i>Regiones de confianza para la media con nivel de confianza</i>		
	<i>0.9</i>	<i>0.95</i>	<i>0.99</i>
<i>Uniforme</i>	(629.83,658.0)	(627.4,661.4)	(622.80,667.88)
<i>Normal</i>	(633.3,653.4)	(631.4,655.4)	(628.0,658.8)
<i>Gamma</i>	(633.5,653.1)	(631.5,655.0)	(628.1,659.4)
<i>Beta</i>	(633.5,653.1)	(632.0,654.5)	(628.2,659.4)
<i>(1)</i>	(633.8,653.0)	(631.8,655.0)	(628.2,658.9)
<i>(2)</i>	(633.8,653.0)	(631.8,655.0)	(628.2,658.9)
<i>(3)</i>	(633.8,653.0)	(631.8,655.0)	(628.2,658.9)

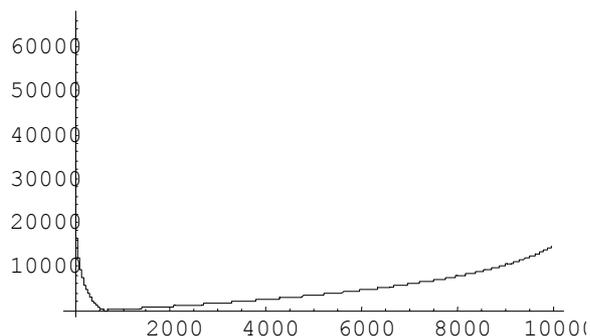
- (1) es la distribución a priori de máxima verosimilitud tipo II de la clase de contaminación con clase contaminante formada por todas las distribuciones de probabilidad.

APROXIMACIÓN BAYESIANA A LA METODOLOGÍA DE LA VEROSIMILITUD EMPÍRICA

- (2) es la distribución a priori de máxima verosimilitud tipo II de la clase de contaminación con clase contaminante formada por todas las distribuciones simétricas y unimodales con la misma moda que π_0 .
- (3) es la distribución a priori de máxima verosimilitud tipo II de la clase de contaminación con clase contaminante formada por todas las distribuciones unimodales con la misma moda que π_0 .

Podemos ver que las distintas regiones de confianza van variando, debido a las distintas elecciones de la distribución a priori, puesto que sus valores varían según la distribución a priori considerada.

Si la distribución a priori es la uniforme y aplicamos el método de la Verosimilitud Empírica para obtener la región de confianza para la media, utilizaremos el teorema de Wilks. Para ello, obtenemos la función $l(\mu)$, que gráficamente viene dada por:



Entonces, las regiones de confianza vienen dadas por:

- Para un nivel de confianza del 90%: (630.0, 658.0).
Para un nivel de confianza del 95%: (628.1, 661.1).
Para un nivel de confianza del 99%: (622.9, 667.7).

Comparando con los resultados obtenidos para las HPD considerando la distribución uniforme como distribución a priori, podemos observar que tenemos una buena aproximación de las regiones de confianza.

C) Contraste de hipótesis.

Queremos contrastar la hipótesis de que la media de los ingresos familiares trimestrales medios es 650000 pesetas. Para ello, consideramos las siguientes hipótesis nula y alternativa:

$$H_0: \mu = 650$$

$$H_1: \mu \neq 650$$

1. Si consideramos como distribución a priori la Uniforme:

- Podemos utilizar la técnica de la Verosimilitud Empírica:

Como el valor 650 está dentro de la región de confianza dada por el teorema de Wilks para la media con nivel de significación del 5%, no rechazamos la hipótesis nula.

- Utilizando la distribución a posteriori obtenida, tenemos dos opciones:

a) Considerar que todos los valores entre 640 y 660 son equivalentes a 650, por lo que la hipótesis nula a considerar será el intervalo (640,660).

Entonces, comparamos las probabilidades a posteriori de las hipótesis nula y alternativa para decidir sobre el contraste:

$$\alpha_0 = P[\Theta_0 | x] = P[\mu \in (640,660)] = 0.621$$

$$\alpha_1 = P[\Theta_1 | x] = 0.379,$$

por lo que no rechazamos la hipótesis nula estudiada.

Además, como la distribución a priori es una uniforme en el intervalo (0,9994), las probabilidades a priori de las hipótesis nula y alternativa son:

$$\pi_0 = \frac{655 - 645}{9994} = 0.001$$

$$\pi_1 = 0.999$$

Si consideramos el factor Bayes:

$$B = \frac{0.621/0.379}{0.001/0.999} = \frac{1.638}{0.001} = 1638,$$

podemos observar que los datos muestrales han hecho que la probabilidad de la hipótesis nula aumente bastante respecto a lo que pensábamos a priori.

b) Si asignamos una probabilidad a priori de 0.3 a la hipótesis nula, tenemos la

siguiente distribución a priori: $\hat{\pi}(\mu) = \begin{cases} 0.3 & \mu = 650 \\ 0.7 \cdot \pi(\mu) & \mu \neq 650 \end{cases}$.

Entonces, las probabilidades a posteriori son:

$$\alpha_0 = P[\Theta_0 | x] = P[\mu = 650] = 0.993.$$

$$\alpha_1 = P[\Theta_1 | x] = P[\mu \neq 650] = 0.007.$$

Por tanto, aceptamos la hipótesis nula estudiada.

APROXIMACIÓN BAYESIANA A LA METODOLOGÍA DE LA VEROSIMILITUD EMPÍRICA

Si consideramos el factor Bayes:

$$B = \frac{0.993/0.007}{0.3/0.7} = \frac{141.857}{0.42857} = 331.001,$$

podemos observar que los datos muestrales han hecho que la probabilidad de la hipótesis nula aumente bastante respecto a lo que pensábamos a priori.

2. Si consideramos como distribución a priori la de máxima verosimilitud tipo II de la clase de contaminación seleccionada, es decir, de la clase formada por todas las distribuciones simétricas y unimodales con moda 643.123.

Tenemos dos opciones de trabajo:

- Considerar que todos los valores entre 640 y 660 son equivalentes a 650, por lo que la hipótesis nula a considerar será el intervalo (640,660).

Entonces, comparamos las probabilidades a posteriori de las hipótesis nula y alternativa para decidir sobre el contraste:

$$\alpha_0 = P[\Theta_0 | x] = P[\mu \in (640,660)] = 0.7121$$

$$\alpha_1 = P[\Theta_1 | x] = 0.2879,$$

por lo que no rechazamos la hipótesis nula estudiada.

