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

TESIS DOCTORAL

Robustez en inferencia bayesiana : un estudio cualitativo

MEMORIA PARA OPTAR AL GRADO DE DOCTOR
PRESENTADA POR

Antonio Cuevas González

DIRECTOR:

Manuel del Rio Bueno

Madrid, 2015

Antonio Cuevas González

ROBUSTEZ EN INFERENCIA BAYESIANA: UN ESTUDIO CUALITATIVO

Departamento de Estadística e Investigación Operativa
Facultad de Ciencias Matemáticas
Universidad Complutense de Madrid
1982

Colección Tesis Doctorales. Nº 185/82

Edita e imprime la Editorial de la Universidad
Complutense de Madrid. Servicio de Reprografía
Noviciado, 3 Madrid-8
Madrid, 1982
Xerox 9200 XB 480
Depósito Legal: M-24218-1982

U N I V E R S I D A D   C O M P L U T E N S E

FACULTAD DE CIENCIAS MATEMATICAS

Sección de Estadística e Investigación Operativa

ROBUSTEZ EN INFERENCIA BAYESIANA: UN ESTUDIO CUALITATIVO

Memoria para optar al grado de Doctor en Ciencias Matemáticas, presentada por Antonio Cuevas González.

Madrid, Noviembre, 1981

INDICE

PROLOGO

PROLOGO

El objetivo de esta memoria es, en esencia, esbozar una teoría matemática de la robustez en inferencia bayesiana respecto a variaciones en la verosimilitud.

En concreto, se propone una definición general de "robustez cualitativa" o "estabilidad" de un modelo bayesiano y, después de analizar su significado y aplicabilidad, se enuncian otras definiciones particulares de robustez para los estimadores, regiones de confianza y tests de hipótesis bayesianos. Paralelamente se demuestran teoremas que proporcionan condiciones suficientes de robustez (según las distintas definiciones mencionadas) y se aplican estos conceptos al análisis de algunos problemas particulares de inferencia.

La bibliografía disponible sobre este tema es relativamente escasa. Los trabajos de Box y Tiao (1962, 1964, 1968 y 1973) constituyen, sin duda, la aportación más conocida. En ellos se plantean algunos modelos y se comparan las distribuciones a posteriori obtenidas con distintas verosimilitudes (la distribución a priori permanece fija) adoptando un enfoque eminentemente cuantitativo en el que incluso se utilizan muestras numéricas concretas.

Por el contrario, nuestro enfoque del problema (desarrollado en los capítulos II y III de esta memoria) es cualitativo, en el sentido de que las definiciones de robustez que se proponen están inspiradas en la idea de continuidad y pretenden recoger la noción intuitiva de que un procedimiento estadístico es "robusto" cuando es poco sensible frente a cambios "infinitesimales" en la verosimilitud sub yacente. Naturalmente, éste es un aspecto parcial de la noción de robustez que, en cada problema estadístico, puede completarse con un análisis de tipo cuantitativo.

En el capítulo I, después de enunciar algunos resultados previos, se hace un breve resumen de la teoría de la robustez en su estado actual indicando las líneas generales de investigación y destacando aquellas que tienen relación directa con nuestro planteamiento, o bien presentan alguna afinidad metodológica con el mismo. Así, por ejemplo, se dedica cierta atención a la "Teoría de la Robustez Cualitativa" de Hampel que, dentro del marco de la Estadística clásica o frecuentista, establece una estrecha relación entre el concepto de robustez de un estimador y la continuidad de un funcional definido sobre un espacio de distribuciones dotado de la topología débil.

En el capítulo II se enuncia la definición fundamental de estabilidad de un modelo bayesiano (constituido por una familia de posibles verosimilitudes y por una distribución a priori). En ella se intenta reflejar la idea de estabilidad concebida como una "continuidad" de las distribuciones a posteriori respecto a los cambios en la verosimilitud. De hecho, se demuestra un teorema que, bajo hipótesis bastante amplias, permite traducir la definición en términos de continuidad de una familia de funcionales definidos en un espacio de distribuciones dotado de la topología débil. Se establece también una versión generalizada de este resultado.

Podemos decir, por tanto, que existe una cierta analogía de fondo con la teoría de Hampel en cuanto al tratamiento matemático, aunque los planteamientos generales son claramente distintos.

En lo referente al manejo práctico de la definición, se demuestran varios teoremas que proporcionan condiciones suficientes de estabilidad. El soporte matemático de estos resultados es el concepto de convergencia débil. En particular, aparece el problema de establecer condiciones que permitan asegurar la convergencia puntual de las funciones de densidad a partir de la convergencia débil de las correspondientes distribuciones. Estas condiciones se obtienen en el capítulo II mediante dos procedimientos distintos y constituyen un importante elemento auxiliar en los resultados sobre estabilidad.

El capítulo III está dedicado a aplicar y desarrollar en varios sentidos la teoría del capítulo anterior: En la primera parte esta teoría se utiliza para estudiar el efecto de ciertas desviaciones de la normalidad en el tratamiento bayesiano del modelo lineal.

A continuación, se proponen definiciones de robustez para los distintos métodos estadísticos: estimación puntual, regiones de confianza y test de hipótesis bayesianos. De acuerdo con nuestro planteamiento, estos conceptos especiales de robustez deberían analizarse como un paso posterior después de haber asegurado la estabilidad genérica propia del modelo (definida en el capítulo II), que constituye el punto de partida natural.

Así, en la metodología seguida para establecer condiciones suficientes de robustez de los procedimientos estadísticos, se impone la estabilidad del modelo y se añaden hipótesis suplementarias para asegurar que dicha estabilidad básica se conserva al "resumir" la información proporcionada por la distribución a posteriori.

Los resultados obtenidos en el capítulo III indican que la robustez de los tests se sigue de forma casi automática a partir de la estabilidad del modelo mientras que en el caso de los estimadores hay que imponer condiciones adicionales (a veces bastante fuertes). Concretamente, se demuestran teoremas de robustez para la media, mediana y moda a posteriori y los estimadores Bayes genéricos.

El problema de las regiones de confianza es, por su propia naturaleza, más complicado. En el texto se establece la robustez de las regiones de confianza unidimensionales definidas mediante cuantiles.

Por último, y en un aspecto diferente, apuntamos la posibilidad de desarrollar una "teoría generalizada" de la robustez cualitativa en inferencia bayesiana que estudiase los efectos de pequeñas variaciones en los tres elementos fundamentales de un problema bayesiano (verosimilitud, distribución a priori, función de pérdida) tomando como punto de partida la teoría desarrollada en los capítulos IIy III. Un paso posterior en esta dirección sería el resultado que establecemos al final del capítulo III demostrando, bajo ciertas condiciones, la "robustez conjunta del operador Bayes" respecto a variaciones simultáneas en la verosimilitud y la distribución a priori. Por otra parte, el artículo de Kadane y Chuang (1978) (que se ocupa de las alteraciones en la función de pérdida y en la distribución a priori) puede considerarse insertado en esta misma línea de investigación que parece bastante prometedora.

Esta visión general del trabajo puede completarse con los sumarios incluidos al principio de cada capítulo. Asimismo, después de los resultados más importantes, se incluyen un breve apartado de discusión y comentarios para aclarar su significado.

Quisiera expresar mi agradecimiento al profesor Manuel del Río por su dirección y su ayuda constante.

Agradezco también a los profesores de la Sección de Estadística e Investigación Operativa el apoyo prestado durante la realización de este trabajo.

Madrid, Noviembre de 1981

CAPITULO I

CAPITULO I

RESUMEN GENERAL DE LA TEORIA DE LA ROBUSTEZ

1. SUMARIO

El propósito fundamental de este capítulo es ofrecer una visión general de la Teoría de la Robustez que permita situar adecuadamente el trabajo desarrollado en los capítulos II y III.

En la sección 2 se enuncian algunos resultados conocidos, (relativos a la convergencia débil, la distancia de Prohorov y la fórmula de Bayes) que utilizaremos en el desarrollo posterior.

La sección 3 se dedica a resumir la Teoría de la Robustez en Inferencia clásica o frecuentista: Se comentan brevemente algunos desarrollos importantes, como los métodos "experimentales" basados en la simulación, la "robustez minimax" de Huber y la teoría cualitativa de Hampel. Esta última se examina con mayor atención, ya que presenta una cierta afinidad metodológica con el planteamiento de nuestro trabajo en lo referente a la interpretación del concepto de robustez en términos de continuidad.

En la sección 4 se resume el estado actual de la Teoría de la Robustez en Inferencia bayesiana en su triple faceta de estudiar los efectos de las alteraciones en la verosimilitud, en la distribución a priori y en la función de pérdida. Concretamente, en el apartado 4.1 se resume el conocido planteamiento de Box-Tiao, en el 4.2 se considera el problema de "sensibilidad" (robustez respecto a la distribución a priori) desde los puntos de vista cualitativo y cuantitativo, y, por último, el apartado 4.3 se dedica a resumir un artículo de Kadane y Chuang en el que se propone una definición de estabilidad frente a pequeñas variaciones simultáneas en la función de pérdida y la distribución a priori.

2. RESULTADOS PREVIOS

En esta sección se enuncian, para referencia posterior, algunos resultados generales conocidos que se utilizarán reiteradamente en el desarrollo de este trabajo.

(a) Convergencia débil

Sea χ un espacio métrico completo y separable y $\mathcal{B}$ la σ-álgebra engendrada por los abiertos de χ.

Sean P_r, $r \in \mathbb{N}$ y P medidas de probabilidad en $(\chi, \mathcal{B})$.

Se dice que la sucesión $\{P_r\}_{r\in\mathbb{N}}$ converge débilmente a P (y se indicará por $P_r \xrightarrow[r\to\infty]{d} P$) cuando, para toda función $u : \chi \to \mathbb{R}$, contínua y acotada, se verifica $\int_\chi u(x)dP_r(x) \xrightarrow[r\to\infty]{} \int_\chi u(x)dP(x)$.

El siguiente teorema ofrece algunas caracterizaciones interesantes de la convergencia débil:

TEOREMA I.1. Las siguientes condiciones son equivalentes:

(a) $P_r \xrightarrow[r\to\infty]{d} P$

(b) $\limsup_{r\to\infty} P_r(C) \leq P(C)$ para todo conjunto cerrado C.

(c) $\liminf_{r\to\infty} P_r(A) \geq P(A)$ para todo conjunto abierto A

(d) $\lim_{r\to\infty} P_r(B) = P(B)$, para todo conjunto B tal que su frontera ∂B cumpla $P(\partial B) = 0$.

Además, cuando $\chi = R^n$, estas cuatro condiciones son equivalentes a

(e) $F_r(x) \xrightarrow[r\to\infty]{} F(x)$ para todo x que sea punto de continuidad de F, donde F_r, $r \in N$ y F son las funciones de distribución asociadas a P_r, $r \in N$ y P respectivamente (se indicará $F_r \xrightarrow[r\to\infty]{d} F$).

La demostración puede verse en Billingsley (1971).

El siguiente resultado, debido a Scheffé (1947), prueba que la convergencia puntual de las funciones de densidad implica la convergencia débil de las correspondientes distribuciones.

<u>TEOREMA I.2</u>. Sea μ una medida σ-finita sobre $(\chi,\mathcal{B})$ y sean P_r, $r \in N$ medidas de probabilidad μ-contínuas que admiten funciones de densidad f_r, $r \in N$, tales que $\forall x \in \chi$, $f_r(x) \xrightarrow[r\to\infty]{} f(x)$ para una cierta función f. Entonces

(i) Si f es una μ-densidad, se verifica que $\int |f_r - f|\, d\mu \xrightarrow[r\to\infty]{} 0$

(ii) Si existe una función g, μ integrable en χ, tal que $|f_r| < g$ $\forall r$, entonces f es una μ-densidad y, según (i), se tiene $\int |f_r - f|\, d\mu \xrightarrow[r\to\infty]{} 0$.

En cualquiera de los dos casos se sigue que $P_r \xrightarrow[r\to\infty]{d} P$ siendo P la medida de probabilidad asociada a f.

Este enunciado es algo más general que la versión original de

Scheffé (ver Rao (1973), p. 135).

Por otra parte, de $P_r \xrightarrow[r \to \infty]{d} P$ no se sigue $f_r \xrightarrow[r \to \infty]{} f$ en c.t.p. como prueba el siguiente contraejemplo:

Sea $f_r(x) = 1+\cos 2\pi rx$, para $x \in [0,1]$ y 0 en el resto ($r \in \mathbb{N}$). Es inmediato comprobar que las f_r son funciones de densidad (respecto a la medida de Lebesgue μ_L) y que sus correspondientes distribuciones convergen débilmente hacia la distribución uniforme en $[0,1]$ cuando $r \to \infty$. Sin embargo, no se verifica, obviamente, la convergencia $f_r \to f$ en c.t.p.

(b) Distancia de Prohorov

Sea $(\mathcal{X},\mathcal{B})$ definido como en (a) y sea $\mathcal{P}$ el conjunto de medidas de probabilidad sobre $(\mathcal{X},\mathcal{B})$.

Definamos, para cada $\varepsilon > 0$ y $A \in \mathcal{B}$, $A^{\varepsilon} = \{x \in \mathcal{X} \mid \inf_{y \in A} \rho(x,y) \leq \varepsilon\}$ (siendo ρ la distancia del espacio métrico $\mathcal{X}$).

Entonces, $\forall P_1,P_2 \in \mathcal{P}$ se define

$$d_p(P_1,P_2) = \inf \{\varepsilon > 0 \mid P_1(A) \leq P_2(A^{\varepsilon}) + \varepsilon, \quad \forall A \in \mathcal{B}\}$$

Es fácil comprobar que d_p es una distancia, (ver Huber (1981) p. 27). Se le denomina distancia de Prohorov.

El resultado fundamental relativo a d_p es el siguiente:

TEOREMA I.3. El conjunto $\mathcal{P}$ dotado de la distancia d_p es un espacio métrico completo y separable. Además, se verifica

$$d_p(P_r,P) \xrightarrow[r \to \infty]{} 0 \iff P_r \xrightarrow[r \to \infty]{d} P$$

La demostración puede verse en Huber (1981) pp. 28 y 29.

Este teorema muestra que d_p proporciona un concepto "natural" de proximidad entre distribuciones.

Cuando $\chi = \mathbb{R}$, la distancia de Lévy (Huber (1981), p. 25) define la misma topología que d_p (llamada topología débil) y es más manejable. Sin embargo, d_p (además de su mayor generalidad) tiene la ventaja de admitir una interesante interpretación estadística que es consecuencia del siguiente teorema de Strassen (1965):

<u>TEOREMA I.4</u>. Las dos afirmaciones siguientes son equivalentes:

(i) $P_1(A) \leq P_2(A^\delta) + \varepsilon \quad \forall A \in \mathcal{B}$

(ii) Existen dos variables aleatorias (dependientes) ξ y η con valores en χ, cuyas distribuciones inducidas son P_1 y P_2 respectivamente, tales que $P\{\rho(\xi,\eta) \leq \delta\} \geq 1-\varepsilon$.

Por tanto, si, en un problema concreto, consideramos que P_2 es la distribución que corresponde al modelo teórico ideal y P_1 es la distribución "verdadera", entonces $d_p(P_1,P_2) \leq \varepsilon$ significa que existe una variable aleatoria (inobservable) η y otra variable aleatoria (observable) ξ con distribuciones respectivas P_2 y P_1 tales que $P\{\rho(\xi,\eta) \leq \varepsilon\} \geq 1-\varepsilon$. Esta relación establece de forma explícita que los errores cometidos al "aproximar" η por ξ serán pequeños con alta probabilidad.

En definitiva, por las razones apuntadas, la distancia de Prohorov desempeña un papel fundamental.

Cuando $\chi = \mathbb{R}^n$ y las distribuciones de probabilidad P_1, P_2

vienen caracterizadas por sus funciones de distribución F_1, F_2 se emplea también la notación $d_p(F_1,F_2)$ para designar a $d_p(P_1,P_2)$.

(c) <u>Fórmula de Bayes</u>

Sean χ y Ω espacios métricos completos y separables, $\mathcal{B}$ y $\mathcal{B}'$ las σ-álgebras en χ y Ω generadas por las respectivas topologías, μ una medida σ-finita sobre $(\chi,\mathcal{B})$, π una medida de probabilidad sobre $(\Omega,\mathcal{B}')$ y $\{P_\theta\}_{\theta\in\Omega}$ una familia de medidas de probabilidad en $(\chi,\mathcal{B})$ tal que $\forall\theta \in \Omega$, P_θ tiene μ-densidad $f(.|\theta)$ siendo $f(x|\theta)$ función $\mathcal{B} \times \mathcal{B}'$-medible.

En estas condiciones (que se supondrán válidas implícitamente en lo que sigue), $\forall x \in \chi$ tal que $0 < \int_\Omega f(x|\theta)d\pi(\theta) < \infty$, la expresión,

$$\pi(\theta|x) = \frac{f(x|\theta)}{\int_\Omega f(x|\theta)d\pi(\theta)} \qquad (1)$$

es la π-función de densidad correspondiente a una distribución π-continua en Ω (distribución a posteriori).

En De Robertis y Hartigan (1979) puede verse un enunciado general del teorema de Bayes que permite interpretar rigurosamente en términos probabilísticos la expresión (1).

Cuando $\Omega = R^k$ y π es absolutamente contínua con densidad p_π respecto a la medida de Lebesgue μ_L, aparece la expresión usual

$$p(\theta|x) = \frac{f(x|\theta)p_\pi(\theta)}{\int_\Omega f(x|\theta)p_\pi(\theta)d\theta}$$

que corresponde a la μ_L-densidad de la distribución a posteriori.

En lo sucesivo utilizaremos preferentemente la expresión (1) co rrespondientes a la π-densidad.

La condición $0 < \int_{\Omega} f(x|\theta)d\pi(\theta) < \infty$ no es restrictiva desde el punto de vista práctico, pues $\int_{\Omega} f(x|\theta)d\pi(\theta) = 0$ indicaría que la información muestral y la información a priori son "incompatibles" en el sentido de que las densidades $f(.|\theta)$ se anulan en x para casi todo valor de θ (respecto a π). Por otra parte, $\mu\{x \mid \int_{\Omega} f(x|\theta)d\pi(\theta) = \infty\} = 0$ (ver De Robertis y Hartigan (1979)).

3. ROBUSTEZ EN INFERENCIA FRECUENTISTA

La llamada "Teoría de la Robustez" constituye, en opinión de Tukey, una "tercera generación de la Estadística" después de la Estadística paramétrica y la no paramétrica. Su objetivo, a grandes rasgos, es enfrentarse a las limitaciones de la Inferencia tradicional mediante la obtención de métodos estadísticos que conserven buenas propiedades cuando las hipótesis iniciales (sobre todo la distribución básica) sufren pequeñas alteraciones en un "entorno" del modelo ideal.

En los últimos años se ha publicado una enorme cantidad de material sobre estos temas adoptando, en la mayor parte de los casos, el enfoque clásico o "frecuentista".

En el libro de Huber (1981) se ofrece una excelente visión general de esta teoría. Otros compendios de menor extensión y con diferentes puntos de vista pueden encontrarse en Hampel (1973), Huber (1977) y Hogg (1979).

No se pretende exponer aquí un resumen detallado de la teoría (lo cual excedería los límites y objetivos de este trabajo), sino úni camente señalar, en forma sumaria, sus principales líneas de desarrollo, dedicando una especial atención a la "Teoría de la Robustez Cualitativa" de Hampel que presenta ciertas afinidades de planteamiento general con nuestro enfoque de la robustez en inferencia bayesiana desarrollado en los capítulos II y III de esta memoria.

Puede mencionarse en primer lugar un esquema de trabajo bastante generalizado que consiste en:

(a) Proponer estimadores que son intuitivamente "robustos" (generalmente, en el sentido de que varianza sufre pocas variaciones cuando se altera la distribución básica).

(b) Estudiar, si es factible, sus propiedades asintóticas.

(c) Frecuentemente las propiedades del estimador para tamaños muestrales finitos (por ejemplo su varianza bajo distintas distribuciones básicas) son difíciles de estudiar por cálculo directo, lo cual conduce al uso de <u>métodos de simulación</u> como el único camino viable para estudiar dichas propiedades.