Además, las probabilidades a priori de las hipótesis nula y alternativa son:

$$\pi_0 = 0.6311$$

$$\pi_1 = 0.3689$$

Si consideramos el factor Bayes:

$$B = \frac{0.7121/0.2879}{0.6311/0.3689} = \frac{2.4734}{1.7107} = 1.445,$$

podemos observar que los datos muestrales han hecho que la probabilidad de la hipótesis nula aumente respecto a lo que pensábamos a priori.

- Si asignamos una probabilidad a priori de 0.3 a la hipótesis nula, tenemos la

siguiente distribución a priori: $\hat{\pi}(\mu) = \begin{cases} 0.3 & \mu = 650 \\ 0.7 \cdot \pi(\mu) & \mu \neq 650 \end{cases}$.

Entonces, las probabilidades a posteriori son:

$$\alpha_0 = P[\Theta_0 | x] = P[\mu = 650] = 0.3302.$$

$$\alpha_1 = P[\Theta_1 | x] = P[\mu \neq 650] = 0.6698.$$

Por tanto, rechazamos la hipótesis nula estudiada.

Si consideramos el factor Bayes:

$$B = \frac{0.3302/0.6698}{0.3/0.7} = \frac{0.4929}{0.42857} = 1.1501,$$

podemos observar que los datos muestrales han hecho que la probabilidad de la hipótesis nula aumente respecto a lo que pensábamos a priori.

Capítulo 4:

Conclusiones y líneas de investigación futura.

En este capítulo vamos a enumerar los resultados más relevantes de nuestra investigación y las que consideramos líneas abiertas de investigación que deberán guiar el futuro trabajo investigador de este doctorando.

El aspecto más relevante a destacar de la memoria es la presentación de un nuevo método de trabajo en el caso del estudio del parámetro valor medio de una variable aleatoria observable. Para el estudio de este parámetro proponemos inicialmente una aproximación bayesiana de trabajo, puesto que consideramos que en este marco podemos incluir toda la información disponible por el investigador.

Proponemos la utilización, como función de verosimilitud del problema, de una generalización de la función que, en el acervo de la literatura estadística usual, recibe el nombre de función de verosimilitud empírica para la media, y para la cual se cuenta con un gran número de resultados de interés. Esta función tiene la particularidad de depender exclusivamente de los datos muestrales obtenidos y no de ninguna decisión subjetiva realizada por el investigador. Además, probamos que la función de verosimilitud empírica para la media toma los mismos valores para cada valor del parámetro que la verosimilitud modificada para la media, también desarrollada en la

CAPÍTULO 4: CONCLUSIONES Y LÍNEAS ABIERTAS

literatura, cuando trabajamos con espacios paramétricos discretos. En el núcleo central de esta memoria se demuestra que la función de verosimilitud empírica y, por tanto, la verosimilitud que proponemos para su utilización, es continua en un espacio paramétrico continuo, lo cual nos va a permitir la combinación de ésta con cualquier distribución a priori continua mediante la utilización del teorema de Bayes, sin demasiadas complicaciones, en un espacio paramétrico continuo para la media y, además, nos va a permitir aplicar todos los resultados relativos a la teoría de la robustez bayesiana desarrollados en éste ámbito. También cabe señalar aquí que la teoría de la verosimilitud empírica es en realidad un caso particular del método propuesto en esta memoria, siempre que consideremos como distribución a priori del parámetro media a la distribución Uniforme en un intervalo de valores que contenga a la muestra.

La elección de la distribución a priori respecto del parámetro objeto de estudio es uno de los aspectos más conflictivos en un estudio bayesiano, puesto que va a depender de la opinión del investigador. Por esta razón, nosotros proponemos un método de trabajo lo más objetivo posible para su elección. Así las cosas, si el investigador conoce la distribución a priori del parámetro por las razones que sean, puede trabajar directamente con ella. En el caso en que el investigador conoce la distribución, pero desconoce los parámetros, estima éstos y trabaja con la distribución obtenida. Sin embargo, y por lo general, la distribución a priori va a ser desconocida, y solo vamos a conocer aspectos generales de ella (la moda, algunos cuantiles, su simetría...) por lo que consideraremos que la distribución a priori está en una clase, que contiene todas las distribuciones de probabilidad que cumplen los requisitos para ser la distribución a priori del parámetro estudiado. Estas clases así definidas pueden ser utilizadas en virtud de la continuidad de la verosimilitud utilizada. A continuación se centra nuestra atención en efectuar un amplio repaso por todas las clases analizadas en la literatura, estudiando su robustez bayesiana, y llegando a la conclusión de que existe una mayor más robustez cuanto más restrictiva es la clase de contaminación considerada. Si la clase elegida es suficientemente robusta, no se producen cambios importantes en los resultados a posteriori obtenidos al cambiar la distribución a priori dentro de la clase, por lo que podemos decantarnos por trabajar con una distribución representativa de la clase, preferentemente la distribución de máxima verosimilitud tipo II de dicha clase, para así obtener los resultados a posteriori relativos al parámetro media estudiado. Puesto que estamos utilizando un método en esencia bayesiano,

APROXIMACIÓN BAYESIANA A LA METODOLOGÍA DE LA VEROSIMILITUD EMPÍRICA

podemos utilizar todos los resultados relativos a la inferencia bayesiana disponible en la literatura, sin demasiadas dificultades y obtener así las conclusiones a posteriori relativas al parámetro de interés.

En resumen, hemos introducido una metodología de trabajo novedosa para realizar un estudio bayesiano exhaustivo relativo a un parámetro, de manera que, en cada etapa, todos los pasos que vamos dando van a depender fundamentalmente de la muestra obtenida, reduciendo al mínimo la subjetividad que puede introducir el investigador. Además, hemos incorporado a nuestro método, como un caso particular, toda la teoría de la verosimilitud empírica para la media.

Como posibles líneas de investigación abiertas para continuar con nuestro estudio puede considerarse la utilización de nuestro método de trabajo en diversos campos, como Medicina, Economía, etcétera. Podemos considerar también la posibilidad de cambiar el parámetro objeto de estudio, por ejemplo, podemos estudiar la varianza; incluso pasando al caso multidimensional (media y varianza, por ejemplo), para lo cual también podríamos basarnos en la función de verosimilitud empírica para dicho/s parámetro/s.

Otra línea de investigación que pudiera resultar de interés sería el tema de la derivabilidad de la verosimilitud que hemos propuesto, pues, de ser cierta, nos permitiría la utilización de más clases de distribuciones a priori y la aplicación de los conceptos relativos a Robustez Local, para los cuales la literatura ofrece multitud de resultados de interés.