Esta metodología, que introduce una componente "experimental" en la Teoría Estadística, ha permitido obtener resultados valiosos e interesantes, sobre todo desde el punto de vista de las aplicaciones prácticas. El trabajo más importante en esta línea se llevó a cabo en el Seminario de Robustez de Princeton (Andrews et. al (1972)) y consistió en analizar mediante simulación las propiedades de los principales estimadores de parámetros de posición. El resultado fue una especie de "guia práctica" para el uso de los estimadores en distintas situaciones.

Los trabajos de Huber (1964, 1972, 1981) sobre "robustez minimax" constituyen otra vertiente de la teoría con un enfoque muy distinto, de mayor elaboración matemática. Su planteamiento es, en resumen, el siguiente:

Sea C un conjunto convexo y compacto de distribuciones F sobre la recta real (completada); por ejemplo C puede ser el conjunto de todas las normales "ε-contaminadas", es decir, el conjunto de todas las distribuciones del tipo $(1-\varepsilon)\phi + \varepsilon H$ donde $\varepsilon \in [0,1)$ es una constante prefijada, ϕ es la función de distribución de la normal $N(0,1)$ y H varía en el conjunto de todas las distribuciones simétricas. El parámetro de interés es θ en $F_\theta(x)=F(x-\theta)$ $(F\in \mathcal{B})$

Se trata de hallar la sucesión $\{T_n\}_{n\in N}$ de estimadores de posición tal que el supremo sobre C de su varianza asintótica sea lo menor posible. Se consideran únicamente distribuciones simétricas y estimadores invariantes por traslación.

Utilizando técnicas de cálculo variacional se calcula la distribución F_o "más desfavorable" de C en el sentido de minimizar la información de Fisher $I(F)$ para $F \in C$. Entonces la sucesión de estimadores de máxima verosimilitud para F_o proporciona la solución del problema.

Surgen así los llamados M-estimadores que son generalizaciones de los estimadores de máxima verosimilitud que, a cambio de una cierta pérdida de eficiencia asintótica, consiguen una mejora en las cualidades de robustez.

La teoría de Huber es "cuantitativa" en el sentido de que no se ocupa de los efectos de alteraciones "infinitesimales" en la distribución básica en contraste con el enfoque "cualitativo" de Hampel que se expondrá en el apartado I.3.1.

Aparte de los desarrollos mencionados hay otras técnicas particulares que no se comentarán aquí (estimadores "jackk-nife", adaptatividad, L y R-estimadores...) y que se consideran incluídas, generalmente, entre los métodos robustos.

3.1. La teoría de la robustez cualitativa de Hampel

El primer y más importante intento de formalizar el concepto de robustez se debe a Hampel (1968, 1971). A continuación se resumen sus ideas principales.

Sean $(\chi, \mathcal{B})$ un espacio medible donde χ es un espacio métrico completo y separable y $\mathcal{B}$ denota el σ-álgebra generada por la topología, y $\mathcal{F}$ la familia de medidas de probabilidad sobre $(\chi, \mathcal{B})$.

Se dice que $\{T_n\}_{n \in \mathbb{N}}$ es una sucesión de estimadores si cada T_n es una función medible $T_n : \mathcal{F}_n \longrightarrow R^k$, siendo $\mathcal{F}_n$ el conjunto de medidas de probabilidad "empíricas" de la forma $F_n = (1/n) \sum_{i=1}^{n} \delta_{x_i}$ (δ_x denota la medida de probabilidad "causal" en $x \in \chi$).

En general, el valor de $T_n(F)$ para $F \in \mathcal{F}_n$ dependerá no sólo de F sino también de n; sin embargo, a veces se tiene que $F \in \mathcal{F}_n \cap \mathcal{F}_m$ implica $T_n(F) = T_m(F)$. Esto ocurre en el importante caso particular de que exista una aplicación $T: \mathcal{F} \longrightarrow R^k$ tal que T_n sea la restricción de T a $\mathcal{F}_n$ $(T_n = T|\mathcal{F}_n)$.

En lo sucesivo, consideremos que $x_1, x_2, \ldots$ son observaciones independientes idénticamente distribuidas (i.i.d.) según la distribución básica F, y F_n es la distribución empírica (aleatoria) correspondiente a las n primeras observaciones. Dada una sucesión de estimadores $\{T_n\}_{n \in N}$, Hampel propone la siguiente definición de robustez:

<u>DEFINICION I.1</u>. Se dice que la sucesión $\{T_n\}_{n \in N}$ es robusta en F si $\forall \varepsilon > 0 \quad \exists \delta > 0, \mid \forall G, \quad \forall n : d_p(F,G) < \delta \Longrightarrow d_p(\mathcal{L}_F(T_n), \mathcal{L}_G(T_n)) < \varepsilon$ donde $\mathcal{L}_F(T_n)$, $\mathcal{L}_G(T_n)$ denotan las medidas de probabilidad en R^k inducidas por T_n bajo F y G respectivamente (i.e. las distribuciones de T_n bajo F y G).

Esta definición tiene un obvio significado intuitivo análogo a la idea de continuidad. La relación entre estos dos conceptos puede formularse explícitamente para una sucesión genérica de estimadores (ver Hampel (1971), teorema 1). No obstante, aquí nos limitaremos al caso particular ya mencionado, de que exista un funcional $T : \mathcal{F} \longrightarrow R^k$ tal que $T_n = T|\mathcal{F}_n \quad \forall n$. Pueden encontrase multitud de ejemplos prácticos en los que los estimadores son de este tipo (ver Hampel (1968, 1971, 1974), Huber (1981), Denian et al. (1977)).

Se dice entonces que la sucesión de estimadores $\{T_n\}_{n \in N}$ con $T_n = T|\mathcal{F}_n$, es consistente en F cuando $T_n\ (F_n) \xrightarrow{P} T(F)$. Esta definición puede relacionarse con la noción usual teniendo en cuenta que, cuando la distribución básica $F = F_\theta$ depende de un parámetro que se quiere estimar, es frecuente que se verifique $T(F_\theta) \equiv \theta$ ("consistencia-Fisher").

A continuación se enuncian algunos de los resultados principales de la teoría. Las demostraciones pueden encontrarse en Hampel (1971).

El calificativo "contínuo" aplicado a un funcional $T : F \longrightarrow \mathbb{R}^k$ se entenderá referido a la continuidad débil (respecto a la topología débil inducida por d_p en F).

<u>TEOREMA I.5</u>. Sea el producto cartesiano $\chi^n = \chi \overset{n}{\times \ldots \times} \chi$ dotado de la topología producto y con la métrica dada por el máximo de la distancia entre las coordenadas.

Sea $T : F \longrightarrow \mathbb{R}^k$ tal que T es contínua en F y $T_n = T|F_n$ considerada como función en χ^n, es contínua con respecto a la métrica definida en χ^n para todo n.

Entonces la sucesión $\{T_n\}_{n\in\mathbb{N}}$ es robusta en F.

La robustez y la consistencia de $\{T_n\}_{n\in\mathbb{N}}$ (con $T_n = T|F_n$) en F no implican la continuidad de T en F, aunque si es cierto el siguiente resultado:

<u>TEOREMA I.6</u>. Supongamos que $\{T_n\}_{n\in\mathbb{N}}$, con $T_n = T|F_n$, es robusta en F y consistente para todo G perteneciente a un entorno de F.

Entonces T es contínuo en F.

La relación más clara entre robustez y continuidad viene dada por el siguiente teorema en el que ambos conceptos se consideran globalmente.

<u>TEOREMA I.7</u>. Sea $T : F \longrightarrow \mathbb{R}^k$ y consideremos la sucesión $\{T_n\}_{n\in\mathbb{N}}$ definida por $T_n = T|F_n$ para todo n.

Entonces T es contínuo en todo F si y sólo si $\{T_n\}_{n\in\mathbb{N}}$ es robusta y consistente en F para toda F.

La teoría de Hampel, que se ha resumido aquí en los aspectos que más nos interesan, tiene una gran riqueza de conceptos que ayudan a clarificar el significado profundo de algunas nociones estadísticas.

Por ejmplo, la idea de considerar estimadores definidos como funcionales, permite, no sólo analizar rigurosamente la robustez en términos de continuidad, sino también utilizar métodos de cálculo diferencial para funcionales (derivadas de Fréchet y Gateaux). Así surge la llamada "curva de influencia" que sirve para "medir", en cierto sentido, la robustez local de los estimadores y como herramienta auxiliar en la construcción de estimadores con propiedades prefijadas (Hampel (1968, 1974)). También, bajo determinadas hipótesis, se puede estudiar el comportamiento asintótico de los estimadores mediante técnicas análogas a los desarrollos de Taylor aplicadas a los funcionales (Serfling (1980)).

Destaquemos, por último, una cuestión interesante: En lo que respecta a los tests de hipótesis y a la estimación por intervalos el desarrollo general de la Teoría de la Robustez es mucho menor que el correspondiente a la estimación puntual. Concretamente no existe una teoría cualitativa general análoga a la desarrollada para estimadores. En el caso de los tests, el tratamiento natural sería estudiar la estabilidad de la función de potencia lo que llevará generalmente a cálculos muy complicados. Nuevamente aquí resultan más asequibles los métodos de simulación.

4. ROBUSTEZ EN INFERENCIA BAYESIANA

En un problema general bayesiano hay tres elementos que son susceptibles de sufrir alteraciones o errores de determinación: La verosimilitud, la distribución a priori y la función de pérdida. Es natural preguntarse por los efectos que ocasionan en los métodos estadísticos las posibles alteraciones en estos tres elementos.

El estudio de estos temas está todavía en sus comienzos y la bibliografía dedicada a ellos es relativamente escasa en comparación con la existente sobre robustez "frecuentista".

En esta sección se dará un resumen (que no pretende ser exhaustivo) de los resultados disponibles en la actualidad poniendo especial énfasis en aquellos que están más relacionados con el contenido de esta memoria (bien sea por referirse a cambios en la verosimilitud o bien por adoptar un enfoque "cualitativo" inspirado en la noción de continuidad).

4.1. Efectos de la alteraciones en la verosimilitud

La aportación más conocida en este aspecto de la teoría se debe a Box y Tiao (1962, 1964, 1968, 1973). Su planteamiento general es bastante sencillo y puede resumirse así:

(a) En muchos problemas de inferencia se supone que la distribución básica es normal. Podemos relajar esta restricción considerando que la verdadera densidad básica pertenece a la familia definida por

$$f_\alpha(x|\theta,\sigma) = C_\alpha \exp\left[-\frac{1}{2}\left|\frac{x-\theta}{\sigma}\right|^{2/(1+\alpha)}\right], \quad \text{para} \quad x \in R, \quad \text{con} \quad \theta \in R,$$

y $\sigma \in (0,\infty)$

siendo $C_\alpha^{-1} = \Gamma[1+\frac{1}{2}(1+\alpha)]2^{[1+\frac{1}{2}(1+\alpha)]}\sigma$, para $\alpha \in (-1,1]$.

Así, para $\alpha = 0$ se tiene la densidad normal, para $\alpha = 1$ la doble exponencial, y cuando $\alpha \to -1$ puede probarse que la distribución tiende a la uniforme, pues

$$\lim_{\alpha \to -1} f_\alpha(x|\theta,\sigma) = \frac{1}{2\sqrt{3}\sigma} \quad \text{para} \quad x \in (\theta - \sqrt{3}\sigma, \ \theta + \sqrt{3}\sigma)$$

La interpretación estadística de las densidades f_α es clara: Se consideran desviaciones de la normalidad debidas a alteraciones en la kurtosis. Esta es, junto con las alteraciones en la simetría, la principal manifestación, en la práctica, de no normalidad. Para $\alpha > 0$ las distribuciones son leptokúrticas y para $\alpha < 0$ son platikúrticas.

(b) Las distribuciones a priori sobre θ y σ se eligen impropias con "densidades" del tipo

$$\pi(\theta) \propto C \qquad \text{y} \qquad \pi(\sigma) \propto \sigma^{-1}$$

(c) El modelo planteado en (a) y (b) se aplica al estudio de dos problemas concretos de inferencia: La estimación del parámetro de posición θ y la comparación de las varianzas σ_1^2 y σ_2^2 correspondientes a dos distribuciones independientes con densidades $f_\alpha(\cdot|\theta_i,\sigma_i)$, $i=1,2$.

(d) En lo que respecta al primer problema, se calcula la densidad a posteriori conjunta (respecto a μ_L) $\pi(\theta,\sigma|x,\alpha_0)$ para un $\alpha_0 \in (-1,1]$ y para una muestra $x = (x_1,\ldots,x_n)$ prefijados y, a partir de ella, la marginal de θ, que resulta ser

$$\pi(\theta|x,\alpha_o) = K_{\alpha_o} \left[M(\theta)\right]^{-\frac{1}{2}n(\alpha_o+1)}$$

donde $M(\theta) = \left[\sum_i [x_i - \theta]^{2/(1+\alpha_o)}\right]$

y $K_{\alpha_o}^{-1} = \int_{-\infty}^{+\infty} \left[M(\theta)\right]^{-\frac{1}{2}n(1+\alpha_o)} d\theta$ es la constante de normalización.

Para $\alpha_o = 0$ se comprueba que

$$\frac{\theta-\bar{x}}{s/\sqrt{n}} \sim t_{n-1}$$

siendo s la cuasivarianza muestral.

Sin embargo, para un α_o arbitrario, $\pi(\theta|x,\alpha_o)$ no puede expresarse generalmente en términos elementales, lo cual dificulta la comparación de estas densidades para distintos valores de α. En el artículo de Box y Tiao (1962) el estudio se limita a la obtención de algunas propiedades de las $\pi(\theta|x,\alpha_o)$. Además, para una muestra numérica concreta procedente de un determinado experimento biológico se obtienen las gráficas aproximadas de $\pi(\theta|x,\alpha)$ para distintos valores de α observándose que hay una variación bastante acusada debida fundamentalmente al desplazamiento que se produce en el centro de estas gráficas al variar α.

(e) Un paso posterior en el análisis consiste en definir una distribución a priori sobre α con densidad:

$$\pi(\alpha) = \omega_a(1-\alpha^2)^{a-1}, \quad \text{para} \quad \alpha \in (-1,1], \quad \text{siendo}$$

$$\omega_a = \Gamma(2a)\left[\Gamma(a)\right]^{-2} e^{-(2a-1)}, \quad a \geq 1$$

y hallar la densidad a posteriori $\pi(\theta|x)$ para distintos valores de la constante a. Haciendo estos cálculos para la muestra numérica mencionada, se obtiene que las inferencias son esencialmente las mismas que corresponderían al caso normal con $\alpha = 0$.

(f) En el problema de comparación de varianzas, se consideran dos muestras $x_{(i)} = (x_{i1},\ldots,x_{in_i})$, $i=1,2$ procedentes de dos distribuciones independientes con densidades $f_\alpha(\cdot|\theta_i,\sigma_i)$, $i=1,2$, y, tomando para θ_i y σ_i las mismas distribuciones a priori que en el caso anterior, se calcula $\pi(\sigma_1,\sigma_2|\alpha,\theta_1,\theta_2,x_{(1)},x_{(2)})$ y, a partir de aquí, $\pi(\sigma_2^2/\sigma_1^2|\alpha,x_{(1)},x_{(2)})$, que cambian notablemente al variar α.

Nuevamente se considera la distribución a priori sobre α, $\pi(\alpha) \propto (1-\alpha^2)^{a-1}$ y se comprueba que las inferencias realizadas a partir de la densidad $\pi(\sigma_2^2/\sigma_1^2 \mid x_{(1)},x_{(2)})$ no difieren esencialmente de las que corresponderían al caso normal con $\alpha = 0$ para muestras numéricas concretas $x_{(1)}$ y $x_{(2)}$.

Como puede verse, este enfoque es eminentemente cuantitativo y, de hecho, ciertas conclusiones se deducen para muestras particulares.

Utilizando una metodología similar (Box-Tiao 1968) se aborda también el problema de los "outliers", u observaciones "anómalas", en el modelo lineal.

4.2. El problema de sensibilidad

En lo sucesivo utilizaremos preferentemente el término "sensibilidad" en el caso de que sea la distribución a priori el elemento sometido a variación.

De los tres problemas mencionados al comienzo de esta sección éste es quizá el que ha sido estudiado con mayor amplitud.

La línea de trabajo más usual tiene como punto de partida el llamado "Principio de la Medida Precisa" (o "de la Estimación Estable") debido a Edwards, Lindman y Savage (ver 4.2.1).

En el subapartado 4.2.2 se expone un enfoque cualitativo que estudia el problema de sensibilidad mediante el concepto de continuidad y que, por tanto, presenta un mayor interfes en relación con la metodología de este trabajo.

4.2.1. Enfoque cuantitativo: el principio de la medida precisa

Sean μ y ν dos medidas σ-finitas sobre $(\chi,\mathcal{B})$ y $(\Omega,\mathcal{B}')$ respectivamente, $\{f(.|\theta)\}_{\theta\in\Omega}$ una familia de μ-densidades en $(\chi,\mathcal{B})$, p una ν-densidad a priori en Ω que se supone acotada y, para $x \in \chi$

$$p(\theta|x) = \frac{f(x|\theta)p(\theta)}{\int_\Omega f(x|\theta)p(\theta)d\nu(\theta)} , \quad \phi(\theta|x) = \frac{f(x|\theta)}{\int_\Omega f(x|\theta)d\nu(\theta)}$$

(supuesto que ambos denominadores son finitos y mayores que 0), las ν-densidades a posteriori correspondientes a elegir la distribución a priori p y la uniforme (propia o impropia).

En esta sitaución el llamado "Principio de la Medida Precisa" viene dado por el siguiente

<u>TEOREMA I.8</u>. Sea $A \in \mathcal{B}'$ tal que $m = \inf_{\theta\in A} p(\theta) > 0$, sean $\alpha \in [0,1)$, $\beta,\gamma \geq 0$ cumpliendo

$$\int_A \phi(\theta|x)d\nu(\theta) \geq 1-\alpha, \quad \sup_{\theta\in A} p(\theta) \leq (1+\beta)m \quad y \quad \sup_{\theta\in A^c} p(\theta) \leq (1+\gamma)m$$

Entonces

$$\int_{\Omega} |p(\theta|x) - \phi(\theta|x)| d\nu(\theta) \leq \varepsilon$$

siendo $\varepsilon = \max \{\frac{\alpha+\beta}{1-\alpha}, \frac{\alpha+\beta+\alpha\gamma}{1+\alpha+\beta+\alpha\gamma}\} + \frac{\alpha(2-\alpha+\gamma)}{1-\alpha}$

La demostración puede encontrarse en De Groot (1970). En Dickey (1976) se ofrece una versión más general que no comentaremos aquí ya que solo nos interesa destacar el planteamiento fundamental del método.

Como consecuencia del teorema I.8 se tiene que para cualquier $B \in \mathcal{B}'$, la diferencia entre las probabilidades a posteriori de B calculadas con $\phi(.|x)$ y $p(.|x)$ es menor o igual que ε. En definitiva, el Principio de la medida precisa ofrece un resultado de sensibilidad desde un punto de vista cuantitativo prescindiendo de consideraciones topológicas.

4.2.2. Enfoque cualitativo

Con la notación usual, sea χ el espacio muestral, Ω el espacio paramétrico, μ una medida σ-finita en $(\chi, \mathcal{B})$, P el conjunto de medidas de probabilidad sobre Ω (dotado de la topología débil) y $\{P_\theta\}_{\theta \in \Omega}$ una familia de medidas de probabilidad μ-contínuas en $(\chi, \mathcal{B})$ con μ-densidades $f(.|\theta)$.