Por otra parte, podríamos continuar estudiando la posibilidad de utilización de otras clases de distribuciones a priori, así como la investigación sobre la simplificación de los resultados de algunas de las ya existentes, o en el estudio de nuevas clases.

En un futuro inmediato nos proponemos también la tarea de trabajar en la mejora del programa informático presentado, para que su utilización sea más sencilla por parte de cualquier usuario, quizás introduciendo un formato interactivo mediante el cual el programa formula una pregunta al usuario que este va respondiendo a medida que el programa precisa de aquella información supuestamente conocida por el usuario, así como completando su eficacia mediante la adición de módulos con más clases de distribuciones a priori.

Apéndice:

Vamos a desarrollar el programa informático utilizado para la aplicación del método desarrollado en la memoria. Este programa ha sido desarrollado por el autor de la memoria y el profesor Nicanor Guerra Quintana de la Universidad de Las Palmas de Gran Canaria.

A.1. Obtención de la verosimilitud del problema, elección de la distribución a priori y obtención de la distribución a posteriori.

En primer lugar, vamos a introducir los datos, elegir la distribución a priori y obtener tanto la verosimilitud del problema como la distribución a posteriori tras combinar la verosimilitud con la distribución a priori elegida.

A.1.1. Introducción de datos y cálculos iniciales.

```
(*DEFINICIONES INICIALES*)  
k=3114; (*FIJAR AQUI EL NUMERO DE VALORES DISTINTOS DETECTADOS*)  
Array[x,k];  
Array[x2,k];  
Array[n,k];  
(*Array[p,k];*)  
ptos=10001; (*FIJAR EL NUMERO DE PUNTOS UTILIZADOS*)  
Array[like,ptos-1];  
Array[thetas,ptos];
```

```

Array[ppostacum,ptos-1];
Array[ppostnor,ptos-1];
Array[percentil,25];
Array[prob,25];
Array[priori,ptos-1];
Array[antes,{5,ptos-1}];
Array[despues,{5,ptos-1}];
entera[x_]=If[x<0,Ceiling[x],Floor[x]]*1.0;

```

(*INTRODUCCION Y ORDENACION DE LOS DATOS*)

```

(*INTRODUCCION DE LOS DATOS DESDE UN FICHERO*)
fichdatos=OpenRead["c:\\imonet.txt"]; (*LOCALIZACION DEL FICHERO*)
i=1;
salida=0;
datomaximo=0;
While[salida==0,
  {dato[i]=Read[fichdatos,RecordSeparators->{"\n"}];
  If[datomaximo<dato[i],datomaximo=dato[i];
  If[dato[i]==EndOfFile,{Print["final"];salida=1;]];
  i=i+1;
  }];
Close[fichdatos]
nobs=i-2;
Print["Numero de observaciones= ",nobs];

```

```

(*DATOS REESCALADOS Y TRUNCADOS A ENTEROS*)
Do[dato[i]=entera[dato[i]/1000},{i,1,nobs}];
datomaximo=entera[datomaximo/1000];

```

```

(*ORDENACION DE LOS DATOS*)
Sum[dato[i},{i,1,nobs}]
valores=Table[0,{i,1,nobs}];
Do[valores[[i]]=dato[i},{i,1,nobs}];
valoresordenados=Sort[valores];
Do[dato[i]=valoresordenados[[i]},{i,1,nobs}];
Sum[dato[i},{i,1,nobs}]

```

(*CALCULO DE LOS x[i] Y SUS FRECUENCIAS n[i]*)

```

Do[{x[i] = 0; n[i] = 0}, {i, 1, k}];
nk = 1;
x[1] = dato[1];
n[1] = 1;
rep = 0;
Do[{If[dato[i] == x[i - rep - 1],
  {n[i - rep - 1] ++;
  rep++;
  } ,
  {x[i - rep] = dato[i];
  n[i - rep] = 1;
  }
  }], {i, 2, nobs}];

```

```
i = i - 1;
Print[i, " ", i - rep - 1, " x[" , i - rep - 1, "]=" , x[i - rep - 1]];
k = nobs - rep;
nk = k;
a = x[1] - 9 (*VALOR DONDE COMENZAMOS EL ESTUDIO*)
xj = (110 / 100) datomaximo;
b = datomaximo + 9 (*VALOR DONDE TERMINAMOS EL ESTUDIO*)
k (*NUMERO DE VALORES MUESTRALES DISTINTOS*)

(*CALCULOS INICIALES*)

(*TAMAÑO MUESTRAL*)
nt = Sum[n[i], {i, 1, k}]
(*MEDIA MUESTRAL*)
cptest = Sum[n[i] * x[i], {i, 1, k}];
cptest = cptest / nt
scalef = entera[N[-Sum[n[i] * Log[n[i] / nt], {i, 1, k}]]];
(*DESVIACION TIPICA MUESTRAL*)
desvtip = Sqrt[Sum[n[i] * (x[i] - cptest)^2, {i, 1, k}] / nt]
(*CUASIDESVIACION TIPICA MUESTRAL*)
cuasidesvtip = Sqrt[Sum[n[i] * (x[i] - cptest)^2, {i, 1, k}] / (nt - 1)]
(*OBTENCION DEL MINIMO Y MAXIMO MUESTRALES*)
xmax = x[1];
xmin = x[1];
Do[{xmax = Max[xmax, x[i + 1]],
    xmin = Min[xmin, x[i + 1]]},
    {i, 1, k - 1}];
Print[xmin, " ", xmax];
Do[x2[i] = x[i] - cptest, {i, 1, k}]
xmin2 = xmin - cptest;
xmax2 = xmax - cptest;
```

A.1.2. Elección de la distribución a priori.

```
<<Statistics`ContinuousDistributions`
(*PARAMETROS DE LA DISTRIBUCION*)
alpha = 643.213
beta = 8.43804
(*DISTRIBUCION A PRIORI*)
g[x_] := PDF[NormalDistribution[alpha, beta], x]
NIntegrate[g[x], {x, a, b}]
(*GRAFICA DE LA DISTRIBUCION*)
Plot[g[x], {x, a, b}, PlotRange -> All]

(*CALCULO DE LA A PRIORI EN PTOS EQUIDISTANTES DEL ESPACIO
PARAMETRICO*)
pasada = 1;
sumprior = 0;
incretheta = (b - a) / (ptos - 2);
theta = a - incretheta;
Do[{theta = incretheta + theta;
    thetas[i] = theta;
    paso = g[theta];
    priori[i] = paso;
    If[Mod[i, 50] == 0, Print[i]];
```

```

}
,{i,1,ptos-1}};

(* LISTADO DE LOS VALORES DE theta Y DE LA PRIORI*)
Do[Print[i," theta=",N[thetas[i]]," priori[i]=",
priori[i]],{i,1,ptos-1}]
Sum[priori[i]*incretheta,{i,1,ptos-1}]

(*MEDIA Y DESVIACION TIPICA DE LA A PRIORI*)
medpriori =  $\sum_{i=1}^{ptos-1} \text{thetas}[i] * \text{priori}[i] * \text{incretheta}$ 
desvtipriori =  $\sqrt{\sum_{i=1}^{ptos-1} (\text{thetas}[i] - \text{medpriori})^2 * \text{priori}[i] * \text{incretheta}}$ 

(*FUNCION DE DISTRIBUCION DE LA A PRIORI*)
prioriacum[1] = 0;
Do[prioriacum[i] = prioriacum[i-1] + priori[i] * incretheta, {i, 2, ptos-1}]

```

A.1.3. Cálculo de la verosimilitud en los valores de theta calculados y la distribución a posteriori en esos puntos.

```

(*BUCLE VARIANDO theta PARA CALCULAR EL MAXIMO DE LA
VEROSIMILITUD, xj ESTA FIJO*)
jj=1;
f[r_] := theta*nt/(1+theta*r) - Sum[n[i]*x2[i]/(1+r*x2[i]),
{i,1,k}];
fprima[r_] := -nt*theta^2/(1+theta*r)^2 +
Sum[n[i]*x2[i]^2/(1+r*x2[i])^2,{i,1,k}];
(*VIENDO r COMO r[theta], LA PENDIENTE (DERIVADA) DE r
CON RESPECTO A theta es acel[r_], QUE SE UTILIZARA
PARA IR ACTUALIZANDO A CADA PASO EL VALOR INICIAL DE r*)
acel[r_] := (-nt/(1+theta*r)^2)/fprima[r];
sumpost=0;
Tolr=.00001;
Tolr001=.001
epsilon=.00000001;
r1:=-1/xmax2+epsilon;
r2:=-1/xmin2-epsilon;
rop=r2-epsilon;
r1aux=r1;
r2aux=r2;
cocientepaso=Table[N[n[i]/(x[i]-xj)],{i,1,nk}];
p=Table[0,{i,1,nk}];
sumacocientepaso=Sum[cocientepaso[[i]],{i,1,nk}];
Print["sumcoc=",sumacocientepaso];
Do[{theta=thetas[jj];
prior=priori[jj];
a2=nt/(theta-xj);
a1=-xj*a2;
p=(1/a2)*cocientepaso;
pj=1-sumacocientepaso/a2;

```

APROXIMACIÓN BAYESIANA A LA METODOLOGÍA DE LA VEROSIMILITUD EMPÍRICA

```
(*SI pj<0, ES IMPOSIBLE LUEGO pj=0 Y SOLO INTERVIENEN LOS
p_i MUESTRALES*)
If[pj<0, (*1 IF*)
  {theta=theta-cptest;
  contador1=0;
  While[Abs[f[rop]]>Tolr && contador1<20,      (*WHILE 1*)
    {incre=-f[rop]/fprima[rop];
    increaux=incre;
    contador1++;
    ropaux=rop;
    rop=rop+incre;
    contador2=1;
    While[(rop>r2 || rop<r1) && contador2<4,    (*WHILE 2*)
      {incre=increaux/contador2;
      rop=ropaux+incre;
      contador2++;
      }];(*CIERRA WHILE 2*)
    If[Abs[f[rop]]<Abs[f[ropaux]],
      {rop=Re[rop];},
      {contador1=20;}
    ];
  }];(*CIERRA WHILE 1*);

(*BISECCION EN EL CASO DE QUE NEWTON NO CONVERJA*)
bandera=0;
If[Abs[f[rop]]>Tolr || rop>r2 || rop<r1,
  {bandera=1;
  r1aux=r1;   r2aux=r2;   paso=2;
  (*While Para quedarnos en el intervalo [r1,r2]*)
  While[f[r1aux]*f[r2aux]>0,
    {r1aux=-1/xmax2+epsilon/(paso^5);
    r2aux=-1/xmin2-epsilon/(paso^5);
    paso++;
    }];
  rop=(r1aux+r2aux)/2;
  i=0;
  While[Abs[f[rop]]>Tolr001 && i<31,
    {(*Acelerador de la biseccion*)
    long=r2aux-r1aux;
    fr1=f[r1aux];
    If[fr1*f[r1aux+long/6]<0,
      {Print["Acelerador ",i];r2aux=r1aux+long/6},
      If[f[r1aux+5*long/6]*f[r2aux]<0,
        {Print["Acelerador ",i];r1aux=r1aux+5*long/6}]];
    (*Fin del acelerador*)
    rop=(r1aux+r2aux)/2;
    If[f[rop]*fr1>0,r1aux=rop,r2aux=rop];
    i=i+1;
    }];
  Print["Salida de biseccion con ",i,
  " iteraciones, rop=",rop," f[rop]=",f[rop]];
}];(*CIERRA IF BISECCION *)
priorr=rop;
lam=nt/(1+theta*rop);
mu=lam*rop;
paso=Re[priorr+acel[priorr]*incre];
If[r1<=paso<=r2,rop=paso];
Do[{p[[i]]=n[i]/(lam+mu*x2[i])},{i,1,k}];
theta=theta+cptest;
```

```

    }); (*CIERRA 1 IF*)

(*REESCALADO PARA QUE LOS VALORES DE LA VEROSIMILITUD NO SE VAYAN A
CERO*)
pasolike=(10^10000)Re[Exp[Sum[n[i]*Log[p[[i]],{i,1,k}]]]];

like[jj]=N[pasolike,12];
sumpost=sumpost+N[prior*pasolike,10];
ppostnor[jj]=N[prior*pasolike,10];
If[Mod[jj,50]==0,Print[jj]];
},{jj,10,ptos-10]];
Do[{like[jj]=0.0;ppostnor[ptos-jj]=0.0;
like[ptos-jj]=0.0;ppostnor[jj]=0.0},{jj,1,9}]

(*NORMALIZACION DE LA VEROSIMILITUD*)
sumlike=Sum[like[i],{i,1,ptos-1}]*incretheta;
numptos=ptos-1;
Do[likenor[i]=like[i]/sumlike,{i,1,ptos-1}];

(*NORMALIZACION DE LA POSTERIORI Y CALCULO DE LA POSTERIORI
ACUMULADA*)
sumpost=Array[ppostnor,{ptos-1},1,Plus]*incretheta
Do[ppostnor[i]=ppostnor[i]/sumpost,{i,1,ptos-1}]
ppostacum[1]=ppostnor[1]*incretheta;
Do[ppostacum[i]=ppostacum[i-1]+ppostnor[i]*incretheta,{i,2,ptos-1}]
Sum[ppostnor[i]*incretheta,{i,1,ptos-1}]

(*DENSIDAD PREDICTIVA*)
densidadpred = 
$$\frac{\sum_{i=1}^{ptos-1} \text{incretheta} * \text{like}[i] * \text{priori}[i]}{10^{10000}}$$


(*VEROSIMILITUD ACUMULADA*)
likeacum[1] = 0
Do[likeacum[i] = likeacum[i-1] + likenor[i] * incretheta, {i, 2, ptos-1}]

(*MEDIA DE LA VEROSIMILITUD*)
medveros = 
$$\sum_{i=1}^{ptos-1} \text{thetas}[i] * \text{likenor}[i] * \text{incretheta}$$


(*LISTADO DE LAS DISTINTAS DISTRIBUCIONES OBTENIDAS*)
Do[Print[N[thetas[i]], " ", priori[i]], {i, 1, ptos-1}]
Do[Print[N[thetas[i]], " ", prioriacum[i]], {i, 1, ptos-1}]
Do[Print[i, " ", N[thetas[i]], " ", like[i]], {i, 1, ptos-1}]
Do[Print[N[thetas[i]], " ", likenor[i]], {i, 1, ptos-1}]
Do[Print[N[thetas[i]], " ", likeacum[i]], {i, 1, ptos-1}]
Do[Print[i, " ", N[thetas[i]], " ", ppostnor[i]], {i, 1, ptos-1}]
Do[Print[N[thetas[i]], " ", ppostacum[i]], {i, 1, ptos-1}]
Do[Print[i, " ", N[thetas[i]], " ", ppostnor[i]], {i, 600, 700}]
Do[Print[N[thetas[i]], " ", ppostacum[i]], {i, 600, 700}]