Para $\pi \in P$, π^x denotará la medida de probabilidad "a posteriori" sobre Ω, cuya π-densidad viene dada por $\pi(\theta|x) =$

$$= [\int_{\Omega} f(x|\theta) d\pi(\theta)]^{-1} f(x|\theta), \quad \forall x \in \chi_\pi = \{x \mid 0 < \int_{\Omega} f(x|\theta) d\pi(\theta) < \infty\}.$$

En estas condiciones, podemos enunciar el siguiente

<u>TEOREMA I.9</u>. Para $x \in \chi$ y $\pi_0 \in \tilde{P}$, definamos el "operador Bayes asociado a x y π_0", $B_{x,\pi_0} : P \longrightarrow P$, por:

$$B_{x,\pi_0}(\pi) = \begin{cases} \pi^x , & \text{si} \quad x \in \chi_\pi \\ \pi_0 , & \text{si} \quad x \notin \chi_\pi \end{cases}$$

Supongamos que, $\forall x \in \chi$, $f(x|\theta)$ es función contínua y acotada en Ω. Entonces, $\forall x \in \chi_{\pi_0}$ el operador B_{x,π_0} es contínuo en π_0.

<u>Demostración</u>:

Sea $\{\pi_r\}_{r \in \mathbb{N}} \subset V$ tal que $\pi_r \xrightarrow[r \to \infty]{d} \pi_0$. Para $x \in \chi_{\pi_0}$ se tiene (por ser $f(x|\theta)$ contínua y acotada):

$$\int_\Omega f(x|\theta)d\pi_r(\theta) \xrightarrow[r \to \infty]{} \int_\Omega f(x|\theta)d\pi_0(\theta) > 0 \qquad (1)$$

y además $\exists r_0 \in \mathbb{N}$ tal que $\forall r \geq r_0 \quad x \in \chi_{\pi_r}$.

Por otra parte, para toda función $u : \Omega \longrightarrow \mathbb{R}$ contínua y acotada, $u(\theta)f(x|\theta)$ también lo es y así

$$\int_\Omega u(\theta)f(x|\theta)d\pi_r(\theta) \xrightarrow[r \to \infty]{} \int_\Omega u(\theta)f(x|\theta)d\pi_0(\theta)$$

que, junto con (1), prueba $B_{x,\pi_0}(\pi_r) \xrightarrow[r \to \infty]{d} B_{x,\pi_0}(\pi_0)$.

En Ríos (1981) puede encontrarse una versión ligeramente distinta de este resultado. Aquí lo hemos establecido en la forma que parece más adaptada a nuestro enfoque posterior. La interpretación intuitiva es sencilla: Si la distribución a priori "teórica" está suficientemente próxima a la "verdadera", las correspondientes distribuciones a posteriori también lo estarán. Este teorema ofrece, por tanto, una garantía frente a las pequeñas perturbaciones en la distribución a priori en contraste con el Principio de la Medida Precisa que se ocupa de cuantificar el efecto de elegir una u otra de dos posibles distribuciones a priori.

4.3. Estabilidad frente a los cambios en la función de pérdida

Hay un interesante artículo de Kadane y Chuang (1978) en el que se estudia el efecto de los cambios simultáneos en la función de pérdida y la distribución a priori en un problema general de decisión. Como veremos, su planteamiento presenta también ciertas analogías de fondo con la "sensibilidad cualitativa" del apartado anterior y con la teoría de la estabilidad bayesiana que desarrollaremos en los capítulos II y III de esta memoria.

Pasamos a resumir el contenido del mencionado artículo:

Consideremos un problema genérico de decisión con espacio paramétrico $\Omega \subset \mathbb{R}^k$ y espacio de decisiones $\mathcal{D} \subset \mathbb{R}^l$. Sean L_o y π_o la función de pérdida y la distribución a priori "teóricas" (y, por tanto, "aproximadas") que utilizamos en el problema. La pérdida esperada (que se supone finita) es, por tanto, $W_o = \inf_{D \in \mathcal{D}} \int_\Omega L_o(\theta,D)d\pi_o(\theta)$.

Entonces, para cada $\varepsilon > 0$, hay una decisión $D_o(\varepsilon)$ que es ε-óptima, es decir:

$$\int_\Omega L_o(\theta,D_o(\varepsilon))d\pi_o(\theta) \leq W_o + \varepsilon \qquad (1)$$

En estas condiciones Kadane y Chuang proponen el siguiente concepto de estabilidad en (L_o,π_o):

DEFINICION I.2. Sea dice que el par (L_o,π_o) es fuertemente estable cuando, para cada sucesión $\{(L_r,\pi_r)\}_{r\in\mathbb{N}}$ tal que

$$L_r \xrightarrow[r\to\infty]{} L_o \text{ uniformemente y } \pi_r \xrightarrow[r\to\infty]{d} \pi_o$$

se verifica que

$$\lim_{\varepsilon\downarrow o}\ \limsup_{r\to\infty} \left[\int_\Omega L_r(\theta,D_o(\varepsilon))d\pi_r(\theta) - \inf_{D\in\mathcal{D}}\int_\Omega L_r(\theta,D)d\pi_r(\theta)\right] = 0 \tag{2}$$

para cualquier $D_o(\varepsilon)$ que verifique (1).

Esta definición tiene sentido ya que el $\limsup_{r\to\infty}$ de la expresión encerrada entre corchetes es función de ε monótona no creciente y acotada inferiormente por 0.

Cuando la expresión (2) se verifica solamente para alguna elección particular de $D_o(\varepsilon)$, diremos que el par (L_o,π_o) es débilmente estable y la $D_o(\varepsilon)$ que cumple la expresión (2) se denominará decisión estabilizadora.

Cuando (L_o,π_o) no es estable en ninguno de los dos sentidos mencionados, se dice que es inestable.

El significado intuitivo de estos conceptos puede expresarse así: Si el "verdadero" (L_r,π_r) está "próximo" al teórico (L_o,π_o) que se maneja, entonces la pérdida "real" $W_r' = \int_\Omega L_r(\theta,D_o(\varepsilon))d\pi_r(\theta)$ ocasionada por la decisión elegida $D_o(\varepsilon)$ está próxima a la pérdida $W_r = \inf_{D\in\mathcal{D}}\int_\Omega L_r(\theta,D)d\pi_r(\theta)$ que se tendría utilizando el verdadero (L_r,π_r). En el caso de estabilidad débil, el decisor debe tomar la decisión estabilizadora $D_o(\varepsilon)$ para evitar errores grandes.

En el artículo de Kadane y Chuan se enuncia una versión equivalente de la definición y se proponen otras dos definiciones algo más restrictivas (equivalentes entre sí) que no se comentarán aquí pues se inspiran esencialmente en las mismas ideas de la versión enunciada.

Se demuestra también una condición suficiente de estabilidad fuerte dada por el siguiente

TEOREMA I.10. Con el planteamiento y notación indicados más arriba, supongamos que:

(i) Existe $C > 0$ tal que $|L_o(\theta,D)| \leq C$ para todo θ y D

(ii) $L_o(\theta,D)$ es contínua en θ uniformemente en D.

Entonces (L_o,π_o) es fuertemente estable.

(Se dice que una función $f(x,y)$ es "continua en x uniformemente en y" cuando $\forall\varepsilon > 0$, $\forall x$ $\exists\delta > 0$ tal que $\forall y$, $|x-x_o| < \delta \Longrightarrow |f(x,y) - f(x_o,y)| < \varepsilon$).

Por último, para el caso en $\Omega = \mathcal{D} = R$ y L_o sea de la forma $L_o(\theta,D) = h(\theta-D)$, se demuestran otras dos condiciones de estabilidad fuerte imponiendo ciertas hipótesis sobre h.

CAPITULO II

CAPITULO II

ESTABILIDAD RESPECTO A LA VEROSIMILITUD EN INFERENCIA BAYESIANA.

1. SUMARIO.

En este capítulo se exponen las ideas esenciales del presente trabajo. Así, en la sección 2, después de unos comentarios preliminares y de establecer rigurosamente los elementos del problema, se enuncia la definición fundamental de estabilidad en un modelo bayesiano (apartado 2.3). A continuación se discuten algunos aspectos formales de esta definición y se analiza su significado estadístico y matemático a través de los teoremas II.1 y II.2 que permiten interpretarla en términos de continuidad. Finalmente, en el apartado 2.6 se incluye un ejemplo de modelo inestable.

La sección 3 está dedicada a estudiar los procedimientos matemáticos que nos permitirán aplicar en la práctica nuestro concepto de estabilidad. En concreto, se demuestran cuatro teoremas que ofrecen condiciones suficientes para asegurar la estabilidad de un modelo en diversas situaciones.

Por último, en la sección 4 se utilizan los resultados de la sección anterior para demostrar la estabilidad del modelo de Box -Tiao y de otros modelos alternativos definidos mediante mixturas infinitas de distribuciones

2. EL CONCEPTO DE ESTABILIDAD EN UN MODELO BAYESIANO.

En el capítulo anterior vimos que el problema de "sensibilidad" (distribución a priori variable, verosimilitud fija) puede analizarse facilmente a través de la continuidad del operador Bayes considerado como un funcional que depende de la distribución a priori. Si intentamos trasladar estas ideas al estudio de la "robustez" (distribución a priori fija, verosimilitud variable) encontramos algunas dificultades que obligan a plantear el problema en forma muy distinta.

Esta sección está dedicada precisamente a proponer y analizar una definición de estabilidad o robustez cualitativa en un modelo bayesiano. En situaciones muy generales este concepto podrá relacionarse también con el de continuidad respecto a la topología débil.

En lo referente a la terminología, emplearemos con preferencia la palabra "estabilidad" para referirnos a la robustez genérica o primaria, propia del modelo, reservando el término "robustez" para referirnos al concepto específico o secundario (propio de estimadores, test de hipótesis y regiones de confianza) que se introducirá en el capítulo siguiente.

2.1. Justificación intuitiva

Para presentar intuitivamente los conceptos que se darán a continuación consideremos, por ejemplo, el modelo de Box - Tiao estudiado en el capítulo I: Disponemos de una muestra aleatoria $x = (x_1,\ldots,x_n)$ y queremos obtener la distribución a posteriori de θ pero no sabemos con certeza cuál es la distribución que ha generado la muestra x; en otras palabras, ignoramos cuál es la verosimilitud exacta que se debe emplear en la correspondiente fórmula de Bayes, aunque la naturaleza del problema nos permite suponer que dicha verosimilitud pertenece a una determinada familia indexada por el parámetro α.

En esta situación tiene sentido plantearse el siguiente problema: Supongamos que la distribución "teórica" (que en general, será solo una aproximación de la "verdadera") es lo que corresponde a $\alpha = \alpha_o$; si la distribución verdadera pertenece a la misma familia y está suficientemente "próxima" a ella, ¿podemos asegurar que la distribución a posteriori teórica estará próxima a la verdadera?. En caso afirmativo, diremos que el modelo es estable o cualitativamente robusto.

Como puede verse, se tratará más bien de una robustez frente a cambios "infinitesimales". Los conceptos implicados en ella serían en principio, bastante análogos a los que se utilizan para estudiar la continuidad de una función.

Pasamos ahora a generalizar y formalizar estas ideas.

2.2. Planteamiento y notación

Sea un espacio paramétrico $\Omega \subset R^k$ sobre el cual se supone definida una medida de probabilidad π (distribución a priori).

Sea $\mathcal{U} = \{F_\alpha\}_{\alpha \in I}$ donde, para cada $\alpha \in I$, F_α es una aplicación

$$F_\alpha : R^n \times \Omega \longrightarrow [0,1]$$
$$(x,\theta) \longrightarrow F_\alpha(x|\theta)$$

tal que, $\forall \theta \in \Omega$ la aplicación inducida $F_\alpha(.|\theta)$ es una función de distribución en R^n.

Supondremos en lo sucesivo que la parametrización elegida "distingue distribuciones" en el sentido de que $F_{\alpha_1}(.|\theta_1) \neq F_{\alpha_2}(.|\theta_2)$ siempre que $(\alpha_1,\theta_1) \neq (\alpha_2,\theta_2)$ y que $\forall(\alpha,\theta) \in I \times \Omega$, $F_\alpha(.|\theta)$ corresponde a una medida de probabilidad continua (dominada por la medida de Lebesgue) o discreta (dominada por una medida "de contar") sin que necesariamente sean todas las $F_\alpha(.|\theta)$ del mismo tipo.

En estas condiciones, el teorema de Radon-Nikodim asegura, para cada $F_\alpha(.|\theta)$, la existencia de una función de densidad $f_\alpha(.|\theta)$ que no es única pues está determinada salvo conjuntos de medida dominante nula. Por tanto, para cada $F_\alpha(.|\theta)$ elijamos una densidad $f_\alpha(.|\theta)$.

Tenemos así definido un conjunto $V = \{f_\alpha\}_{\alpha \in I}$ en el que cada elemento f_α es, formalmente, una aplicación de $R^n \times \Omega$ en $[0,\infty)$ aunque será útil interpretarlo más bien como el conjunto $\{f_\alpha(.|\theta),$ $\theta \in \Omega\}$ de funciones de densidad obtenidas cuando θ varía en Ω; lo representaremos indistintamente por f_α o por $f_\alpha(.|\theta)$ (entendiendo que θ es variable). Con este convenio, también usaremos las notaciones más explícitas:

$$U = \{F_\alpha(.|\theta),\ \theta \in \Omega\}_{\alpha \in I}, \qquad V = \{f_\alpha(.|\theta),\ \theta \in \Omega\}_{\alpha \in I}$$

Utilizando, por extensión, el lenguaje habitual en inferencia bayesiana llamaremos verosimilitudes a los elementos de V, aunque, más estrictamente, este término se suele emplear para designar a la función de θ definida por $f_\alpha(x|\theta)$ cuando x está fijo. En algunos casos, nosotros usaremos también la palabra "verosimilitud" en esta segunda acepción lo cual no dará lugar a confusión pues se trata de una diferencia de matiz que quedará clara a partir del contexto.

I es un conjunto arbitrario de índices. Lo más usual será consi derar $I \subset R^m$ $(m \geq 1)$ como en el modelo de Box-Tiao (ver cap. I) pero ésto no tiene porqué ocurrir así necesariamente; de hecho puede suceder que I no aparezca explícitamente, como, por ejemplo, en el caso de que V fuese de la forma $\{f(x-\theta)\}$ donde f varía en el conjunto de todas las funciones de densidad continuas en R y $\Omega = R$.

En lo sucesivo, diremos que V y la distribución a priori π determinan un modelo al que denotaremos brevemente por (π, V). Según se deduce de lo expuesto, α y θ juegan un papel totalmente distinto: θ es el parámetro de interés mientras que, en la práctica, elegir un valor concreto de α equivale a elegir una verosimilitud para, a través de la fórmula de Bayes, calcular la distribución a posteriori.

Naturalmente, al plantear un modelo concreto es importante elegir bien el conjunto V para asegurarnos de que contiene las desviaciones de la verosimilitud "teórica" que convenga estudiar.

En lo referente a la estructura de las $f_\alpha(x|\theta)$, el caso más frecuente será aquel en que $x = (x_1, \ldots, x_n)$ es una muestra correspondiente a n observaciones de variables aleatorias independientes idénticamente distribuidas (i.i.d) con función de densidad unidimensional $p_\alpha(.|\theta)$ de forma que $f_\alpha(x|\theta) = \prod_{i=1}^{n} p_\alpha(x_i|\theta)$. Sin embargo, esta situación no es la única interesante: Así, por ejemplo, en la aplicación a los modelos lineales (ver III.2) la muestra corresponde a observaciones independientes pero no i.d. En cualquier caso, no necesitaremos hipótesis de este tipo para obtener los resultados esenciales. Bastará considerar a $f_\alpha(.|\theta)$ como una función de densidad n-variante (para α y θ fijos) sobre la que impondremos cuando sea necesario hipótesis de continuidad, acotación,... etc.

El valor de $n \in \mathbb{N}$ se supone fijo en principio.

En cuanto a la forma de dependencia de cada F_α con respecto a θ, los casos más interesantes son:

(a) θ es un parámetro de posición. Pueden darse diversas variantes:

(a_1) $\theta \in \Omega \subset \mathbb{R}$, las $x_1,\ldots,x_n$ corresponden a observaciones i.i.d. y $f_\alpha(x|\theta) = \prod_{i=1}^{n} p_\alpha(x_i-\theta)$, $\alpha \in I$, donde $\forall\alpha \in I$, y $p_\alpha : \mathbb{R} \longrightarrow \mathbb{R}^+$ es una función de densidad.

(a_2) $\theta \in \Omega \subset \mathbb{R}^n$, $x = (x_1,\ldots,x_n)$ se considera como una observación de una distribución n-variante y $f_\alpha(x|\theta) = f_\alpha(x_1-\theta_1,\ldots,x_n-\theta_n)$ siendo $f_\alpha : \mathbb{R}^n \longrightarrow \mathbb{R}^+$ una función de densidad.

(a_3) $\theta \in \Omega \subset \mathbb{R}^k$, $n = mk$ $(m \in \mathbb{N})$, $x = (x_{11},\ldots,x_{1k},\ldots,x_{m1},\ldots,x_{mk})$ se considera como una muestra de tamaño m de una distribución k-variante y $f_\alpha(x|\theta) = \prod_{j=1}^{k} f_\alpha(x_{1j}-\theta_1,\ldots,x_{kj}-\theta_k)$ donde $f_\alpha : \mathbb{R}^k \longrightarrow \mathbb{R}^+$ es una función de densidad.

(b) θ es un parámetro de escala. Se presentan variantes análogas a las de (a).

(c) $\forall(\alpha,\theta) \in I \times \Omega$ la función de distribución asociada a $f_\alpha(.|\theta)$ es de la forma $F_\alpha(.|\theta) = F_\alpha(u_\theta(.))$, donde $\forall\alpha \in I$ $F_\alpha : \mathbb{R}^n \longrightarrow [0,1]$ es una función de distribución y $\forall\theta \in \Omega$ $u_\theta : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ es una función tal que $F_\alpha(u_\theta(.))$ sea también función de distribución $\forall(\alpha,\theta)$.

Obsérvese que esta situación incluye como casos particulares (a) y (b) con todas sus variantes.

No será necesario para el desarrollo de la mayor parte de la teoría suponer que V es de este tipo. De hecho, pueden darse situaciones más complicadas en las que la forma de dependencia de las F_α respecto a θ varíe para los distintos valores de α. Sin embargo, como se verá más adelante, la interpretación de los conceptos será más sencilla y natural en los casos (a), (b) ó (c) (que abarcan la gran mayoría de los problemas reales). Si nuestro modelo no es de este tipo y la estructura de V es excesivamente complicada, puede suceder que el concepto de estabilidad que se introducirá en 2.3 pierda significado estadístico aunque conserve su sentido matemático.

Representaremos por $\pi_\alpha(.|x)$ la función de densidad respecto a π de la distribución a posteriori, cuando se elige la verosimilitud dada por α. Es decir

$$\pi_\alpha(\theta|x) = \frac{f_\alpha(x|\theta)}{\int_\Omega f_\alpha(x|\theta)d\pi(\theta)}$$

supuesto que $0 < \int_\Omega f_\alpha(x|\theta)d\pi(\theta) < \infty$.

Denotaremos por $F_\pi^{(\alpha)}(.|x)$ la función de distribución correspondoente a la densidad a posteriori $\pi_\alpha(.|x)$.

La familia V se considerará fija en lo sucesivo, salvo que se indique explícitamente lo contrario. Cada elemento f_α de V lleva asociada una familia de distribuciones a posteriori

$$\{F_\pi^{(\alpha)}(.|x)\}_{x\in X_\alpha}, \quad \text{siendo } X_\alpha = \{x \in R^n \mid o < \int_\Omega f_\alpha(x|\theta)d\pi(\theta) < \infty\}$$

2.3. Definición de modelo estable

Sea el modelo definido por $V = \{f_\alpha(.|\theta),\ \theta \in \Omega\}_{\alpha \in I}$ y por la distribución a priori π sobre $\Omega \subset R^k$.

Diremos que este modelo es ESTABLE o CUALITATIVAMENTE ROBUSTO en la verosimilitud f_α cuando, para toda sucesión $\{f_{\alpha_r}\}_{r \in N} \subset V$ tal que $F_{\alpha_r}(.|\theta) \xrightarrow[r \to \infty]{d} F_\alpha(.|\theta)$ $\forall \theta \in \Omega$, se verifica que $F_\pi^{(\alpha_r)}(.|x) \xrightarrow[r \to \infty]{d} F_\pi^{(\alpha)}(.|x)$ $\forall x \in R^n$ tal que $0 < \int_\Omega f_\alpha(x|\theta)d\pi(\theta) < \infty$ y $0 < \int_\Omega f_{\alpha_r}(x|\theta)d\pi(\theta) < \infty$ $\forall r \in N$.

2.4. Discusión y análisis de la definición

2.4.1. Discusión

La condición $0 < \int_\Omega f_\alpha(x|\theta)d\pi(\theta) < \infty$ y $0 < \int_\Omega f_{\alpha_r}(x|\theta)d\pi(\theta) < \infty$ $\forall r$ se impone obviamente para que tenga sentido la fórmula de Bayes en todas las verosimilitudes de la sucesión y en la verosimilitud límite. La interpretación estadística de esta condición ya se ha comentado en I.

Puede darse el caso de que alguna verosimilitud $f_\alpha \in V$ sea "punto aislado" en el sentido de que no exista ninguna sucesión $\{F_{\alpha_r}(.|\theta)\}$ tal que $F_{\alpha_r}(.|\theta) \xrightarrow{d} F_\alpha(.|\theta)$ $\forall \theta \in \Omega$. En este caso, por razones de coherencia lógica, consideraremos que el modelo es estable en f_α, aunque es claro que esta situación no tiene interés práctico y la excluiremos implícitamente en lo sucesivo.

Hay otro aspecto que merece comentarse, al menos desde el punto de vista teórico: Con la definición propuesta en 2.3 la estabilidad depende esencialmente de la familia $\mathcal{V} = \{f_\alpha(.|\theta),\ \theta \in \Omega\}_{\alpha \in I}$ elegida para representar a $\mathcal{U} = \{F_\alpha(.|\theta),\ \theta \in \Omega\}_{\alpha \in I}$. En otras palabras, puede ocurrir que el modelo $(\pi, \mathcal{V})$ sea estable en f_α y el modelo $(\pi, \tilde{\mathcal{V}})$ sea inestable, siendo $\mathcal{V}$ y $\tilde{\mathcal{V}}$ dos representaciones de la misma familia $\mathcal{U}$. Para ilustrar ésto, consideremos el siguiente ejemplo sencillo:

Sea $\Omega = \{1,2\}$, $\pi(1) = \pi(2) = 1/2$, $n=1$, y sea $\mathcal{V}$ un conjunto de verosimilitudes correspondientes a distribuciones univariantes Lebesgue-continuas que contiene a

$$f_{\alpha_o}(x|\theta) = \theta e^{-\theta x}, \quad x > o, \ (\theta = 1,2)$$

Supongamos que f_{α_o} no es un punto aislado en $\mathcal{V}$ y que el modelo $(\pi, \mathcal{V})$ es estable en f_{α_o}.

Consideremos ahora la familia $\tilde{\mathcal{V}}$ en la que todas las verosimilitudes son iguales a las de $\mathcal{V}$ excepto la f_{α_o} que se modifica así:

$$\tilde{f}_{\alpha_o}(x|\theta) = \theta e^{-\theta x}, \quad x > o, \ x \neq 1 \quad (\theta = 1,2)$$

$$\tilde{f}_{\alpha_o}(1|1) = 2e^{-2}; \quad \tilde{f}_{\alpha_o}(1|2) = e^{-1}$$

$\mathcal{V}$ y $\tilde{\mathcal{V}}$ representan obviamente el mismo conjunto de distribuciones básicas, verificándose además:

$$\tilde{\pi}_\alpha(.|x) \equiv \pi_\alpha(.|x) \quad \forall \alpha \neq \alpha_o, \quad \forall x \in X_\alpha = \{x \quad o < \int_\Omega f_\alpha(x|\theta)\, d\pi(\theta) < \infty\}$$

$$\tilde{\pi}_\alpha(1|1) = \frac{2e^{-2}}{e^{-1} + 2e^{-2}} \neq \pi_{\alpha_o}(1|1) = \frac{e^{-1}}{e^{-1} + 2e^{-2}} = \tilde{\pi}_{\alpha_o}(2|1) \neq \pi_{\alpha_o}(2|1)$$

Así, si $(\pi, \mathcal{V})$ era estable en f_{α_o}, $(\pi, \tilde{\mathcal{V}})$ ya no lo será en f_{α_o}.

En general, para cada $\theta = 1,2$ podemos modificar todas las $f_\alpha(.|\theta)$ en conjuntos de medida de Lebesgue nula obteniéndose una nueva familia de verosimilitudes $\hat{\mathcal{V}}$ cuyas distribuciones a posteriori asociadas verifican:

$\forall\alpha \in I$, $\hat{F}_\pi^{(\alpha)}(\cdot|x) = F_\pi^{(\alpha)}(\cdot|x)$ $\forall x \in X_\alpha$ excepto, a lo sumo, en un conjunto de medida de Lebesgue nula en el que algunas $\hat{F}_\pi^{(\alpha)}(.|x)$ pueden no estar definidas).

Estas observaciones sugieren la idea de modificar la definición 2.3 manteniendo idéntico texto y añadiendo al final la expresión "excepto, a lo sumo, en un conjunto de medida de Lebesgue nula". Con ésto habríamos conseguido, en el caso del ejemplo, definir la estabilidad como un concepto intrínseco a la familia $\mathcal{U} = \{F_\alpha(.|\theta),\ \theta \in \Omega\}_{\alpha\in I}$, no dependiendo de la familia $\mathcal{V}$ de verosimilitudes elegida.

Esta nueva definición "ampliada" es aplicable a otras muchas situaciones recogiendo sustancialmente la idea de estabilidad, pero cuando $\mathcal{V}$ contiene funciones de probabilidad, (es decir, densidades respecto a una medida "de contar") o cuando la estructura de π es más complicada, ya no sería adecuada y nos veríamos obligados a entrar en una casuística complicada y artificial que oscurecería el sentido de la definición.

En definitiva, parece conveniente mantener la definición 2.3 (dependiendo de la familia concreta $\mathcal{V}$ que se ha elegido) por las siguientes razones:

(a) El enunciado 2.3 recoge lo esencial de la idea en la forma más sencilla posible y es aplicable a situaciones sumamente generales.

(b) En la inmensa mayoría de los casos prácticos la familia V viene dada de una forma natural y nos interesa estudiar el problema para esa familia concreta, lo cual está de acuerdo, por otra parte, con la metodología bayesiana usual en la que la verosimilitud viene fijada.

2.4.2. Significado estadístico de la definición

La estabilidad que se ha definido 2.3 puede concebirse como una robustez fundamental o primaria, propia del modelo, ya que, de acuerdo con la filosofía implícita en los métodos bayesianos, la distribución a posteriori, en su conjunto, recoge toda la información disponible relativa al parámetro de interés y, por tanto, el primer requerimiento debe ser la estabilidad de dicha distribución. Sin embargo, lo usual en un problema de inferencia es condensar o resumir la información proporcionada por la distribución a posteriori en un sólo valor extraído de ésta, que es la estimación puntual, o bien determinar regiones de confianza o realizar contrastes de hipótesis. Por consiguiente, resulta natural dedicar una especial atención a la robustez de los estimadores, las regiones de confianza y los tests de hipótesis (ver III.3 y III.4).

A continuación se demuestra un teorema sencillo aplicable a un caso particular, pero suficiente amplio, y que es fundamental para ilustrar el significado de la definición ya que <u>permite interpretar la estabilidad en términos de continuidad</u> con respecto a la distancia de Prohorov.

<u>TEOREMA II.1</u>. Sea el modelo definido por la distribución a priori π sobre $\Omega \subset \mathbb{R}^k$ y la familia de verosimilitudes $V = \{f_\alpha(.|\theta)\ \theta \in \Omega\}_{\alpha \in I}$ Supongamos que

(i) La función de distribución correspondiente a $f_\alpha(.|\theta)$ es de la forma $F_\alpha(.|\theta) = F_\alpha(u_\theta(.))$ siendo $C = \{F_\alpha\}_{\alpha \in I}$ un conjunto de funciones de distribución absolutamente continuas en $\mathbb{R}^n$ y $\forall \theta \in \Omega$ $u_\theta : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ función biyectiva y bicontinua tal que $F_\alpha(u_\theta(.))$ es función de distribución $\forall(\alpha,\theta) \in I \times \Omega$.

(ii) Existe un conjunto $X \subset \mathbb{R}^n$ tal que, $\forall \alpha \in I$, $X_\alpha = X$ siendo $X_\alpha = \{x \in \mathbb{R}^n \mid 0 < \int_\Omega f_\alpha(x|\theta)d\pi(\theta) < \infty\}$.

Entonces el modelo (π, V) <u>es estable en</u> f_{α_o} <u>si y sólo si todos los operadores</u> $\{B_x\}_{x \in X}$ <u>son continuos en</u> F_{α_o}, siendo:

$$\forall x \in X \qquad C \xrightarrow{B_x} P$$

$$F_\alpha \longrightarrow B_x(F_\alpha) = F_\pi^{(\alpha)}(.|x)$$

donde P denota el conjunto de todas las distribuciones de probabilidad sobre Ω, al que se considera dotado (al igual que C) de la topología débil (inducida por la métrica de Prohorov; ver I.2).

<u>Demostración</u>:

En primer lugar obsérvese que

$$F_{\alpha_r} \xrightarrow{d} F_{\alpha_o} \text{ cuando } r \to \infty \iff \forall y \in \mathbb{R}^n \lim_{r\to\infty} F_{\alpha_r}(y) = F_{\alpha_o}(y) \iff$$

$$\iff \forall x \in \mathbb{R}^n,\ \forall \theta \in \Omega \lim_{r\to\infty} F_{\alpha_r}(u_\theta(x)) = F_{\alpha_o}(u_\theta(x))$$

La primera equivalencia se verifica por ser F_{α_o} continua. La implicación hacia la derecha en la segunda equivalencia se obtiene trivialmente haciendo $y = u_\theta(x)$. La aimplicación hacia la izquierda se obtiene haciendo $x = u_\theta^{-1}(y)$ para un valor cualquiera, prefijado, de θ.

Así se ha probado: $F_{\alpha_r} \xrightarrow{d} F_{\alpha_o} \Leftrightarrow F_{\alpha_r}(.|\theta) \xrightarrow{d} F_{\alpha_o}(.|\theta)$ $\forall\theta \in \Omega$ (1).

Por otra parte, la caracterización de la continuidad por sucesiones nos permite afirmar que:

$\forall x \in \chi$, B_x es continuo en $F_{\alpha_o} \Leftrightarrow$ Para toda sucesión $\{F_{\alpha_r}\}_{r\in N} \subset C$ tal que $F_{\alpha_r} \xrightarrow{d} F_{\alpha_o}$ se verifica que $B_x(F_{\alpha_r}) = F_\pi^{(\alpha_r)}(.\ x) \xrightarrow{d} B_x(F_{\alpha_o}) = F_\pi^{(\alpha_o)}(.\ x)$ cuando $r \to \infty$, y este último enunciado es a su vez equivalente, por (1) y por la definición 2.3, a la estabilidad de (π, V) en F_{α_o} c.q.d.

La interpretación intuitiva de este teorema es sencilla: El modelo planteado corresponde a una situación en que todas las posibles distribuciones básicas $F_\alpha(.|\theta)$, para θ fijo se diferencian sólo en la "forma" dada por F_α, (por ejemplo $f_\alpha(.|\theta) \equiv N(\theta,\alpha)$, $\alpha > 0$). En este caso la estabilidad indicaría, traducida a términos de continuidad, que, variaciones suficientemente pequeñas en la "forma" de la verosimilitud subyacente provocarán variaciones pequeñas en la distribución a posteriori (con respecto a la distancia de Prohorov).

La hipótesis (ii) puede eliminarse mediante un artificio similar al utilizado más adelante en el teorema II.2 que consiste en considerar la familia de operadores $\{B_x\}_{x \in \chi_{\alpha_o}}$, definiendo $B_x(F_\alpha) = F^{(\alpha_o)}(.|x)$ cuando $x \notin \chi_\alpha$. No obstante, se ha preferido mantener el enunciado en su forma actual para hacer resaltar más claramente su significado.

2.5. Interpretación matemática general

Según indicábamos al principio de esta sección, el concepto matemático de continuidad es la idea esencial en que se inspira la definición propuesta de estabilidad. Se trata, por tanto, de establecer formalmente la relación entre estos dos conceptos parafraseando la definición de estabilidad en términos de continuidad mediante una equivalencia lógica. En este apartado obtenemos dicha equivalencia para modelos (π, V) cuya familia U asociada verifique la siguiente condición de regularidad:

DEFINICION II.1

Se dice que la familia $U = \{F_\alpha(.|\theta),\ \theta \in \Omega\}_{\alpha \in I}$ está uniformemente parametrizada si y sólo si para toda sucesión $\{F_{\alpha_r}(.|\theta)\}_{r \in \mathbb{N}} \subset U$ tal que $d_p(F_{\alpha_r}(.|\theta),\ F_\alpha(.|\theta)) \longrightarrow 0$ cuando $r \to \infty$ $\forall \theta \in \Omega$, se verifica que $\sup_{\theta \in \Omega} d_p(F_{\alpha_r}(.|\theta),\ F_\alpha(.|\theta)) \longrightarrow 0$ cuando $r \to \infty$ (siendo d_p la distancia de Prohorov).

Intuitivamente, esta definición tiene una interpretación similar a la de la convergencia uniforme de una sucesión de funciones y se introduce para evitar anomalías en el comportamiento de las sucesiones que verifican $F_{\alpha_r}(.|\theta) \xrightarrow{d} F_\alpha(.|\theta)$ $\forall \theta \in \Omega$ imponiendo que todas las sucesiones $\{F_{\alpha_r}(.|\theta)\}_{r\in N}$, obtenidas fijando θ, tengan una "velocidad mínima común" de convergencia.

El siguiente teorema establece la equivalencia anunciada:

TEOREMA II.2. Sea un modelo (π, V) tal que la correspondiente familia U está uniformemente parametrizada. Definamos:

$$\forall f_{\alpha_1}, f_{\alpha_2} \in V, \quad d_p^*(f_{\alpha_1}, f_{\alpha_2}) = \sup_{\theta\in\Omega} d_p(F_{\alpha_1}(.|\theta), F_{\alpha_2}(.|\theta))$$

Entonces:

(i) d_p^* es una métrica en V

(ii) El modelo (π, V) es estable en f_{α_o} si y sólo si $\forall x \in \chi_{\alpha_o}$ el operador $B_x^{(\alpha_o)} : V \longrightarrow P$, definido por

$$B_x^{(\alpha_o)}(f_\alpha) = \begin{cases} F_\pi^{(\alpha)}(.|x) & \text{si } x \in \chi_\alpha \\ F^{(\alpha_o)}(.|x) & \text{si } x \notin \chi_\alpha \end{cases}$$

es continuo en f_{α_o}, siendo P el conjunto de distribuciones de probabilidad sobre Ω (con la métrica de Prohorov) y $\chi_\alpha = \{x \in R^n \mid 0 < \int_\Omega f_\alpha(x|\theta)d\pi(\theta) < \infty\}$.

Demostración:

Obsérvese en primer lugar que d_p^* está bien definida pues la distancia de Prohorov está acotada por 1.

Teniendo en cuenta que d_p es una métrica resulta trivial comprobar que $d_p^*(f_{\alpha_1}, f_{\alpha_2}) = 0 \iff f_{\alpha_1} = f_{\alpha_2}$, $d_p^*(f_{\alpha_1}, f_{\alpha_2}) = d_p^*(f_{\alpha_2}, f_{\alpha_1})$, $d_p^*(f_{\alpha_1}, f_{\alpha_2}) \leq d_p^*(f_{\alpha_1}, f_{\alpha_3}) + d_p^*(f_{\alpha_3}, f_{\alpha_2})$

$\forall f_{\alpha_1}, f_{\alpha_2}, f_{\alpha_3} \in V$, lo cual prueba (i).

La parte (ii) se prueba también de manera sencilla utilizando la caracterización de la continuidad por sucesiones (que es válida en espacios métricos):

$B_x^{(\alpha_o)}$ es continuo en f_{α_o} $\Longleftrightarrow$ $\forall \{f_{\alpha_r}\}_{r \in \mathbb{N}} \subset V \mid d_p^*(f_{\alpha_r}, f_{\alpha_o}) \xrightarrow[r \to \infty]{} 0$, se verifica que $d_p(B_x^{(\alpha_o)}(f_{\alpha_r}), B_x^{(\alpha_o)}(f_{\alpha_o})) \xrightarrow[r \to \infty]{} 0$.

Ahora bien, por la parametrización uniforme, $d_p^*(f_{\alpha_r}, f_{\alpha_o}) \xrightarrow[r \to \infty]{} 0$ $\Longleftrightarrow$ $d_p(F_{\alpha_r}(.|\theta).\ F_{\alpha_o}(.|\theta)) \xrightarrow[r \to \infty]{} 0$ $\forall \theta \in \Omega$ y esto equivale a $F_{\alpha_r}(.|\theta) \xrightarrow[r \to \infty]{d} F_{\alpha_o}(.|\theta)$ $\forall \theta \in \Omega$.

Por otra parte $d_p(B_x^{(\alpha_o)}(f_{\alpha_r}), B_x^{(\alpha_o)}(f_{\alpha_o})) \xrightarrow[r \to \infty]{} 0$ $\forall x \in X_{\alpha_o}$ $\Longleftrightarrow$ $F_\pi^{(\alpha_r)}(.|x) \xrightarrow[r \to \infty]{d} F_\pi^{(\alpha_o)}(.|x)$ $\forall x \in \bigcap_{i=o}^{\infty} X_{\alpha_i}$, ya que, para $x \notin \bigcap_{i=o}^{\infty} X_{\alpha_i}$, las $F_\pi^{(\alpha_i)}(.|x)$ que no estén definidas no presentan problema pues se sustituyen por $F_\pi^{(\alpha_o)}(.|x)$.

En definitiva, según la definición de estabilidad, hemos obtenido la equivalencia propuesta.

<u>Discusión y comentarios</u>:

(a) En esencia, el teorema afirma que la estabilidad del modelo en f_{α_o} es equivalente a la continuidad en f_{α_o} de los "operadores Bayes" $\{B_x \mid x \in X_{\alpha_o}\}$ con respecto a la topología inducida de d_p^* (que debe considerarse como la topología "natural" de V).

(b) Comparando este resultado con el teorema II.1 del apartado anterior, vemos que allí se nos ofrecía otra traducción de la estabilidad en términos de continuidad de una forma más directa aunque, en contrapartida, su ámbito de aplicación era más restringido. Así, en las hipótesis del teorema II.1

$F_\alpha(\cdot|\theta) = F_\alpha(u_\theta(\cdot))$, luego

$$d_p(F_{\alpha_1}(\cdot|\theta),\ F_{\alpha_2}(\cdot|\theta)) = \inf\{\varepsilon > 0 \mid F_{\alpha_1}(A|\theta) \leq F_{\alpha_2}(A^\varepsilon|\theta) + \varepsilon$$

$$\forall A \in \mathcal{B}\} = \inf\{\varepsilon > 0 \mid F_{\alpha_1}(u_\theta(A)) \leq F_{\alpha_2}(u_\theta(A^\varepsilon)) + \varepsilon \quad \forall A \in \mathcal{B}\} =$$

$$= \inf\{\varepsilon > 0 \mid F_{\alpha_1}(A) \leq F_{\alpha_2}(A^\varepsilon) + \varepsilon \quad \forall A \in \mathcal{B}\} = d_p(F_{\alpha_1}, F_{\alpha_2})$$

ya que u_θ es biyectiva y bicontinua.

Por tanto, en este caso d_p^* se reduce a la distancia de Prohorov corriente entre las distribuciones F_{α_1} y F_{α_2} que dan la "forma" y la parametrización uniforme se verifica de inmediato.

2.6. Ejemplo de modelo inestable

Vamos a plantear a continuación un ejemplo no trivial de modelo inestable que nos permitirá extraer algunas conclusiones interesantes.