```

APROXIMACIÓN BAYESIANA A LA METODOLOGÍA DE LA VEROSIMILITUD EMPÍRICA

(*GRAFICAS TOTALES Y PARCIALES DE LAS DISTINTAS DISTRIBUCIONES*)

(*GRAFICA DE LA VEROSIMILITUD*)

```
numptos=ptos-1;
tablalike:=Table[{thetas[i],N[likenor[i]]},{i,1,numptos}];
glike=ListPlot[tablalike,PlotJoined->True,PlotRange->All];
tablalikeparcial:=Table[{thetas[i],N[likenor[i]]},{i,600,700}];
glikeparcial=ListPlot[tablalikeparcial,PlotJoined->True,
  >All]
```

(*GRAFICA DE LA A PRIORI*)

```
tablaprior:=Table[{thetas[i],priori[i]},{i,1,numptos}];
gprior=ListPlot[tablaprior,PlotJoined->True,PlotRange->All];
tablapriorparcial:=Table[{thetas[i],N[priori[i]]},{i,600,700}];
gpriorparcial=ListPlot[tablapriorparcial,
  >True,PlotRange->All]
```

(*GRAFICA DE LA A POSTERIORI*)

```
tablapost:=Table[{thetas[i],ppostnor[i]},{i,1,numptos}];
gpost=ListPlot[tablapost,PlotJoined->True,PlotRange->All];
tablapostparcial:=Table[{thetas[i],N[ppostnor[i]]},{i,600,700}];
gpostparcial=ListPlot[tablapostparcial,
  >True,PlotRange->All]
```

(*GRAFICA DE LA A POSTERIORI ACUMULADA*)

```
tablapostacum:=Table[{thetas[i],ppostacum[i]},{i,1,numptos}];
gpostacum=ListPlot[tablapostacum,PlotJoined->True,PlotRange->All];
tablapostacumparcial:=Table[{thetas[i],N[ppostacum[i]]},{i,600,700}];
gpostacumparcial=ListPlot[tablapostacumparcial,
  >True,PlotRange->All]
```

(*EN UNA MISMA GRAFICA VARIAS DISTRIBUCIONES*)

```
Show[gpriorparcial,gpostparcial]
>Show[gpostparcial,glikeparcial]
>Show[glikeparcial,gpriorparcial]
>Show[glikeparcial,gpriorparcial,gpostparcial]
```

(*CÁLCULO Y GRAFICO DE LA RAZÓN DE VEROSIMILITUD EMPÍRICA*)

$$ctel = N\left[\frac{3114^{3114}}{10^{10000} \prod_{i=1}^k n[i]^{n[i]}}\right]$$

```
Do[{rlike[jj] = ctel * like[jj]}, {jj, 10, ptos - 10}]
Do[Print[N[thetas[i]], " ", rlike[i]], {i, 10, ptos - 10}]
Do[{lgrlike[jj] = -2 * Log[rlike[jj]]; logrlike[jj] = N[lgrlike[jj], 12]}, {jj, 10, ptos - 10}]
Do[Print[N[thetas[i]], " ", lgrlike[i]], {i, 10, ptos - 10}]
tablaogrlike := Table[{thetas[i], logrlike[i]}, {i, 10, ptos - 10}]
glogrlike = ListPlot[tablaogrlike, PlotJoined->True, PlotRange->All];
tablaogrlikeparcial := Table[{thetas[i], N[logrlike[i]]}, {i, 600, 700}];
glogrlikeparcial = ListPlot[tablaogrlikeparcial, PlotJoined->True, PlotRange->All]
```