Sea $V = \{f_\alpha(\cdot|\theta),\ \theta \in \Omega = [1,2]\}_{\alpha \in \mathbb{R}}$ donde $\forall(\alpha,\theta) \in \mathbb{R}\times[1,2]$, $f_\alpha(\cdot|\theta)$ es una función de densidad univariante definida por

$$f_\alpha(x|\theta) = (1-\varepsilon)\,\frac{1}{2\theta}\, I_{(-\theta,\theta)}(x) + \varepsilon\, g_\alpha(x)$$

siendo $I_{(-\theta,\theta)}$ la función indicatriz del intervalo $(-\theta,\theta)$, $\varepsilon \in (0,1)$ una constante conocida y $\{g_\alpha\}_{\alpha \in \mathbb{R}}$ una familia de funciones de densidad en R, correspondientes a distribuciones distintas, con la única condición de que contenga a las densidades $g_r(x) = [1 + \cos 2\pi r x] \cdot I_{(0,1)}(x)$, para $r=1,2,\dots$ y $g_o(x) = I_{(0,1)}(x)$.

Tomamos como distribución a priori la uniforme en $\Omega = [1,2]$.

Queremos estudiar la estabilidad del modelo en $f_o(.|\theta)$. Hablando en términos intuitivos, la situación puede resumirse así: La distribución básica es una uniforme en $(-\theta,\theta)$ "perturbada" en un porcentaje $100\varepsilon\%$ por una distribución que no depende de θ y de la cual conjeturamos que puede ser la uniforme en (0,1) o una muy próxima a ella. La pregunta que nos planteamos es: ¿Tendrán mucha influen--ia en la distribución a posteriori los pequeños errores en esta distribución "perturbadora"? La respuesta es afirmativa porque el sistema es inestable en $f_o(.|\theta)$. En efecto, es inmediato comprobar que

$$F_r(.|\theta) \xrightarrow[r\to\infty]{d} F_o(.|\theta) \qquad \forall\theta \in [1,2]$$

y además $\forall x \in (0,1)$ se tiene:

$$F_\pi^{(r)}(\theta|x) = \frac{1/2(1-\varepsilon)\ln\theta + \varepsilon(\theta-1)(\cos 2\pi r x+1)}{1/2(1-\varepsilon)\ln 2 + \varepsilon(\cos 2\pi r x+1)}$$

$$F_\pi^{(o)}(\theta|x) = \frac{1/2(1-\varepsilon)\ln\theta + \varepsilon(\theta-1)}{1/2(1-\varepsilon)\ln 2 + \varepsilon} \qquad \forall\theta \in [1,2]$$

y es claro que no se verifica $\forall x \in (0,1)$,

$F_\pi^{(r)}(\theta|x) \xrightarrow[r\to\infty]{} F_\pi^{(o)}(\theta|x) \quad \forall\theta \in (1,2)$. Por ejemplo, para $x=1/2$ y r impar se tiene: $F_\pi^{(r)}(\theta|1/2) = \dfrac{1/2(1-\varepsilon)\ln\theta}{1/2(1-\varepsilon)\ln 2}$ que, obviamente no es igual a $F_\pi^{(o)}(\theta|1/2) \quad \forall\theta \in (1,2)$.

Este contraejemplo sugiere algunas observaciones:

(a) La inestabilidad de este modelo se debe, en esencia, al mal comportamiento de las densidades frente a la convergencia débil. Este problema ya se ha mencionado en el capítulo I, y se analizará con más detalle en II.3.

(b) Este es un caso particular de "modelo de contaminación" en el cual la distribución "contaminante" no depende de θ, lo cual no es frecuente, (comparar, por ejemplo con Huber (1977), pág. 30). De todas formas, este modelo muestra la conveniencia de desarrollar la teoría en el ámbito más general posible, sin limitarnos a las $F_\alpha(.|\theta)$ del tipo $F_\alpha(.|\theta) = F_\alpha(u_\theta(.))$ (ver Teorema II.1).

3. CONDICIONES SUFICIENTES DE ESTABILIDAD

3.1. Métodos de Análisis

Cuando intentamos establecer condiciones de fácil comprobación que nos permitan asegurar la estabilidad de un modelo, nos enfrentamos en definitiva al problema de garantizar la convergencia débil de las distribuciones a posteriori. Un primer análisis de este problema nos conduce a la siguiente observación: Las distribuciones $F_\pi^{(\alpha)}(.|x)$ vienen definidas a través de su π-densidad, dada por la fórmula de Bayes, que, a su vez, depende de $f_\alpha(.|\theta)$. Por lo tanto, la manera más cómoda de establecer la convergencia débil de las $F_\pi^{(\alpha)}(.|x)$ sería utilizar el teorema de Scheffé, demostrando previamente $\pi_{\alpha_r}(\theta|x) \xrightarrow[r \to \infty]{} \pi_\alpha(\theta|x)$. Esta condición sería fácil de probar a partir de $F_{\alpha_r}(.|\theta) \xrightarrow{d} F_\alpha(.|\theta)$ si se verificase la implicación

"$F_{\alpha_r}(.|\theta) \xrightarrow{d} F_\alpha(.|\theta) \Longrightarrow f_{\alpha_r}(x|\theta) \longrightarrow f_\alpha(x|\theta) \quad \forall x$"; en este caso necesitaríamos simplemente añadir alguna hipótesis adecuada que nos asegurase $\int_\Omega f_{\alpha_r}(x|\theta)d\pi(\theta) \longrightarrow \int_\Omega f_\alpha(x|\theta)d\pi(\theta)$ para tener $\pi_{\alpha_r}(\theta|x) \longrightarrow \pi_\alpha(\theta|x) \quad \forall\theta$. Sin embargo, dicha implicación no es cierta en general y está es una de las mayores dificultades que aparece en el desarrollo posterior.

En resumen, cualquier condición adicional que nos permita (a partir de la convergencia débil) concluir la <u>convergencia puntual de las densidades</u> será útil para obtener condiciones suficientes de estabilidad. De hecho, los resultados principales de esta sección son los teoremas II.4, II.5 y II.6, que se basan en esta idea.

En principio, también podría intentarse analizar el problema directamente mediante la definición de convergencia débil (o por alguna caracterización equivalente). Como hemos visto en el capítulo I, esta metodología resulta útil para analizar la robustez del "operador de Bayes" respecto a la distribución a priori, pero, en nuestro caso, sólo es eficaz en situaciones muy particulares como la que se plantea en el apartado siguiente. Esto se debe, en esencia, a la mayor complejidad en la estructura de los operadores Bayes cuando se consideran como funcionales que dependen de la verosimilitud.

3.2. <u>Estabilidad en algunos modelos con parámetros de posición</u>

A continuación establecemos un resultado que es aplicable en situaciones bastante restringidas, aunque, por otra parte, tiene el interés de mostrar un caso en el que puede probarse la estabilidad utilizando directamente la definición de convergencia débil, sin necesidad de pasar por la convergencia puntual de las verosimilitudes.

TEOREMA II.3. Sea π una distribución de probabilidad μ_L-contínua sobre $\Omega = R^n$ que admite una función de densidad, p, continua y acotada en R^n.

Sea $V = \{f_\alpha(.|\theta),\ \theta \in \Omega\}_{\alpha \in I}$ donde $f_\alpha(x|\theta) = f_\alpha(x-\theta)$ y, para todo $\alpha \in I$, $f_\alpha : R^n \longrightarrow R^+$ es una función de densidad (con respecto a μ_L). Entonces, el modelo (π, V) es ESTABLE en V. (*)

Demostración:

Sea una sucesión de verosimilitudes tal que

$$F_{\alpha_r}(.|\theta) \xrightarrow{d} F_\alpha(.|\theta) \qquad \forall\theta \in \Omega$$

si queremos probar $F_\pi^{(\alpha_r)}(.|x) \xrightarrow{d} F_\pi^{(\alpha)}(.|x)$, bastará demostrar que, para toda función $u : R^n \longrightarrow R$ continua y acotada, se verifica:

$$\int_{R^n} u(\theta)dF_\pi^{(\alpha_r)}(\theta|x) \xrightarrow[r \to \infty]{} \int_{R^n} u(\theta)dF_\pi^{(\alpha)}(\theta|x) \qquad (1)$$

En efecto, denotando $K_r(x) = \int_{R^n} f_{\alpha_r}(x-\theta)p(\theta)d\theta$ y haciendo el cambio de variable $x-\theta = w$, tenemos

$$\int_{R^n} u(\theta)d\, F_\pi^{(\alpha_r)}(\theta|x) = \frac{1}{K_r(x)} \int_{R^n} f_{\alpha_r}(x-\theta)p(\theta)u(\theta)d\theta =$$

$$= \frac{1}{K_r(x)} \int_{R^n} f_{\alpha_r}(w)p(x-w)u(x-w)dw, \quad \text{pero}$$

$$K_r(x) = \int_{R^n} f_{\alpha_r}(x-\theta)p(\theta)d\theta = \int_{R^n} f_{\alpha_r}(w)p(x-w)dw \xrightarrow[r \to \infty]{}$$

$$\int_{R^n} f_\alpha(w)p(x-w)dw = \int_{R^n} f_\alpha(x-\theta)p(\theta)dw$$

(*) La expresión "el modelo es estable en V" significa, naturalmente, que (π, V) es estable en f_α, $\forall f_\alpha \in V$.

ya que, para x fijo, $p(x-w)$ es función continua y acotada de w.

Análogamente

$$\int_{R^n} f_{\alpha_r}(x-\theta)p(\theta)u(\theta)d\theta = \int_{R^n} f_{\alpha_r}(w)p(x-w)u(x-w)dw \xrightarrow[r \to \infty]{}$$

$$\longrightarrow \int_{R^n} f_{\alpha}(w)p(x-w)u(x-w)dw = \int_{R^n} f_{\alpha}(x-\theta)p(\theta)u(\theta)d\theta$$

y así queda probada la relación (1) y con ella la estabilidad.

<u>Discusión y comentarios</u>:

(a) La restricción más fuerte que se impone en este teorema es suponer que el espacio muestral y el paramétrico tienen ambos la misma dimensión. Esto significa, en la práctica, que el problema de inferencia se plantea con "muestras de tamaño 1": Por ejemplo, consideremos distribucions n-dimensionales del tipo $f_{\alpha}(x-\theta)$, donde f_{α} varía en un entorno de la normal $N(0,I)$. Queremos estimar el "centro de simetría" $\theta=(\theta_1,\ldots,\theta_n)$ y la información muestral viene dada por una sola observación $x = (x_1,\ldots,x_n)$.

(b) En este teorema hemos necesitado imponer ciertas condiciones sobre la distribución a priori; sin embargo, no se impone ninguna restricción sobre las distribuciones que dan "la forma" de las verosimilitudes (salvo el hecho de ser μ_L-continuas). Esta situación contrasta con el planteamiento de los teoremas II.4 y II.5 y II.6 que veremos más adelante.

(c) Puede notarse, por último que el resultado es generalizable fácilmente a otros casos en que $F_{\alpha}(x|\theta) = F_{\alpha}(u_{\theta}(x))$ aunque habría que añadir algunas hipótesis sobre las u_{θ} que complicarían el enunciado sin aportar ideas nuevas.

$$\int_\Omega f_{\alpha_r}(x|\theta)d\pi(\theta) \xrightarrow[r\to\infty]{} \int_\Omega f_\alpha(x|\theta)d\pi(\theta)$$

En definitiva, hemos probado "$F_{\alpha_r}(.|\theta) \xrightarrow{d} F_\alpha(.|\theta)$ $\forall\theta \in \Omega$" $\Longrightarrow$ $\Rightarrow$ "$\pi_{\alpha_r}(\theta|x) \longrightarrow \pi_\alpha(\theta|x)$ $\forall\theta \in \Omega$ y $\forall x \in R^n$ que verifique las condiciones impuestas en la definición 2.3" y de aquí, aplicando nuevamente el teorema de Scheffé, concluimos $F_\pi^{(\alpha_r)}(.|x) \xrightarrow{d} F_\pi^{(\alpha)}(.|x)$.

El siguiente corolario prueba que, en el caso particular importante de que las $f_\alpha(x|\theta)$ sean verosimilitudes correspondientes a observaciones i.i.d., (es decir, $f_\alpha(x|\theta) = \prod_{i=1}^n p_\alpha(x_i|\theta)$, será suficiente con imponer las hipótesis del teorema para $n=1$, pues, si se verifican para este valor, se verifican también $\forall n \in N$.

<u>COROLARIO II.4.1</u>.

Sea un modelo (π,V_n) donde las verosimilitudes de V_n son todas de la forma $f_\alpha(x|\theta) = \prod_{i=1}^n p_\alpha(x_i|\theta)$ $\forall x = (x_1,\ldots,x_n) \in R^n$.

Supongamos que el conjunto de verosimilitudes "univariantes" $V_1 = \{p_\alpha(.|\theta),\ \theta \in \Omega\}_{\alpha\in I}$ cumple las tres hipótesis del teorema II.4. Entonces el modelo (π,V_n) es estable en V_n.

<u>Demostración</u>:

(π,V_n) cumple obviamente la hipótesis (i) pues el producto de funciones continuas es continua.

En cuanto a la hipótesis (ii) obsérvese que, si la sucesión de distribuciones definidas por $\prod_{i=1}^n p_{\alpha_r}(x_i|\theta)$ convergiese débilmente cuando $\alpha_r \to c \in \hat{I}\setminus I$ hacia la distribución definida por $\prod_{i=1}^n p_\alpha(x_i|\theta)$ se verificaría (ver lema III.1) $p_{\alpha_r}(.|\theta) \xrightarrow{d} p_\alpha(.|\theta)$.

Así pues, también basta con imponer (ii) para el caso n=1.

Finalmente, con respecto a (iii), si existe $g_{x_i} : \Omega \longrightarrow R$ π-integrable tal que $p_\alpha(x_i|\theta) \leq g_{x_i}(\theta)$, entonces, para todo $x = (x_1,\ldots,x_n)$ existe $g_x = \prod_{i=1}^{n} g_{x_i} : \Omega \longrightarrow R$, π-integrable, tal que $f_\alpha(x|\theta) = \prod_{i=1}^{n} p_\alpha(x_i|\theta) \leq g_x(\theta)$ $\forall\theta \in \Omega$. Aplicando el teorema II.4 concluimos la estabilidad de (π, V_n).

Discusión y comentarios

(a) La estabilidad del modelo se asegura únicamente dentro de la familia paramétrica V. No hay ninguna garantía de que dicha estabilidad se conserve cuando $f_\alpha(.|\theta)$ se vea sometida a variaciones "infinitesimales" que la lleven fuera de V. Este hecho constituye, desde luego, una limitación a la aplicabilidad práctica del teorema que, aún así, será útil en los casos en que la información disponible nos permita elegir una familia V adecuada para "confinar" la verosimilitud básica. Este es, en esencia, el planteamiento del ya mencionado modelo de Box-Tiao, que se estudiará con más detalle en la sección 4.

(b) La hipótesis (ii) es una especie de "condición de frontera": exige que no podamos obtener una verosimilitud "repetida" (de las que ya están en V) como límite de una sucesión $\{f_{\alpha_r}\}$ haciendo tender α_r hacia un punto de la frontera "ampliada" de I.

Esta hipótesis puede suprimirse exigiendo, a cambio, que $I \subset R^m$ sea compacto.

(c) La hipótesis (iii) se impone obviamente para asegurar la convergencia de las integrales que aparecen en los denominadores de las fórmulas de Bayes aplicando el teorema de la convergencia dominada. Se da en la forma más complicada y directa porque también es la más general.

Cuando Ω es compacto esta hipótesis se cumple automáticamente siempre que, para todo x, $f_\alpha(x|\theta)$ sea continua en $I \times \Omega$ (como función de (α,θ)) pues, en este caso, cuando $\alpha_r \to \alpha$, $\exists K_x$ tal que $f_{\alpha_r}(x|\theta) \leq K_x$, $\forall r$, y las constantes son π-integrables.

(d) Bajo las hipótesis de este teorema, se puede definir fácilmente una topología natural en $\mathcal{U}$ a partir de la distancia:

$$d(f_{\alpha_1}, f_{\alpha_2}) = \|\alpha_1 - \alpha_2\|$$

siendo $\|\ .\ \|$ una norma cualquiera en R^m.

3.4. Condiciones suficientes de estabilidad a través de la función característica

En este apartado establecemos unos resultados que nos permitirán concluir la estabilidad demostrando, como paso previo, la convergencia de las $f_{\alpha_r}(x|\theta)$ a través de una conocida propiedad de las funciones características.

Los resultados fundamentales vienen dados por los Teoremas II.5 y II.6.

TEOREMA II.5. Sean π una distribución a priori sobre $\Omega \subset R^k$, $\mathcal{U} = \{F_\alpha(.|\theta),\ \theta \in \Omega\}_{\alpha \in I}$ y $\phi_{\alpha,\theta}$ la función característica correspondiente a $F_\alpha(.|\theta)$.

Dado $\alpha_o \in I$, supongamos que, para toda sucesión $S = \{\alpha_r\}_{r \in N}$

tal que $F_{\alpha_r}(.|\theta) \xrightarrow[r \to \infty]{d} F_{\alpha_o}(.|\theta) \quad \forall\theta \in \Omega$, se verifica:

(i) Para todo $\theta \in \Omega$ existe una función integrable $h_\theta^s : \mathbb{R}^n \longrightarrow [0,+\infty)$ tal que $|\phi_{\alpha_r,\theta}(t)| \leq h_\theta^s(t)$, $\forall t \in \mathbb{R}^n$, $\forall r \in \mathbb{N} \cup \{o\}$.

(ii) Para todo $x \in \mathbb{R}^n$ existe una función π-integrable $g_x^s : \Omega \longrightarrow [0,+\infty)$ tal que $f_{\alpha_r}(x|\theta) \leq g_x^s(\theta)$ $\forall\theta \in \Omega$, $\forall r \in \mathbb{N}$, siendo $f_\alpha(.|\theta)$ la función de densidad asociada a $F_\alpha(.|\theta)$ dada por la fórmula de inversión de Fourier:

$$f_\alpha(x|\theta) = \left(\frac{1}{2\pi}\right)^n \int_{\mathbb{R}^n} \exp\ (-it'x)\ \phi_{\alpha,\theta}(t)dt.$$

Entonces el modelo definido por la distribución a priori π y por la familia $V = \{f_\alpha(.|\theta),\ \theta \in \Omega\}_{\alpha\in I}$ es estable en f_{α_o}.

Demostración:

El teorema de inversión de Fourier afirma que, cuando $\phi_{\alpha,\theta}$ es absolutamente integrable en $\mathbb{R}^n$, entonces la correspondiente distribución admite una función de densidad continua y acotada dada por

$$f_\alpha(x|\theta) = \left(\frac{1}{2\pi}\right)^n \int_{\mathbb{R}^n} \exp\ (-it'x)\ \phi_{\alpha,\theta}(t)dt$$

Obsérvese que, si F_{α_o} no está "aislada" (es decir, si existe alguna sucesión $\{F_{\alpha_r}\}$ tal que $F_{\alpha_r}(.|\theta) \xrightarrow{d} F_{\alpha_o}(.|\theta)$ $\forall\theta \in \Omega$, lo cual se supone siempre implícitamente), la hipótesis (i) implica que todas las $F_\alpha(.|\theta)$ tienen función característica absolutamente integrable y, por tanto, tiene sentido considerar la familia de verosimilitudes V.

Ahora, sea $\{F_{\alpha_r}\}_{r\in\mathbb{N}}$ tal que $F_{\alpha_r}(.|\theta) \xrightarrow{d} F_{\alpha_o}(.|\theta)$ $\forall\theta \in \Omega$.

Esto equivale a $\phi_{\alpha_r,\theta}(t) \xrightarrow[r\to\infty]{} \phi_{\alpha_o,\theta}(t)$ $\forall t \in R^n$, es decir $|\phi_{\alpha_r,\theta}(t) - \phi_{\alpha_o,\theta}(t)| \xrightarrow[r\to\infty]{} 0$.

Según la hipótesis (i) $\exists h_\theta^s$ tal que $|\phi_{\alpha_r,\theta}(t)| \leq h_\theta^s(t)$ y $|\phi_{\alpha_r,\theta}(t) - \phi_{\alpha_o,\theta}(t)| \leq 2h_\theta^s(t)$.

Así pues, según el teorema de la convergencia dominada $|\phi_{\alpha_r,\theta}(t) - \phi_{\alpha_o,\theta}(t)| \xrightarrow[r\to\infty]{} 0$ implica $\int_{R^n} |\phi_{\alpha_r,\theta}(t) - \phi_{\alpha_o,\theta}(t)|dt \to 0$. Por otra parte, aplicando el teorema de inversión de Fourier

$$\forall x \in R^n \; |f_{\alpha_r}(x|\theta) - f_{\alpha_o}(x|\theta)| \leq \left(\frac{1}{2\pi}\right)^n \int_{R^n} |\phi_{\alpha_r,\theta}(t) - \phi_{\alpha_o,\theta}(t)|dt \to 0 .$$

Luego

$$f_{\alpha_r}(x|\theta) \xrightarrow[r\to\infty]{} f_{\alpha_o}(x|\theta) \quad \forall x \in R^n, \quad \forall\theta \in \Omega \qquad (1)$$

(nótese además que, $\forall\theta$ fijo la convergencia de $f_{\alpha_r}(.|\theta)$ a $f_{\alpha_o}(.|\theta)$ es <u>uniforme</u> en R^n).

Ahora, a partir de ésto, la hipótesis (ii) implica, por el teorema de la convergencia dominada:

$$\int_\Omega f_{\alpha_r}(x|\theta)d\pi(\theta) \longrightarrow \int_\Omega f_{\alpha_o}(x|\theta)d\pi(\theta) \qquad (2).$$

De (1) y (2) deducimos $\pi_{\alpha_r}(\theta|x) \longrightarrow \pi_{\alpha_o}(\theta|x)$ y, en consecuencia, la estabilidad en f_{α_o}.

El siguiente teorema es una variante del resultado anterior que ofrece quizá un mayor interés práctico pues nos permite suprimir la hipótesis (ii) a cambio de limitarnos a un cierto tipo de familias $\mathcal{U}$ (ya estudiadas en el teorema II.1) en que las correspondientes distribuciones básicas sólo se diferencian por la "forma" (es decir, $F_\alpha(x|\theta) = F_\alpha(u_\theta(x))$).

TEOREMA II.6. Sea π una distribución a priori sobre $\Omega \subset R^k$ y sea $\mathcal{U} = \{F_\alpha(.|\theta),\ \theta \in \Omega\}_{\alpha \in I}$, donde $\forall(\alpha,\theta) \in I \times \Omega$, $F_\alpha(.|\theta)$ es una función de distribución en R^n de la forma $F_\alpha(.|\theta) = F_\alpha(u_\theta(.))$ siendo $F_\alpha : R^n \longrightarrow [0,1]$ una función de distribución cuya función característica denotaremos por ϕ_α.

Supongamos que

(i) Para toda sucesión $S = \{\alpha_r\}_{r \in N}$ tal que $F_{\alpha_r} \xrightarrow[r \to \infty]{d} F_{\alpha_o}$, existe una función $h^s : R^n \longrightarrow [0,\infty)$ integrable, tal que

$$|\phi_{\alpha_r}(t)| \leq h^s(t), \quad \forall t \in R, \quad \forall r \in N \cup \{o\}$$

(ii) Para todo $\theta \in \Omega$, $u_\theta : R^n \longrightarrow R^n$ es función biyectiva, bicontinua y diferenciable tal que, $\forall x \in R^n$ el valor absoluto del jacobiano $|J_x(\theta)| = |\det \frac{\partial u_\theta^i}{\partial x_j}(x)|$ es función continua y acotada de θ.

Entonces, el modelo definido por la distribución a priori π y por la familia de verosimilitudes $V = \{f_\alpha(.|\theta) = f_\alpha(u_\theta(.))|J_x(\theta)|,\ \theta \in \Omega\}_{\alpha \in I}$ (donde $f_\alpha(x) = (\frac{1}{2\pi})^n \int_{R^n} e^{-its}\phi_\alpha(t)dt$) es <u>estable</u> en f_{α_o}.

<u>Demostración</u>:

Con un razonamiento idéntico al utilizado en la demostración anterior tenemos:

$$F_{\alpha_r} \xrightarrow[r \to \infty]{d} F_{\alpha_o} \Longrightarrow |\phi_{\alpha_r}(t) - \phi_{\alpha_o}(t)| \xrightarrow[r \to \infty]{} 0 \Longrightarrow \int_{R^n} |\phi_{\alpha_r}(t) -$$

$$- \phi_{\alpha_o}(t)|dt \xrightarrow[r \to \infty]{} 0 \Longrightarrow f_{\alpha_r} \xrightarrow[r \to \infty]{} f_{\alpha_o}$$ uniformemente en R^n.

Ahora, si tenemos en cuenta que $f_\alpha(x|\theta) = f_\alpha(u_\theta(x))|J_x(x)|$ resulta, para todo x fijo:

$$\sup_{\theta \in \Omega} |f_{\alpha_r}(x|\theta) - f_{\alpha_o}(x|\theta)| \leq \sup_{\theta \in \Omega} |J_x(\theta)| \sup_{\theta \in \Omega} |f_{\alpha_r}(u_\theta(x)) -$$

$$- f_{\alpha_o}(u_\theta(x))| \xrightarrow[r \to \infty]{} 0,$$ ya que $J_x(\theta)$ está acotada, por hipótesis, y $f_{\alpha_r} \xrightarrow[r \to \infty]{} f_{\alpha_o}$ uniformemente en R^n.

Por lo tanto hemos probado que $\forall x \in R^n$ $f_{\alpha_r}(x|.) \xrightarrow[r \to \infty]{} f_{\alpha_o}(x|.)$ uniformemente en Ω y esto implica (por tener Ω π-medida finita):

$$\int_\Omega f_{\alpha_r}(x|\theta)d\pi(\theta) \xrightarrow[r \to \infty]{} \int_\Omega f_{\alpha_o}(x|\theta)d\pi(\theta).$$

En definitiva, se tiene: $\pi_{\alpha_r}(\theta|x) \xrightarrow[r \to \infty]{} \pi_{\alpha_o}(\theta|x)$, con lo que queda probada la estabilidad.

En relación con estos dos teoremas, se puede enunciar también un resultado cuyo significado es totalmente análogo al del corolario II.4.1.:

COROLARIO II.6.1.

Sea un modelo (π, V_n) donde las verosimilitudes de V_n son todas de la forma $f_\alpha(x|\theta) = \prod_{i=1}^{n} p_\alpha(x_i|\theta)$ $\forall x = (x_1,\ldots,x_n) \in R^n$.

Supongamos que el conjunto de verosimilitudes "univariantes"

$\mathcal{V}_1 = \{p_\alpha(.|\theta),\ \theta \in \Omega\}_{\alpha\in I}$ cumple las hipótesis del teorema II.5 (o bien las del th. II.6).

Entonces $(\pi, \mathcal{V}_n)$ es estable en $\mathcal{V}_n$.

Demostración:

Basta comprobar que, si las hipótesis se cumplen para n=1, entonces se cumplen también para cualquier tamaño muestral n. Las comprobaciones relativas a las hipótesis (i) y (ii) del teorema II.5 y (i) del teorema II.6 son triviales.

Respecto a la hipótesis (ii) del teorema II.6, observemos que, si $p_\alpha(x_i|\theta) = p_\alpha(v_\theta(x_i))$, entonces $f_\alpha(x|\theta) = f_\alpha(u_\theta(x))$, donde $f_\alpha(z_1,\dots,z_n) = \prod_{i=1}^{n} p_\alpha(z_i)$ y $u_\theta(x_1,\dots,x_n) = (v_\theta(x_1),\dots,v_\theta(x_n))$ que será función biyectiva y bicontinua si lo es v_θ y el valor absoluto de su jacobiano $|J_x(\theta)| = |\prod_{i=1}^{n} v_\theta(x_i)|$ será función continua y acotada de θ para todo $x \in R^n$ si lo es $v_\theta(x_i)$ $\forall x_i \in R$.

Discusión y comentarios:

(a) Los teoremas II.5 y II.6 son, en cierto modo, más potentes que el II.4 pues nos aseguran la estabilidad, bajo ciertas condiciones de regularidad, para verosimilitudes cualesquiera, sin necesidad de restringirnos a familias paramétricas, lo cual está más acorde con la idea intuitiva de estabilidad. En contrapartida, las hipótesis de los teoremas II.5 y II.6 serán, en algunos casos, de comprobación más complicada que las del teorema II.4.

(b) El teorema II.6 cubre una amplia gama de situaciones prácti cas y es más cómodo de manejar que el II.5. Sin embargo, este último tiene la ventaja de ser aplicable también al caso en que la distribución a priori π sea impropia (ver sección III.1).

(c) La verificación de la hipótesis (i) en un caso concreto (tanto en el teorema II.5 como en el II.6) se reduce a comprobar que una sucesión de funciones características $\{\phi_r\}_{r\in N}$ cuyas correspondientes funciones de distribución cumplen $F_r \xrightarrow[r\to\infty]{d} F$, está mayorada en módulo por una función integrable. Sería interesante dar una condición suficiente para que se cumpla esta hipótesis <u>en términos de las densidades</u>. Esto puede conseguirse (en el caso univariante) utilizando algunas propiedades conocidas de las transformadas de Fourier que nos permiten obtener el siguiente resultado:

<u>LEMA II.1</u>. Sean F_r, $(r \in N)$ y F funciones de distribución absolutamente continuas en R con funciones características ϕ_r, $r \in N$; y ϕ respectivamente. Supongamos que

(i) $F_r \xrightarrow[r\to\infty]{d} F$

(ii) Las F_r y F admiten densidades f_r y f con segundas derivadas tales que existe una función integrable $q : R \longrightarrow [0,\infty)$ cumpliendo $|f_r''(x)| \leq q(x)$ y $|f''(x)| \leq q(x)$ $\forall r \in N$, $\forall x \in R$.

Entonces, existe una función integrable $h : R \longrightarrow [0,\infty)$ y $r_0 \in N$ tales que $|\phi_r(t)| \leq h(t)$ y $|\phi(t)| \leq h(t)$ $\forall r \geq r_0$, $\forall t \in R$.

<u>Demostración</u>:

Como las densidades tienen, por la hipótesis (ii), segundas derivadas integrables, se cumple:

$$\phi_r(t) = \left(\frac{1}{it}\right)^2 \int_R e^{itx} f_r''(x)dx \quad y \quad \phi(t) = \left(\frac{1}{it}\right)^2 \int_R e^{itx} f''(x)dx$$

(Esta es una propiedad muy conocida y de fácil demostración: Ver, por ejemplo, Kolmogorov y Fomin (1975), p. 474. Feller (1971), p. 514).

Por tanto,

$$|\phi_r(t)| \leq \frac{1}{t^2} \int_R q(x)dx \qquad y \qquad |\phi(t)| \leq \frac{1}{t^2} \int_R q(x)dx.$$

Además sabemos que $F_r \xrightarrow[r \to \infty]{d} F$ implica $\phi_r \xrightarrow[r \to \infty]{d} \phi$ uniformemente en todo intervalo acotado; así dado $\varepsilon > 0$, $\exists r_o \in N$ cumpliendo

$$\sup_{t \in [-1,1]} \left||\phi_r(t)| - |\phi(t)|\right| \leq \sup_{t \in [-1,1]} |\phi_r(t) - \phi(t)| < \varepsilon$$

$\forall r \geq r_o$

Luego $\forall r \geq r_o$ $\quad |\phi(t)| - \varepsilon \leq |\phi_r(t)| \leq |\phi(t)| + \varepsilon \quad \forall t \in [-1,1]$.

Si definimos

$$h(t) = \begin{cases} |\phi(t)| + \varepsilon & \text{si} \quad t \in [-1,1] \\ M/t^2, \quad (\text{siendo } M = \int_R q(x)dx), & \text{si} \quad t \notin [-1,1] \end{cases}$$

Esta función es integrable en R y verifica las acotaciones requeridas.

(d) Es interesante notar que en el teorema II.6 la distribución a priori π no interviene en la formulación de las hipótesis. De hecho, todas las restricciones recaen sobre la familia U de distribuciones básicas y la estabilidad se afirma para el modelo (π, V) siendo V la familia de verosimilitudes "natural" (dada por la fórmula de inversión de Fourier) y π una distribución a priori arbitraria.

4. EJEMPLOS DE APLICACION

Intentaremos ahora aplicar los resultados de la sección anterior al análisis de algunos modelos concretos referentes al clásico problema de estimación de un parámetro de posición en una distribución simétrica.

4.1. El modelo de Box-Tiao

Sea π una distribución a priori sobre $\Omega \subset \mathbb{R}$ y sea

$$V_1 = \{f_\alpha(.|\theta),\quad \theta \in \Omega \subset \mathbb{R}\}_{\alpha \in [\alpha_o,\alpha_1]}$$

donde $\alpha_o,\alpha_1 \in (-1,1)$, $\alpha_o < \alpha_1$ y $\forall\alpha \in [\alpha_o,\alpha_1]$, $f_\alpha(x|\theta) =$
$= K_\alpha \exp\{-\frac{1}{2}|x-\theta|^{2/1+\alpha}\}$, $\forall x \in \mathbb{R}$, siendo $K_\alpha^{-1} =$
$= \Gamma[1 + \frac{1}{2}(1+\alpha)]2^{[1+\frac{1}{2}(1+\alpha)]}$.

El modelo (π,V_1) resultará adecuado para estudiar la estabilidad en las proximidades de la normal con respecto a variaciones en la kurtosis. Puede darse una versión más general incluyendo un parámetro de escala σ; en este caso podría tomarse (θ,σ), en principio, como parámetro de interés y, una vez calculada la distribución a posteriori $\pi(\theta,\alpha|x)$, considerar σ como un "parámetro perturbador" y manejar la distribución a posteriori marginal de θ. Estos "procedimientos de marginalización", junto con el posible uso de distribuciones a priori impropias, tienen ciertas implicaciones teóricas que se analizan en la sección III.2.

Comprobaremos ahora que (π,V_1) verifica las tres hipótesis del teorema II.4, lo cual nos llevará a concluir su estabilidad:

La continuidad de $f_\alpha(x|\theta)$, considerada como función de α, $\forall(x,\theta) \in \mathbb{R} \times \Omega$ se sigue trivialmente de la propia expresión analíti ca de $f_\alpha(x|\theta)$.

La comprobación de la hipótesis (ii) es también inmediata, por ser I compacto.

Por otra parte $\max_{x \in \mathbb{R}} f_\alpha(x|\theta) = K_\alpha$, $\forall\theta \in \Omega$; como K_α es fun ción de α acotada en $[\alpha_0,\alpha_1]$ (por ser continua), se tiene

$$\exists K \in \mathbb{R} \mid f_\alpha(x|\theta) \leq K, \quad \forall(\alpha,x,\theta) \in [\alpha_0,\alpha_1] \times \mathbb{R} \times \Omega$$

Esto prueba que se cumple también la hipótesis (iii), pues las funcio nes constantes son π-integrables en Ω.

Además, el corolario II.4.1. nos permite asegurar también la estabilidad del modelo n-variante (π,V_n) con verosimilitudes del ti po $\psi_\alpha(x|\theta) = \prod_{i=1}^{n} f_\alpha(x_i|\theta)$ (correspondientes a observaciones $x = (x_1,\ldots,x_n)$ de variables aleatorias i.i.d.) cualquiera que sea $n > 1$.

4.2. Modelos de "mixturas infinitas"

Una posibilidad alternativa al modelo de Box-Tiao, sugerida por Girón (1981), viene dada por las llamadas "mixturas infinitas". Las familias de verosimilitudes definidas por este procedimiento tienen, entre otras propiedades interesantes, una interpretación estadística directa, permiten cubrir una amplia gama de verosimilitudes, y son relativamente sencillas en lo que se refiere a su manejo matemático.

Consideremos, por ejemplo

$$\mathcal{V}_1 = \{f_\alpha(.|\theta),\quad \theta \in \Omega \subset \mathbb{R}\}_{\alpha\in[0,\infty)}$$

donde, $\forall\alpha > 0 \quad f_\alpha(x|\theta) = \int_0^\infty N(x;\theta,1/v)\gamma(v;1/\alpha,1/\alpha)dv$, siendo

$$N(x;\theta,1/v) = v^{1/2}(2\pi)^{-1/2} \exp\left[-\frac{v}{2}(x-\theta)^2\right],\quad x\in\mathbb{R},\quad \theta\in\Omega,$$

$$\gamma(v;1/\alpha,1/\alpha) = \frac{(1/\alpha)^{1/\alpha}}{\Gamma(1/\alpha)} v^{1/\alpha-1} \exp\left(-\frac{v}{\alpha}\right),\quad v > 0,$$

y, para $\alpha = 0$, $f_0(x|\theta) = N(x;\theta,1)$.

$\mathcal{V}_1$ puede interpretarse intuitivamente considerando que, para $\alpha > 0$, la verosimilitud $f_\alpha(.|\theta)$ corresponde a una normal de media θ y precisión aleatoria v que se elige "sorteando" según una gamma de parámetros $a = 1/\alpha$, $p = 1/\alpha$. La verosimilitud correspondiente a $\alpha = 0$ es $N(x;\theta,1)$. Mediante cálculo directo obtenemos

$$f_\alpha(x|\theta) \propto \int_0^\infty v^{1/2} \exp\left[-\frac{v}{2}(x-\theta)^2\right] v^{1/\alpha-1} \exp\left(-\frac{v}{\alpha}\right)dv =$$

$$= \int_0^\infty \exp\left[-v\left(\frac{1}{2}(x-\theta)^2 + \frac{1}{\alpha}\right)\right] v^{1/\alpha-1/2}\, dv$$

así, $f_\alpha(x|\theta) = \dfrac{C_\alpha}{\left[\dfrac{(x-\theta)^2}{2} + \dfrac{1}{\alpha}\right]^{1/\alpha+1/2}}$, donde $C_\alpha = \dfrac{\left(\frac{1}{\alpha}\right)^{1/\alpha}\Gamma\left(\frac{1}{\alpha}+\frac{1}{2}\right)}{(2\pi)^{1/2}\,\Gamma\left(\frac{1}{\alpha}\right)}$.

Por tanto, para $\alpha = 2/n$ $(n\in\mathbb{N})$, $f_\alpha(x|\theta)$ es la t de Student de n grados de libertad "centrada" en θ. En general, $f_\alpha(x|\theta)$ puede considerarse como una t de Student "generalizada" con "grados de libertad" $2/\alpha$.

En resumen, la familia $\mathcal{V}_1$ refleja la siguiente situación: La verosimilitud "teórica" es $N(x;\theta,1)$ (correspondiente a $\alpha = 0$) y estudiamos las desviaciones de ella en el sentido de la t de Student "generalizada".

Ahora, si consideramos definida una distribución a priori π, podemos utilizar de nuevo el teorema II.4 y el corolario II.4.1 y llegar a las mismas conclusiones de estabilidad que en el modelo anterior; para ello debemos comprobar que (π, V_1) verifica las tres hipótesis del teorema II.4:

(i) $f_\alpha(x|\theta)$ es, obviamente, función continua de α en $(0,\infty)$ $\forall(x,\theta) \in R \times \Omega$ prefijado. Para estudiar la continuidad en $\alpha = 0$ consideremos una sucesión $\{\alpha_r\} \subset [0,\infty)$ tal que $\lim_{r\to\infty} \alpha_r = 0$; tenemos

$$\forall\theta \in \Omega, \quad \forall x \neq \theta \quad f_{\alpha_r}(x|\theta) = \int_0^\infty N(x;\theta,1/v)\gamma(v;1/\alpha_r,1/\alpha_r)dv \longrightarrow$$

$$\longrightarrow N(x;\theta,1) = f_0(x|\theta)$$

ya que la distribución con densidad $\gamma(.;1/\alpha_r,1/\alpha_r)$ converge débilmente a la distribución causal en 1 cuando $\alpha_r \to 0$ y $N(x;\theta,1/v) = v^{1/2}(2\pi)^{-1/2}\exp(-\frac{v}{2}(x-\theta)^2)$ es (para $x \neq \theta$) función contínua y acotada de v (pues $\lim_{v\to\infty} N(x;\theta,1/v) = 0$).