A.1.4. Cálculo de medidas a posteriori.

```
(*PERCENTILES*)
prob[1] = .01;
prob[2] = .025;
prob[22] = .975;
prob[23] = .99;
prob[24] = 1.0;
Do[prob[i] = (i-2) * .05;, {i, 3, 21}];
tol3 = .00001;
sumpr = 0;
i = 1;
While[sumpr < .025,
  {sumpr = sumpr + priori[i];
  i = i + 1;
  }];
q025priori = N[10000 * (thetas[i] + thetas[i-1]) / 2, 7];
While[sumpr < .975,
  {sumpr = sumpr + priori[i];
  i = i + 1;
  }];
q975priori = N[10000 * (thetas[i] + thetas[i-1]) / 2, 7];
i = 1;
j = 1;
Do[{While[ppostacum[j] < prob[i], j = j + 1];
  If[j == 1 && ppostacum[1] > prob[i],
    {m = ppostacum[1] / (thetas[1] - a);
    If[Abs[m] < tol3, percentil[i] = thetas[1],
      percentil[i] = thetas[1] +
        (prob[i] - ppostacum[1]) / m;
    }];
  }];
If[j > 1,
  {
  m = (ppostacum[j] - ppostacum[j-1]) /
    (thetas[j] - thetas[j-1]);
  (* SI m ES MUY PEQUEÑO PUEDE DAR PROBLEMAS
    AL DIVIDIR POR m (m=0) *)
  If[Abs[m] < tol3, {percentil[i] = thetas[j]; j = j - 1},
    percentil[i] = thetas[j] + ((prob[i] - ppostacum[j]) / m)];
  percentil[i] = N[percentil[i], 7];
  }
```

```

}
  ]
}
, {i, 1, 23}];

(*MEDIA, MODA Y VARIANZA*)
media = 0

$$\sum_{i=1}^k \text{ppostnor}[i] * \text{incretheta}$$

Do[media = media + thetas[i] * ppostnor[i] * incretheta, {i, 1, ptos - 1}];
varianza = 0;
Do[varianza = varianza + ((thetas[i] - media) ^ 2) * ppostnor[i] * incretheta
, {i, 1, ptos - 1}];
desvtip = Sqrt[varianza];
mediapost = media;
sigma = desvtip;
mediapost
Do[If[ppostnor[j - 1] < ppostnor[j],
{modapost = N[thetas[j]},
}]
, {j, 2, ptos - 1}]

Print["Media a posteriori=", " ", media];
Print["Desviación típica=", " ", desvtip];
Do[Print[prob[i], " ", percentil[i]];
, {i, 1, 23}]
Print["Media muestral= ", N[cpstest]];
Print["Media a posteriori=", " ", media];
Print["Moda a posteriori=", " ", modapost];

```

A.2. Robustez Bayesiana.

Vemos los programas desarrollados para estudiar la robustez bayesiana en varias clases de contaminación.

A.2.1. Clase formada por todas las distribuciones.

```

sump = Sum[priori[i], {i, 1, ptos - 1}] * incretheta
sumpost = Sum[priori[i] * like[i], {i, 1, ptos - 1}] * incretheta
ep = 0.3 (*VALOR DE EPSILON CONSIDERADO*)
A = sumpost * (1 - ep) / ep
medprior = 0;
Do[medprior = medprior + thetas[i] priori[i] * incretheta, {i, 1, ptos - 1}]
medprior
medlike = 0;
Do[medlike = medlike + thetas[i] likenor[i] * incretheta, {i, 1, ptos - 1}]
medlike
medpost = 0;
Do[medpost = medpost + thetas[i] ppostnor[i] * incretheta, {i, 1, ptos - 1}]
medpost
A0 = A * medpost
Do[valor[i] = (A0 + thetas[i] * like[i]) / (A + like[i]), {i, 1, ptos - 1}]
(*Do[Print[N[thetas[i]], " ", valor[i]], {i, 1, ptos - 1}]*)
numptos;
(*GRAFICA DEL VALOR*)
tablavalor := Table[{thetas[i], valor[i]}, {i, 1, numptos}]
gvalor = ListPlot[tablavalor, PlotJoined -> True, PlotRange -> All];
tablavalorparcial := Table[{thetas[i], N[valor[i]]}, {i, 600, 700}];
gvalorparcial = ListPlot[tablavalorparcial, PlotJoined -> True, PlotRange -> All]
(*MAXIMO Y MINIMO*)
vmax = valor[1];
vmin = valor[1];
Do[{vmax = Max[vmax, valor[i + 1]], vmin = Min[vmin, valor[i + 1]]}, {i, 1, ptos - 2}];
vmax
vmin
(*SENSIBILIDAD RELATIVA*)
rs = (vmax - vmin) / (2 * medpost) * 100

```

A.2.2. Clase formada por todas las distribuciones unimodales en θ_0 .

```

sumpost
ep = 0.3
A = sumpost * (1 - ep) / ep
medpost = 0;
Do[medpost = medpost + thetas[i] ppostnor[i] increthetha, {i, 1, ptos - 1}]
medpost
A0 = A * medpost
increthetha
th = 652 (*ES EL ORDEN QUE OCUPA LA MODA*)
H01[z_] := (1/z) * (Sum[Like[i], {i, th - z, th}])
H0[0] = like[th]
H0[z_] := (1/z) * (Sum[like[i], {i, th, th + z}])
Hg1[z_] := (1/z) * (Sum[thetas[i] like[i], {i, th - z, th}])
Hg[0] = like[th] * thetas[th]
Hg[z_] := (1/z) * (Sum[thetas[i] like[i], {i, th, th + z}])
Do[val[i] = (A0 + Hg1[th - i]) / (A + H01[th - i]), {i, 1, th - 1}]
Do[val[i] = (A0 + Hg[i - th]) / (A + H0[i - th]), {i, th, ptos - 8000}]
(*Do[Print[i, " ", val[i]], {i, 1, ptos - 1}]*)
(*GRAFICA DEL VALOR*)
tablavalor := Table[{thetas[i], val[i]}, {i, 1, ptos - 8000}]
gvalor = ListPlot[tablavalor, PlotJoined -> True, PlotRange -> All];
tablavalorparcial := Table[{thetas[i], N[valor[i]]}, {i, 600, 700}];
gvalorparcial = ListPlot[tablavalorparcial, PlotJoined -> True, PlotRange -> All]
(*CALCULO DEL MAXIMO Y MINIMO*)
vmax = val[1];
vmin = val[1];
Do[{vmax = Max[vmax, val[i + 1]], vmin = Min[vmin, val[i + 1]]}, {i, 1, ptos - 8000}];
vmax
vmin
(*SENSIBILIDAD RELATIVA*)
rs = (vmax - vmin) / (2 * medpost) * 100

(*CALCULO DE LA DISTRIBUCIÓN A PRIORI ML-II DE LA CLASE*)

(*OBTENCION DE Z TAL QUE MAXIMIZA H0 (Z)*)
zmax = H0[0];
Do[{zmax = Max[zmax, H0[i]]}, {i, 1, ptos - th}];
zmax
Do[Print[i, " ", H0[i]], {i, 0, ptos - th}]

```

```
(*OBTENCION DE Z TAL QUE MAXIMIZA H01 (Z)*)
zmax1 = H01[1];
Do[{zmax1 = Max[zmax1, H01[i+1]]}, {i, 1, th-1}];
zmax1
Do[Print[i, " ", H01[i]], {i, 1, th}]
```

A.2.3. Clase formada por todas las distribuciones unimodales en θ_0 y simétricas.

```
sumpost
ep = 0.3
A = sumpost * (1 - ep) / ep
medpost = 0;
Do[medpost = medpost + thetas[i] ppostnor[i] incretheta, {i, 1, ptos-1}]
medpost
A0 = A * medpost
incretheta
th = 652 (*ES EL ORDEN QUE OCUPA LA MODA*)
H0[0] = like[th]
H0[z_] := (1 / (2 * z)) * (Sum[like[i], {i, th-z, th+z}])
Hg[0] = like[th] * thetas[th]
Hg[z_] := (1 / (2 * z)) * (Sum[thetas[i] like[i], {i, th-z, th+z}])
Do[like[i] = 0, {i, -ptos, 0}]
Do[like[i] = 0, {i, ptos, ptos + ptos}]
ptos1 = Min[th, ptos - th]
Do[val[i] = (A0 + Hg[i-1]) / (A + H0[i-1]), {i, 1, ptos1+1000}]
(*Do[Print[i, " ", val[i]], {i, 1, ptos1+100}]*
(*GRAFICA DEL VALOR*)
tablavalor := Table[{thetas[i], val[i]}, {i, 1, ptos1}]
gvalor = ListPlot[tablavalor, PlotJoined -> True, PlotRange -> All];
tablavalorparcial := Table[{thetas[i], N[val[i]]}, {i, 1, 100}];
gvalorparcial = ListPlot[tablavalorparcial, PlotJoined -> True, PlotRange -> All]
(*CALCULO DEL MAXIMO Y MINIMO*)
vmax = val[1];
vmin = val[1];
Do[{vmax = Max[vmax, val[i+1]], vmin = Min[vmin, val[i+1]]}, {i, 1, ptos1-100}];
vmax
vmin
(*SENSIBILIDAD RELATIVA*)
rs = (vmax - vmin) / (2 * medpost) * 100
```

APROXIMACIÓN BAYESIANA A LA METODOLOGÍA DE LA VEROSIMILITUD EMPÍRICA

(*CALCULO DE LA DISTRIBUCIÓN A PRIORI ML-II DE LA CLASE*)

(*OBTENCION DE Z TAL QUE MAXIMIZA H0 (Z)*)

zmax = H0[0] ;

Do[{zmax = Max[zmax, H0[i]]}, {i, 1, ptosl}];

zmax

Do[Print[i, " ", H0[i]], {i, 0, ptosl + 100}];

Bibliografía:

- AMOR PULIDO, R. (1998). *Verosimilitud Empírica*. Trabajo de Investigación. Dpto. Estadística e I.O. Universidad de Granada.
- BASU, S. (1994). *Variations of Posterior Expectations for Symmetric Unimodal Priors in a Distribution Band*. *Sankhya, A*, 56, pg 320-334.
- BASU, S. (1995). *Ranges of posterior Probability over a Distribution Band*. *Journal of Statistical Planning and Inference*, 44, pg 149-166.
- BASU, S. y DASGUPTA, A. (1995). *Robust Bayesian Analysis with Distribution Bands*. *Statistics and Decisions*, 13, pg 333-349.
- BERGER, J.O. (1985). *Statistical Decision Theory and Bayesian Analysis*. Ed. Springer-Verlag. Second Edition.
- BERGER, J.O. (1994). *An Overview of Robust Bayesian Analysis*. *Test*, 3, 1, pg 5-124.
- BERGER, J.O. y BERLINER, L. (1986). *Robust Bayes and Empirical Bayes Analysis with ε -contaminated Priors*. *The Annals of Statistics*, 14, 2, pg 461-86.
- BERGER, J.O.; RÍOS INSUA, D. y RUGGERI, F. (2000). *Bayesian Robustness*. In *Robust Bayesian Analysis* (D. Ríos Insua y F. Ruggeri, eds.). New York: Springer-Verlag.
- BERGER, J.O. y O'HAGAN, A. (1988). *Ranges of Posterior Probabilities for Unimodal Priors with Specified Quantiles*. In *Bayesian Statistics 3* (J.M. Bernardo y otros, eds.) pg 45-66. Oxford: Oxford University Press.

BIBLIOGRAFÍA

- BETRO, B. y GUGLIELMI, A. (2000). *Methods for Global Prior Robustness under Generalized Moment Conditions*. In Robust Bayesian Analysis (D. Ríos Insua y F. Ruggeri, eds.). New York: Springer-Verlag.
- BETRO, B. y RUGGERI, F. (1994). *Robust Bayesian Analysis under Generalized Moments Conditions*. Journal of Statistical Planning and Inference, 41, pg 257-266.
- CANO, J.A.; HERNÁNDEZ BASTIDA, A. y MORENO, E. (1986). *Posterior Measure under Partial Prior Information*. Statistica, 46, pg 219-30.
- CANO, J.A.; HERNÁNDEZ BASTIDA, A. y MORENO, E. (1987a). *Conditioning with upper Integral.Upper-Sufficiency*. Pub Inst Stat Univ, 32 (3), pg 15-24.
- CANO, J.A.; HERNÁNDEZ BASTIDA, A. y MORENO, E. (1987b). *On Minimum Risk Parameter Estimation under Partial Prior Information*. Pub Inst Stat Univ, 32 (4), pg 7-22.
- CANO, J.A.; HERNÁNDEZ BASTIDA, A. y MORENO, E. (1987c). *A Short Note on the J-Minimax and Mean-max Risk Functions*. Pub Inst Stat Univ, 32 (3), pg 25-32.
- CANO, J.A.; HERNÁNDEZ BASTIDA, A. y MORENO, E. (1989). *On L-sufficiency concept of partial sufficiency*. Statistica, 49 (4), pg 519-528.
- DeGROOT, M. (1970). *Optimal Statistical Decisions*. Ed. McGraw-Hill.
- DiCICCIO, T.; HALL, P. y ROMANO, J.P. (1989). *Comparison of Parametrics and Empirical Likelihood Functions*. Biometrika, 76, 3, pg 447-56.
- *Encuesta continua de Presupuestos familiares correspondientes al primer trimestre de 1997*. INE. Madrid.
- GÓMEZ DÉNIZ, E. (1996). *Estadística Bayesiana en Credibilidad con Aplicación a la Fijación de Primas de Seguros*. Tesis Doctoral. Dpto. Economía Aplicada. Universidad de Las Palmas de G.C.
- HALL, P.; LA SCALA, B. (1990). *Methodology and Algorithms of the Empirical Likelihood*. International Statistical Review, 58, 2, pg 273-89.
- JAYNES, E.T. (1968). *Prior probabilities*. IEEE Transactions on Systems Science and Cybernetics SSc-4, pg 227-241.
- LAVINE, M. (1991). *Sensitivity in Bayesian Statistics: the Prior and the Likelihood*. Journal of the American Statistical Association, 86, pg 396-399.