Para $x = \theta$ se tiene que $f_{\alpha_r}(\theta|\theta) = \int_0^\infty v^{1/2}(2\pi)^{-1/2}$ $\gamma(v;1/\alpha_r,1/\alpha_r)dv$ es el "momento no central" de orden 1/2 de la distribución $\gamma(.;1/\alpha_r,1/\alpha_r)$, cuya media (momento de orden 1) es igual a 1 cualquiera que sea α_r; así pues, aplicando el lema III.3 de convergencia de momentos, concluimos

$$\int_0^\infty v^{1/2}(2\pi)^{-1/2}\,\gamma(v;1/\alpha_r,1/\alpha_r)dv \longrightarrow 1 = f_0(\theta|\theta)$$

Por tanto, hemos probado que $f_\alpha(x|\theta)$ es función contínua de α en

$[0,\infty)$ $\forall(x,\theta) \in R \times \Omega$.

(ii) La segunda hipótesis del teorema se verifica también en este caso, pues el único punto límite a examinar es ∞ y se comprueba de inmediato que, $\forall\theta$, $\lim_{\alpha\to\infty} f_\alpha(x|\theta) = 0$ uniformemente en cualquier intervalo acotado que no contenga a θ. Esto excluye la posibilidad de que $\alpha_o \in [0,\infty)$ tal que $F_\alpha(.|\theta) \xrightarrow{d} F_{\alpha_o}(.|\theta)$ $\forall\theta \in \Omega$, cuando $\alpha \to \infty$.

(iii) Por último, observemos que, para toda sucesión $\{\alpha_r\} \subset (0,\infty)$ tal que $\alpha_r \to \alpha_o$, se tiene $\forall x \in R$ prefijado:

$$f_{\alpha_r}(x|\theta) \leq \int_0^\infty v^{1/2}\,(2\pi)^{-1/2}\,\gamma(v;\,1/\alpha_r,1/\alpha_r)dv$$ y, aplicando,

como antes, el lema de convergencia de momentos obtenemos que esta sucesión de integrales es acotada por ser convergente.

Por otra parte, $f_0(x|\theta) \leq (2\pi)^{-1/2}$. Así, para toda sucesión convergente $\{\alpha_r\} \subset [0,\infty)$ se tiene que $\exists K \in R$, $f_{\alpha_r}(x|\theta) \leq K$ $\forall\theta \in \Omega$ y, como $g_x(\theta) = K$ $\forall\theta \in \Omega$, es función π-integrable, se concluye que el modelo verifica también la hipótesis (iii).

Se pueden definir de manera análoga otros modelos de mixturas infinitas. Por ejemplo, tomando la varianza (en lugar de la precisión) como parámetro de mixtura o cambiando la distribución de mixtura. En algunos de los modelos resultantes será difícil calcular explícitamente las densidades, pero puede ser más sencillo manejar las funciones características (apoyándonos en la propiedad que tienen éstas de "respetar" el proceso de mixtura). En estas situaciones resultará más adecuado abordar el análisis del modelo a través del teorema II.6:

Consideremos, por ejemplo, el modelo definido por la distribución a priori π sobre $\Omega \subset \mathbb{R}$ y por la familia:

$$W_1 = \{f_\alpha(.|\theta),\quad \theta \in \Omega\}_{\alpha \in (0,\infty)\times(0,\infty)}$$

donde $f_\alpha(x|\theta) = \int_1^\infty N(x;\theta,v)d\ G_\alpha(v),\quad \forall x \in \mathbb{R},$ siendo

$$G_\alpha(v) = \varepsilon_\alpha \delta(v) + (1-\varepsilon_\alpha)\ G(v;a,p),\quad \alpha = (a,p),\quad (a > 0,\ p > 0)$$

$$\varepsilon_\alpha = \int_0^1 \gamma(v;a,p)dv,\quad \delta(v) = \begin{cases} 0, & \text{si}\quad v < 1 \\ 1, & \text{si}\quad v \geq 1 \end{cases}$$ y $G(v;a,p)$ la

función de distribución correspondiente a $\gamma(v;a,p)$.

Para dar una interpretación estadística a $f_\alpha(.|\theta)$, podemos considerar que la verosimilitud básica es $N(x;\theta,1)$ con probabilidad ε_α y, con probabilidad $1-\varepsilon_\alpha$, tomamos la verosimilitud entre las $N(x;\theta,v)$ eligiendo v en $[1,\infty)$ según una gamma truncada de parámetros $(a,p) = \alpha$.

La función característica correspondiente a $f_\alpha(.|\theta)$ es

$$\phi_{\alpha,\theta}(t) = e^{it\theta}\int_1^\infty e^{-1/2t^2v}\ d\ G_\alpha(v).$$

Utilizando el teorema II.6 deducimos de inmediato la estabilidad de (π,W_1) ya que

$$|\phi_{\alpha,\theta}(t)| \leq e^{-\frac{1}{2}t^2}\int_1^\infty d\ G_\alpha(v) = e^{-\frac{1}{2}t^2},\quad \forall(\alpha,\theta) \in I \times \Omega,$$ y

$h(t) = e^{-\frac{1}{2}t^2}$ es función integrable en $\mathbb{R}$.

Además, en virtud del corolario II.6.1 también son estables todos los modelos (π,V_n) con verosimilitudes de la forma $\psi_\alpha(x|\theta) = \prod_{i=1}^{n} f_\alpha(x_i|\theta)$.

CAPITULO III

CAPITULO III

APLICACIONES. ROBUSTEZ DE LOS METODOS DE INFERENCIA

1. SUMARIO

El objetivo de este capítulo es analizar las derivaciones de la teoría del capítulo II, en un triple aspecto: Aplicaciones directas de la misma (sección 2), estudio de la robustez de los métodos de inferencia (secciones 3 y 4) y generalizaciones (sección 5).

Así, en la sección 2, dedicada al modelo lineal bayesiano, se consideran las desviaciones de la hipótesis de normalidad que aparecen usualmente en esta teoría y se proponen dos tipos de modelos (de "contaminación" y de "convolución") para estudiarlas. Aplicando los resultados del capítulo II se demuestra la estabilidad de distintas variantes de estos modelos. Asimismo se examinan los problemas derivados del uso de distribuciones a priori impropias y de distribuciones a posteriori marginales.

En la sección 3 se propone una definición de robustez cualitativa de un estimador que se sigue de forma natural del concepto de estabilidad de un modelo enunciado en II.2.3.. La robustez de los estimadores se puede considerar, a grandes rasgos, como una propiedad "secundaria" en el sentido de que debe estudiarse después de analizar la estabilidad "genérica" o "primaria" del modelo. En los apartados 2, 3, 4 y 5 de esta sección se demuestran varios teoremas que proporcionan condiciones suficientes para asegurar la robustez cualitativa de los principales estimadores.

La sección 4 está dedicada al estudio de la robustez cualitativa de las regiones de confianza y tests de hipótesis bayesianos. Se proponen las definiciones que aparecen más adecuadas a nuestro enfoque y se discute su significado y aplicabilidad. Se concluye, en resumen, que (a diferencia de lo que ocurre con los estimadores), la robustez cualitativa de los tests es una propiedad intrínseca al modelo, en el sentido de que se cumple automáticamente cuando el modelo subyacente es estable.

El caso de las regiones de confianza es, por su propia naturaleza, más complicado; en el texto se analiza la definición, inspirada, como las restantes, en la noción de "continuidad" y se demuestra la robustez cualitativa de los intervalos de confianza bayesianos definidos mediante cuantiles.

Por último en la sección 5 se sugiere la posibilidad de desarrollar una "teoría generalizada" de la estabilidad en inferencia bayesiana que estudiase los efectos de pequeñas variaciones simultáneas en los tres elementos de un problema de inferencia bayesiana (distribución a priori, verosimilitud,función de pérdida) tomando como punto de partida nuestro enfoque de la robustez y los resultados que se exponen en el capítulo I. En concreto, como un primer paso en esta dirección, se demuestra un teorema que da una condición suficiente para asegurar la "estabilidad conjunta" de la distribución a posteriori frente a variaciones simultáneas de la distribución a priori y de la verosimilitud.

2. APLICACION A LOS MODELOS LINEALES

Se trata ahora de estudiar la estabilidad de los métodos bayesianos usuales en uno de los problemas de inferencia más frecuentes en la práctica: el modelo lineal.

Para ello haremos previamente un resumen de lo más significativo de esta teoría.

2.1. El modelo lineal bayesiano: planteamiento y resumen de la teoría

El planteamiento del problema es ampliamente conocido:

Sea $Y = X\beta + e$, donde $Y = (Y_1,\ldots,Y_n)'$ es una variable aleatoria n-dimensional (variable observable). Las observaciones concretas de las Y_i se denotan por letras minúsculas $y_1,\ldots,y_n$.

$X = (x_{ij})$ es una matriz nxk de rango k (matriz de diseño) que permanece constante durante todo el problema de inferencia. $\beta = (\beta_1,\ldots,\beta_k)'$ vector de "coeficientes de regresión" (desconocidos) que se trata de estimar.

$e = (e_1,\ldots,e_n)'$ es un vector de variables aleatorias (vector de error). Suponemos que sigue una distribución normal n-variante $N(0,\sigma^2 I)$ (por tanto, las e_i son independientes idénticamente distribuidas según una $N(0,\sigma^2)$). La varianza σ^2 es, en general, desconocida.

El problema consiste en estimar β y σ dada una observación $(y_1,\ldots,y_n)'$ de $(Y_1,\ldots,Y_n)'$.

Podemos llegar a resultados equivalentes a los obtenidos mediante el enfoque frecuentista si utilizamos distribuciones a priori impropias: En concreto, se supone que las β_i y $\log \sigma$ tienen distribuciones "uniformes" e "independientes" lo cual implica:

$$\pi(\beta,\sigma) \propto \frac{1}{\sigma} \quad (\beta_i \in R, \quad \sigma \in R^+)$$

El empleo de distribuciones a priori impropias tiene ciertas implicaciones teóricas que se comentarán en el apartado siguiente.

Bajo las suposiciones anteriores la distribución a posteriori que resulta, después de ciertos cálculos, es:

$$\pi(\beta,\sigma|y) \propto \frac{1}{\sigma^{n+1}} \exp\{-\frac{1}{2\sigma^2}[\nu s^2 + (\beta-\hat{\beta})'X'X(\beta-\hat{\beta})]\}$$

En el caso de que estemos interesados principalmente en β, se puede comprobar con facilidad que $\pi(\beta|\sigma,y)$ es una normal k-variante con media $\hat{\beta}$ y matriz de covarianzas $(X'X)^{-1}\sigma^2$. No obstante, como σ no es conocida en general, el procedimiento usual consiste en integrar respecto a σ en $\pi(\beta,\sigma|y)$ para obtener la marginal a posteriori:

$$\pi(\beta|y) \propto \int_0^\infty \pi(\beta,\sigma|y)d\sigma \propto [\nu s^2 + (\beta-\hat{\beta})'X'X(\beta-\hat{\beta})]^{-n/2}$$

que es una t de Student k-variante.

Análogamente puede obtenerse la marginal de σ, $\pi(\sigma|y)$ (que resulta ser una gamma invertida) y la de $\beta_{r+1},\ldots,\beta_s$ para $0 \leq r < s \leq n$ (ver Lindley (1971) y Zellner (1971)). A partir de $\pi(\beta_{r+1},\ldots,\beta_s|y)$, pueden hacerse inferencias sobre los correspondien tes parámetros en términos de la distribución F de Snedecor obteniéndose regiones de confianza bayesianas que tienen exactamente la misma forma que las regiones (del mismo coeficiente) obtenidas por métodos frecuentistas. También se puede obtener una interpretación bayesiana de los tests usuales basados en la F de Snedecor.

2.2. Distribuciones a priori impropias

Como se indicó en el apartado anterior, el enfoque bayesiano del modelo lineal, utiliza frecuentemente "distribuciones" a priori impropias sobre el espacio paramétrico Ω, entendiendo por tales, las medidas sobre Ω que verifican $\int_\Omega d\pi(\theta) = \infty$.

Se plantea, por tanto, la conveniencia de estudiar cómo afecta esta nueva situación al desarrollo de nuestra teoría en la que, hasta ahora, sólo se han empleado distribuciones a priori propias.

A este respecto pueden hacerse las siguientes observaciones:

(a) El uso de distribuciones a priori impropias lleva asociadas algunas dificultades de tipo general, como es la posible aparición de las llamadas "paradojas de marginalización" y el hecho de que la distribución a posteriori que se calcula mediante la fórmula de Bayes ya que no puede interpretarse como la distribución condicionada de θ dado x, cuando π es impropia.

(b) En lo que se refiere directamente a nuestra teoría, debe señalarse en primer lugar que la definición propuesta para la estabilidad de un modelo (apartado II.2.3) se mantiene en la nueva situación con una interpretación idéntica. Lo mismo se puede afirmar de las definiciones de robustez de los estimadores, regiones de confian za y tests de hipótesis bayesianos que se proponen más adelantes (secciones III.3 y III.4).

(c) Igualmente los teoremas II.3, II.4 y II.5 siguen siendo válidos con iguales enunciados y demostraciones (y serán de especial utilidad en esta sección).

Sin embargo, cuando π es impropia, ya no es válido el teorema II.6 pues la convergencia uniforme de las funciones del integrando no implica la convergencia de las correspondientes integrales cuando el espacio es de medida infinita.

2.3. Estabilidad de la inferencia bayesiana en el modelo lineal

El primer problema que se plantea es establecer qué tipos de desviaciones (respecto de las hipótesis usuales) conviene estudiar o, en otros términos, que conjuntos V de verosimilitudes se deben considerar.

Las principales desviaciones de la normalidad pueden resumirse en los siguientes fenómenos:

(a) Aparición ocasional (con una cierta probabilidad) de observaciones fuertemente erróneas ("outliers"), en el sentido de que están generadas por una distribución distinta de la teórica (en general, con mayor varianza). Esta situación puede concretarse mediante los "modelos de contaminación" ("Gross - error model") que estudiamos, con diversas variantes, en el primer apartado de esta sección.

(b) La distribución básica está sistemáticamente perturbada por un "error de medida" aleatorio de distribución desconocida reflejado en una variable aleatoria δ, de forma que ahora el "error" en $Y = X\beta + e$ es suma de dos variables aleatorias independientes $e = \xi+\delta$ donde ξ se distribuye, según las hipótesis usuales, $N(0,\sigma^2 I)$. Imponiendo ciertas condiciones sobre la distribución de δ obtendremos la estabilidad del modelo (apartado 2.3.2).

2.3.1. Modelos de contaminación

A continuación demostraremos, utilizando el teorema II.4, la estabilidad de una primera variante de estos modelos.

TEOREMA III.1. Sea $V_1 = \{f_c(.|\theta),\ \theta \in \Omega\}_{c\in I}$, donde $\theta = (\beta,\sigma) \in \Omega = R^k \times (0,+\infty)$, $I = [1,+\infty)$ y $f_c(y|\theta) = f_c(y|\beta,\sigma) = (1-\varepsilon)N(y;X\beta,\sigma^2 I) + \varepsilon N(y;X\beta,c^2\sigma^2 I_n)$, siendo $N(y;\mu,\Sigma)$ la función de densidad en el punto $y = (y_1,\dots,y_n)'$ de la normal n-variante de media μ y matriz de covarianzas Σ, I_n la matriz unidad de orden n y $\varepsilon \in (0,1)$ una constante conocida que indica la "proporción de contaminación".

Entonces el modelo definido por V_1 y por la distribución a priori impropia $\pi(\beta,\sigma) \propto \frac{1}{\sigma}$ es estable en V_1.

Demostración:

Bastará comprobar que se verifican en este caso las hipótesis del teorema II.4:

La hipótesis (i) se cumple pues, obviamente $f_c(y|\beta,\sigma) = (1-\varepsilon)N(y;X\beta,\sigma^2 I_n) + \varepsilon N(y;X\beta,c^2\sigma^2 I_n)$ es función continua de c en $[1+\infty)$ para y,β,σ fijos.

La "condición de frontera" dada por la hipótesis (ii) se verifica también pues $F_c(.|\beta,\sigma)$ no converge débilmente hacia ninguna distribución cuando $c \to \infty$.

Por otra parte, para toda sucesión $\{c_r\} \subset [1,\infty)$ tal que $c_r \xrightarrow[r \to \infty]{} c^*$, $\exists c_o$ tal que $c_r \leq c_o$ $\forall r \in \mathbb{N}$, luego

$$f_{c_r}(y|\beta,\sigma) = (1-\varepsilon)\frac{1}{\sigma^n(2\pi)^{n/2}} \exp\left[-\frac{1}{2\sigma^2}(y-X\beta)'(y-X\beta)\right] +$$

$$+ \varepsilon \frac{1}{c_r^n\sigma^n(2\pi)^{n/2}} \cdot \exp\left[-\frac{1}{2c_r^2\sigma^2}(y-X\beta)'(y-X\beta)\right] \leq$$

$$\leq \frac{1}{\sigma^n(2\pi)^{n/2}} \exp\left[-\frac{1}{2\sigma^2}(y-X\beta)'(y-X\beta)\right] + \frac{1}{\sigma^n(2\pi)^{n/2}}$$

$$\exp\left[-\frac{1}{2c_o^2\sigma^2}(y-X\beta)'(y-X\beta)\right]$$

que es π-integrable en Ω, pues es suma de dos verosimilitudes de V_1 (salvo constantes multiplicativas). Así pues, se cumple también la hipótesis (iii) del teorema II.4 y puede concluirse la estabilidad del modelo en V_1.

De manera absolutamente análoga se puede deducir la estabilidad del modelo que se obtiene considerando que α es fijo $(\alpha > 1)$ y ε es variable en $[0,\varepsilon^*]$. El correspondiente conjunto de verosimilitudes sería, por tanto,

$V_2 = \{f_\varepsilon(.|\theta),\ \theta \in \Omega\}_{\varepsilon\in[0,\varepsilon^*]}$, siendo $\varepsilon^* \in (0,1)$ una constante prefijada y $f_\varepsilon(y|\theta) = f_\varepsilon(y|\beta,\sigma) = (1-\varepsilon)N(y;X\beta,\sigma^2 I_n) +$

$$+ \varepsilon N(y;X\beta,c^2\sigma^2 I_n)$$

Para estudiar el caso en que ε y c son variables, debemos elegir adecuadamente el conjunto de índices I con objeto de evitar "duplicaciones" en la parametrización (i.e. para que $F_{\alpha_1}(.|\theta_1) \neq F_{\alpha_2}(.|\theta_2)$ cuando $(\alpha_1,\theta_1) \neq (\alpha_2,\theta_2)$) y de incluir la verosimilitud "teórica" $N(y;X\beta,\sigma^2 I)$ ya que principalmente nos interesará considerar la estabilidad del modelo en ella.

El resultado que obtenemos es el siguiente:

<u>TEOREMA III.2</u>. Sea $V_3 = \{f_\alpha(.|\theta),\ \theta \in \Omega\}_{\alpha \in I}$, donde

$\alpha = (\varepsilon,c) \in I = (0,1] \times (1,c^*] \cup \{(0,1)\} \subset \mathbb{R}^2$

$f_\alpha(y|\theta) = f_\alpha(y|\beta,\sigma) = (1-\varepsilon)\, N(y;X\beta,\sigma^2 I_n) + \varepsilon N(y;X\beta,c^2\sigma^2 I_n)$.

Entonces el modelo definido por V_3 y por la distribución a priori impropia $\pi(\beta,\sigma) \propto \frac{1}{\sigma}$ (en $\Omega = \mathbb{R}^k \times (0,+\infty)$) es estable en V_3.

<u>Demostración</u>:

En primer lugar, obsérvese que en este caso no es aplicable el teorema II.4, pues no se cumple la segunda hipótesis al aparecer en la frontera distribuciones "repetidas". En efecto: $\forall c \in (1,c^*)$ y para toda sucesión $\{\alpha_r\} \subset I$, con $\alpha_r = (\varepsilon_r,c_r) \xrightarrow[r\to\infty]{} (0,c)$ se tiene $f_{\alpha_r}(y|\beta,c) \xrightarrow[r\to\infty]{} f_{(0,1)}(y|\beta,\sigma) \equiv N(y;\beta,\sigma)$.