BIBLIOGRAFÍA

- LAVINE, M. (1991). *An Approach to Robust Bayesian Analysis for Multidimensional Parameter Spaces*. Journal of the American Statistical Association, 86, pg 400-403.
- LAVINE, M.; PERONE-PACIFICO, M.; SALINETTI, G. y TARDELLA, L. (2000). *Linearization Techniques in Bayesian Robustness*. In Robust Bayesian Analysis (D. Ríos Insua y F. Ruggeri, eds.). New York: Springer-Verlag.
- LEÓN, C.J. y VÁZQUEZ POLO, F.J. (1998). *A Bayesian Approach to Double Bounded Contingent Valuation*. Environmental and Resource Economics, 11, pg 197-215.
- LISEO, B.; PETRELLA, L y SALINETTI, G. (1996). *Robust Bayesian Analysis: An Interactive Approach*. In Bayesian Statistics 5 (J.O. Berger y otros eds.) pg 661-666. Oxford: Oxford University Press.
- MARTEL ESCOBAR, M.C. (1996). *Aportaciones al Estudio de Técnicas Bayesianas en Auditoría*. Tesis Doctoral. Dpto. Economía Aplicada. Universidad de Las Palmas de G.C.
- McCRAY, J. (1997). *A Practical Heuristic Distribution Free Multiparameter Most Likely Probability Density Function*. Manuscrito no publicado. The College of William and Mary. School of Business Administration.
- McCRAY, J. (1998). *Most Likely Probability Density Function for Simultaneous Inference of Multiple Parameters*. Manuscrito no publicado. The College of William and Mary. School of Business Administration.
- McCRAY, J.; HERNÁNDEZ BASTIDA, A. y VÁZQUEZ POLO, F.J. (1995). *Modifying Likelihood by Prior Information in Business Statistics*. Dpto. Economía Aplicada. Documentos de Trabajo. Universidad de Las Palmas de G.C.
- MORENO, E. (2000). *Global Bayesian Robustness for some Classes of Prior Distributions*. In Robust Bayesian Analysis (D. Ríos Insua y F. Ruggeri, eds.). New York: Springer-Verlag.
- MORENO, E. y CANO, J.A. (1991). *Robust Bayesian Analysis with ε -contaminations Partially Known*. J R Statist Soc B, 53, 1, pg 143-55.
- MORENO, E. y GONZÁLEZ CARMONA, A. (1990). *Empirical Bayes Analysis for ε -contaminated Priors with shape and quantile Constraints*. Revista Brasileira de Probabilidade e Estatística, 4, pg 177-200.

BIBLIOGRAFÍA

- MORENO, E.; MARTINEZ, C. y CANO, J.A. (1996). *Local Robustness and Influence for Contamination Classes of Prior Distributions*. In Bayesian Robustness IMS Lectures-Notes Monograph Series (J.O. Berger y otros, eds.), Vol 29, 137-154. Hayward: IMS.
- MORENO, E. y PERICCHI, L. (1992). *Bands of Probability Measures: a Robust Bayesian Analysis*. In Bayesian Statistics 4 (J.O. Berger y otros, eds.) pg 607-713. Oxford: Oxford University Press.
- MORENO, E. y PERICCHI, L. (1993). *Prior Assessments for Bandas of Probability Measures: Empirical Bayes Analysis*. Test, 2, 1-2, pg 101-110.
- MORENO CARRETERO, M.F.; HERNÁNDEZ BASTIDA, A. y VÁZQUEZ POLO, F.J. (1997). *Estimación del error en una contabilidad: procedimientos no paramétricos y uso de la distribución multinomial*. Universidad de Almería. Servicio de Publicaciones.
- O'HAGAN, A. y BERGER, J.O. (1988). *Ranges of Posterior Probabilities for Quasiunimodal Priors with Specified Quantiles*. Journal of the American Statistical Association, 83, pg 503-508.
- OWEN, A. (1988). *Empirical Likelihood Ratio Confidence Intervals for a single Functional*. Biometrika, 75, 2, pg 237-49.
- OWEN, A. (1990). *Empirical Likelihood Ratio Confidence Regions*. The Annals of Statistics, 18, 1, pg 90-120.
- PERONE-PACIFICO, M.; SALINETTI, G. y TARDELLA, L. (1996). *Bayesian Robustness on Constrained Density Band Classes*. Test, 5, 2, pg 395-409.
- RIOS INSUA, S. (1988). *Investigación Operativa. Optimización*. Ed. Centro de Estudios Ramón Areces.
- SIVAGANESAN, S. (1988). *Range of Posterior Measures for Priors with Arbitrary Contaminations*. Commun Statist-Theory Meth, 17 (5), pg 1591-612.
- SIVAGANESAN, S. (1989). *Sensitivity of Posterior Mean to Unimodality Preserving Contaminations*. Statistics and Decisions, 7, pg 77-93.
- SIVAGANESAN, S. (1991). *Sensitivity of some Standard Bayesian Estimates to Prior Uncertainty: A Comparison*. Journal of Statistical Planning and Inference, 27, pg 85-103.

BIBLIOGRAFÍA

- SIVAGANESAN, S. (1991). *Sensitivity of some Posterior Summaries when the Prior is Unimodal with Specified Quantiles*. The Canadian Journal of Statistics, 19, 1, pg 57-65.
- SIVAGANESAN, S. (1993). *Range of the Posterior Probability of an Interval for Priors with Unimodality Preserving Contaminations*. Ann Inst Statist Math, 45, 1, pg 187-199.
- SIVAGANESAN, S. (1994). *Bounds on Posterior Expectations for Density Bounded Class with Constant Bandwidth*. Journal of Statistical Planning and Inference, 40, pg 331-343.
- SIVAGANESAN, S. (2000). *Global and Local Robustness Approaches: Uses and Limitations*. In Robust Bayesian Analysis (D. Ríos Insua y F. Ruggeri, eds.). New York: Springer-Verlag.
- SIVAGANESAN, S. y BERGER, J. (1989). *Ranges of Posterior Measures for Priors with unimodal Contaminations*. The Annals of Statistics, 17, 2, pg 868-89.
- VÁZQUEZ POLO, F.J. (1992). *Técnicas Estadísticas Bayesianas en Auditoría. Un Análisis de Robustez*. Tesis Doctoral. Dpto. Economía Aplicada. Universidad de Las Palmas de G.C.
- WILKS, S.S. (1938). *The Large-sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses*. Ann Math Statist, 9, pg 60-62.
- ZACKS, S. (1971). *The theory of Statistical Inference* .Ed. John Wiley and Sons.