Sin embargo, podemos asegurar que la estabilidad de este modelo utilizando el teorema II.4:

La hipótesis (i) se verifica pues

$$|\phi_{\alpha,\theta}(t)| = |(1-\varepsilon)\exp\left[it'X\beta - \frac{\sigma^2}{2}t't\right] + \varepsilon \exp\left[it'X\beta - \frac{c^2\sigma^2}{2}t't\right]| \leq$$
$$\leq |\exp\left[it'X\beta - \frac{\sigma^2}{2}t't\right]| + |\exp\left[it'X\beta - \frac{\sigma^2}{2}t't\right]|$$

que es integrable (como función de $t = (t_1,\ldots,t_n)'$) en R^n.

En cuanto a la hipótesis (ii), se tiene

$$f_\alpha(y|\beta,\sigma) \leq \frac{1}{\sigma^n(2\pi)^{n/2}} \exp\left[-\frac{1}{2\sigma^2}(y-X\beta)'(y-X\beta)\right] +$$

$$+ \frac{1}{\sigma^n(2\pi)^{n/2}} \exp\left[-\frac{1}{2c^{*2}\sigma^2}(y-X\beta)'(y-X\beta)\right]$$

y esta función es π-integrable en Ω.

Se concluye, por tanto, que (π, V_3) es estable en V_3.

Discusión y comentarios

(a) El rango de variación de c ha debido limitarse a un intervalo finito $(1,c^*]$ (lo que no supone una fuerte restricción en la práctica), pues si fuera $(1,+\infty)$, la hipótesis (i) seguiría verificándose pero, en principio, ya no podríamos garantizar (ii) pues $N(y;X\beta,\sigma^2 I_n)$ podría aparecer también como límite de $(1-\varepsilon_r)N(y;X\beta,\sigma^2 I_n) +$ $+ \varepsilon_r N(y;X\beta,c_r^2\sigma^2 I_n)$ cuando $\varepsilon_r \to 0$ y $c_r \to \infty$, y no parece fácil acotar esta sucesión.

(b) Como puede observarse, la idea intuitiva que subyace en los modelos V_1, V_2 y V_3, implica que, o bien TODAS las observaciones de la muestra $y_1,\ldots,y_n$ proceden de la distribución teórica (con probabilidad ε), o bien de la distribución perturbada $N(X\beta,c^2\sigma^2 I_n)$, y así, la "contaminación" se da "en bloque" (para toda la muestra) o no se da. Sin embargo, en algunos casos puede ser más realista suponer que la contaminación puede darse elemento a elemento, es decir, que las $Y_1,\ldots,Y_n$ son variables aleatorias independientes verificando:

$$Y_i \sim (1-\varepsilon)N\left(\sum_{j=1}^k x_{ij}\beta_j,\sigma^2\right) + \varepsilon N\left(\sum_{j=1}^k x_{ij}\beta_j,c^2\sigma^2\right), \quad i=1,\ldots,n.$$

Por tanto, las correspondientes verosimilitudes serán del tipo

$$f_{(\varepsilon,c)}(y|\beta,\sigma) = \prod_{i=1}^{n} \Big[(1-\varepsilon)N(y_i; \sum_{j=1}^{k} x_{ij}\beta_j,\sigma^2) + $$
$$+ \varepsilon N(y_i; \sum_{j=1}^{k} x_{ij}\, \beta_j, c^2\sigma^2)\Big]$$

Podemos considerar, como antes, los casos:

(b_1) $c \in [1,+\infty)$ y $\varepsilon \in (0,1)$ fijo y conocido

(b_2) $\varepsilon \in [0,1]$ y $c \in (1,+\infty)$ conocido

(b_3) $(\varepsilon,c) \in (0,1] \times (1,c^*] \cup \{(0,1)\}$

En las situaciones (b_1) y (b_2) se puede concluir la estabilidad de los respectivos modelos de manera completamente similar a como se hizo anteriormente con $\mathcal{V}_1$ y $\mathcal{V}_2$ (utilizando el teorema II.4). En cuanto al caso (b_3) puede deducirse también la estabilidad mediante el teorema II.5; el proceso es análogo al que se ha utilizado para demostrar la estabilidad en $\mathcal{V}_3$ y no se repetirá aquí.

(c) En todo el análisis anterior nos hemos limitado al caso de que la variable "alternativa" sea del tipo $N(X\beta, c^2\sigma^2 I_n)$. Esta es una hipótesis plausible aunque, en algún aspecto, puede parecer demasiado restrictiva y tiene el inconveniente (ya comentado en el capítulo II) de restringirse a una familia "paramétrica". En cualquier caso, puede afirmarse que dicha hipótesis no es esencial: Basta con que las funciones características de todas las posibles distribuciones alternativas estén mayoradas (en módulo) por una misma función integrable y que cualquier sucesión convergente de verosimilitudes alternativas (para todo y fijo) esté acotada por una función π-integrable en

Ω; entonces, estas mismas propiedades se verifican en las distribuciones completas $(1-\varepsilon)\ N(y;X\beta,\sigma^2 I_n) + \varepsilon f(y;\beta,\sigma)$, y la aplicación inmediata del teorema II.5 nos lleva a concluir la estabilidad.

2.3.2. Modelo de convolución.

Este modelo se basa en la idea de que la variable aleatoria $Y = (Y_1,\dots,Y_n)'$ es suma de otras dos η y δ siendo $\eta \sim N(X\beta,\sigma^2 I_n)$ y δ, independiente de η, la variable aleatoria "perturbadora". Imponiendo ciertas condiciones sobre la distribución de ésta última podemos deducir la estabilidad del modelo. En concreto, obtenemos el siguiente resultado:

TEOREMA III.3. Sea el modelo (π,V), con $V = \{f_\alpha(.|\beta,\sigma),\ (\beta,\sigma) \in R^n \times R^+\}_{\alpha \in I}$ cumpliendo que $\forall\alpha \in I$ la distribución de Y_α, con densidad $f_\alpha(.|\beta,\sigma)$, corresponde a la suma de dos variables aleatorias independientes, $Y_\alpha = \eta + \delta_\alpha$, con $\eta \sim N(X\beta,\sigma^2 I_n)$.

Supongamos que las distribuciones de las variables perturbadoras $\{\delta_\alpha\}_{\alpha\in I}$, son absolutamente continuas con densidades $\{p_\alpha\}_{\alpha\in I}$ y verifican

(i) Existe una función integrable $g : R^n \longrightarrow R^+$ tal que $|\phi_\alpha| \le g$, $\forall\alpha \in I$ siendo ϕ_α la función característica de δ_α.

(ii) Existe una función integrable $h : R^n \longrightarrow R^+$ tal que $p_\alpha \le h$, $\forall\alpha \in I$ siendo además $\psi(y;\beta,\sigma) = \int_{R^n} N(z;X\beta,\sigma^2 I_n)h(y-z)dz$ función de (β,σ) π-integrable en $R^n \times R^+$, $\forall y \in R^n$.

Entonces el modelo $(\pi, \mathcal{V})$ es estable en $\mathcal{V}$.

Demostración:

La función característica de $Y_\alpha = \eta + \delta_\alpha$, $\forall\theta = (\beta,\sigma)$ fijo, es:

$$\phi_{\alpha,\theta}(t) = \exp\left[it'X\beta - \frac{1}{2}\sigma^2 t't\right] \cdot \phi_\alpha(t), \quad y$$

$$|\phi_{\alpha,\theta}(t)| = \exp\left[-\frac{1}{2}\sigma^2 t't\right] |\phi_\alpha(t)| \leq \exp\left[-\frac{1}{2}\sigma^2 t't\right] g(t)$$

Por otra parte, la densidad de Y_α es

$$f_\alpha(y|\beta,\sigma) = \int_{\mathbb{R}^n} N(z;X\beta,\sigma^2 I_n)\, p_\alpha(y-z)dz,$$

$$|f_\alpha(y|\beta,\sigma)| \leq \int_{\mathbb{R}^n} N(z;X\beta,\sigma^2 I_n)\, h(y-z)dz,$$ que es π-integrable en $\Omega = \mathbb{R}^n \times \mathbb{R}^+$, $\forall y \in \mathbb{R}^n$.

Por tanto, nuevamente la aplicación del teorema II.5 nos lleva a concluir la estabilidad.

Obsérvese que la expresión exacta de las p_α y ϕ_α puede ser, en general, desconocida ya que, para establecer el teorema sólo necesitamos asegurar $p_\alpha \leq h$ y $|\phi_\alpha| \leq g$.

2.4. Estabilidad en las marginales

Queda por resolver ahora un problema interesante: Frecuentemente (ver apartado II.2.1) no se maneja la densidad a posteriori conjunta $\pi(\beta_1,\ldots,\beta_k,\sigma|y)$ sino que algunos de los parámetros (por ejemplo, σ) se consideran parámetros secundarios ("nuisance parameters"). En consecuencia, se utiliza la densidad a posteriori marginal resultante de integrar con respecto a ellos.

Naturalmente, la pregunta que se plantea es: ¿Se conserva la estabilidad de la distribución conjunta en las marginales? Concretando más, si una sucesión de distribuciones a posteriori "conjuntas" converge débilmente hacia otra distribución a posteriori, ¿convergerán también las sucesiones de marginales hacia las correspondientes marginales del límite?. La respuesta es afirmativa como consecuencia directa del siguiente resultado general:

<u>LEMA III.1</u>. Sea $\{F_r\}_{r\in N}$, con $F_r : R^n \longrightarrow [0,1]$, una sucesión de funciones de distribución. Supongamos que $F_r \xrightarrow[r\to\infty]{d} F_o$ (F_o propia). Entonces $F_{r,j} \xrightarrow[r\to\infty]{d} F_{o,j}$, siendo $F_{r,j}$ para $j=1,\ldots,n$ la función de distribución marginal unidimensional j-ésima de F_r, $\forall r \in N \cup \{o\}$.

<u>Demostración</u>:

$F_{r,j} \xrightarrow[r\to\infty]{d} F_{o,j} \iff$ Para toda función $u : R \longrightarrow R$ contínua y acotada, se tiene $\int_R u(x_j)\, dF_{r,j}(x_j) \xrightarrow[r\to\infty]{} \int_R u(x_j) dF_{o,j}(x_j)$.

Por otra parte se cumple

$$\int_R u(x_j) d\, F_{r,j}\,(x_j) = \int_{R^n} u(x_j)\; d\; F_r(x_1,\ldots,x_n)$$

Esta igualdad puede comprobarse fácilmente verificándola en primer lugar para funciones indicatrices y siguiendo después el proceso usual (funciones simples, paso al límite,...).

A partir de aquí la demostración es trivial, pues por cumplirse $F_r \xrightarrow[r\to\infty]{d} F_o$, se tiene también

$$\int_{R^n} u(x_j) dF_r(x_1,\ldots,x_n) \xrightarrow[r\to\infty]{} \int_{R^n} u(x_j) dF_o(x,\ldots,x_n) =$$
$$= \int_R u(x_j) d\; F_{o,j}(x_j).$$

3. ROBUSTEZ DE LOS ESTIMADORES

3.1. Planteamiento del problema. Concepto de Robustez.

Hasta ahora nos hemos limitado a estudiar la estabilidad del modelo, que puede considerarse como una robustez "básica" o "primaria" inherente a la distribución a posteriori, según se indicó en el apartado 2.4.2 del capítulo anterior.

Cuando, a partir de la distribución a posteriori, obtenemos un estimador puntual pasa el parámetro θ, tiene sentido preguntarnos por la "robustez" de este nuevo proceso.

Consideremos, por tanto, el problema planteado con los mismos elementos y notación del apartado II.2.2.

Dado un modelo $(\pi, \mathcal{V})$, un estimador es una función $\hat{\theta} = \hat{\theta}(\alpha, x)$ (con valores en $\mathbb{R}^k$) de la densidad a posteriori $\pi_\alpha(.\mid x)$.

La definición de robustez que proponemos es la que parece más acorde con nuestro planteamiento general:

DEFINICION III.1. Sea un modelo $(\pi, \mathcal{V})$ y un estimador $\hat{\theta} = \hat{\theta}(\alpha, x)$ con $\alpha \in I$ y $x \in X_\alpha = \{x \in \mathbb{R}^n \mid 0 < \int_\Omega f_\alpha(x|\theta)\, d\pi(\theta) < \infty\}$.

Diremos que $\hat{\theta}$ es cualitativamente robusto en $f_{\alpha_o} \in \mathcal{V}$ cuando, para toda sucesión de verosimilitudes $\{f_{\alpha_r}\}_{r \in \mathbb{N}}$ tal que

$F_{\alpha_r}(.|\theta) \xrightarrow[r \to \infty]{d} F_{\alpha_o}(.|\theta) \quad \forall \theta \in \Omega$ se verifica

$$\hat{\theta}_r = \hat{\theta}(\alpha_r, x) \xrightarrow[r \to \infty]{} \hat{\theta}_o = \hat{\theta}(\alpha_o, x) \quad \forall x \in \bigcap_{r=o}^{\infty} X_{\alpha_r}.$$

Puede observarse que esta definición no exige la estabilidad del modelo; incluso podrían darse situaciones en las que el modelo fuese inestable y un determinado estimador fuese robusto. Con todo, estos casos deben considerarse como "patológicos" y de escasa importancia práctica. La situación que parece más interesante es aquella en que el modelo es estable y nos preguntamos por la robustez de un estimador concreto; este es el punto de vista que adoptamos a lo largo de esta sección.

En resumen, el proceso natural en el análisis de la robustez de un método de estimación puntual sería, según nuestro enfoque, estudiar en primer lugar la estabilidad del modelo y a continuación la robustez de los estimadores que exigirá, en general, condiciones adicionales.

Por otra parte, obsérvese que la definición III.1 puede traducirse formalmente en términos de continuidad siguiendo un procedimiento análogo al de los teoremas II.1 y II.2 que no repetiremos aquí.

En los restantes apartados de esta sección se establecen teoremas que proporcionan condiciones suficientes para asegurar la robustez de los principales estimadores.

3.2. Moda

La moda a posteriori es un estimador ampliamente utilizado en inferencia bayesiana donde puede considerarse que juega un papel análogo al de los estimadores de máxima verosimilitud en la teoría frecuentista.

Con objeto de que el estimador esté bien definido nos limitaremos a considerar densidades a posteriori que alcanzan su máximo en un valor único (densidades unimodales).

Cuando π es μ_L-contínua, la moda será, en general, distinta según que se elija la densidad a posteriori respecto a π o respecto a μ_L (ver cap. I). Esta última representación es la más frecuente en la práctica. No obstante, el teorema que demostraremos a continuación puede aplicarse indistintamente a uno u otro caso.

TEOREMA III.4. Dado un modelo (π, V), con $\Omega = R^k$ sea $\phi_{\alpha,x}$ la función característica de la distribución a posteriori $F_\pi^{(\alpha)}(.|x)$. Supongamos que

(i) (π, V) es estable en $f_{\alpha_o} \in V$.

(ii) Para toda sucesión $S = \{f_{\alpha_r}\}_{r \in N} \subset V$ tal que $F_{\alpha_r}(.|\theta) \xrightarrow[r \to \infty]{d} F_{\alpha_o}(.|\theta)$ $\forall \theta \in \Omega$ existe, $\forall x \in \bigcap_{i=o}^{\infty} \chi_{\alpha_i}$, una función $g_x^s : R^k \longrightarrow R$ tal que $|\phi_{\alpha_r, x}| \leq g_x^s$ $\forall r \in N \cup \{o\}$.

(iii) Las densidades a posteriori son unimodales y continuas y verifican $\lim_{\|\theta\| \to \infty} \pi_\alpha(\theta|x) = 0$ $\forall \alpha \in I$, $\forall x \in \chi_\alpha$.

Demostración:

Sea $S = \{f_{\alpha_r}\}_{r \in N}$ tal que $F_{\alpha_r}(.|\theta) \xrightarrow[r \to \infty]{d} F_{\alpha_o}(.|\theta)$ $\forall \theta \in \Omega$
Esto implica por la hipótesis (ii) (ver la demostración del teorema II.5) que

$$\pi_{\alpha_r}(.|x) \xrightarrow[r \to \infty]{} \pi_{\alpha_o}(.|x) \text{ uniformemente en } \Omega, \quad \forall x \in \bigcap_{r=o}^{\infty} \chi_{\alpha_r}$$

Sea $\hat{\theta}_r = \hat{\theta}(\alpha_r, x)$ la moda de $\pi_{\alpha_r}(.|x)$ $\forall r \in N \cup \{o\}$.

Probaremos en primer lugar $\pi_{\alpha_r}(\hat{\theta}_r|x) \xrightarrow[r \to \infty]{} \pi_{\alpha_o}(\hat{\theta}_o|x)$.

En efecto

Si $\pi_{\alpha_r}(\hat\theta_r|x) \le \pi_{\alpha_o}(\hat\theta_o|x)$ se tiene $|\pi_{\alpha_r}(\hat\theta_r|x) - \pi_{\alpha_o}(\hat\theta_o|x)| \le$ $\le |\pi_{\alpha_r}(\hat\theta_o|x) - \pi_{\alpha_o}(\hat\theta_o|x)|$. Si $\pi_{\alpha_r}(\hat\theta_r|x) > \pi_{\alpha_o}(\hat\theta_o|x)$ entonces $|\pi_{\alpha_r}(\hat\theta_r|x) - \pi_{\alpha_o}(\hat\theta_o|x)| < |\pi_{\alpha_r}(\hat\theta_r|x) - \pi_{\alpha_o}(\hat\theta_r|x)|$. En cualquier caso

$$\forall r \in \mathbb{N},\ |\pi_{\alpha_r}(\hat\theta_r|x) - \pi_{\alpha_o}(\hat\theta_o|x)| \le \text{máx}\,\{|\pi_{\alpha_r}(\hat\theta_o|x) - \pi_{\alpha_o}(\hat\theta_o|x)|,$$

$$|\pi_{\alpha_r}(\hat\theta_r|x) - \pi_{\alpha_o}(\hat\theta_r|x)|\} \le \sup_{\theta\in\Omega} |\pi_{\alpha_r}(\theta|x) - \pi_{\alpha_o}(\theta|x)| \xrightarrow[r\to\infty]{} 0$$

por la convergencia uniforme.

Supongamos ahora que no se cumpliese $\hat\theta_r \xrightarrow[r\to\infty]{} \hat\theta_o$.

Obsérvese que la sucesión $\{\hat\theta_r\}_{r\in\mathbb{N}}$ está acotada pues, en caso contrario, existiría una subsucesión $\{\hat\theta_{r_p}\}_{p\in\mathbb{N}}$ con $\|\hat\theta_{r_p}\| \xrightarrow[p\to\infty]{} \infty$ y, por la hipótesis (iii) se tendría $\pi_{\alpha_{r_p}}(\hat\theta_{r_p}|x) \xrightarrow[p\to\infty]{} 0 = \pi_{\alpha_o}(\hat\theta_o|x)$ lo cual es imposible por ser $\hat\theta_o$ la moda de $\pi_{\alpha_o}(.|x)$.

Así pues, existe una subsucesión $\{\hat\theta_{r_m}\}$ y un $\tau \in \Omega$, $\tau \ne \hat\theta_o$ tales que $\hat\theta_{r_m} \xrightarrow[m\to\infty]{} \tau$; entonces

$$|\pi_{\alpha_{r_m}}(\hat\theta_{r_m}|x) - \pi_{\alpha_o}(\tau|x)| \le |\pi_{\alpha_{r_m}}(\hat\theta_{r_m}|x) - \pi_{\alpha_o}(\hat\theta_{r_m}|x)| +$$

$$+ |\pi_{\alpha_o}(\hat\theta_{r_m}|x) - \pi_{\alpha_o}(\tau|x)| \le \sup_{\theta\in\Omega} |\pi_{\alpha_{r_m}}(\theta|x) - \pi_{\alpha_o}(\theta|x)| +$$

$$+ |\pi_{\alpha_o}(\hat\theta_{r_m}|x) - \pi_{\alpha_o}(\tau|x)|.$$

El primer sumando tiende a cero cuando $m \to \infty$ por la convergencia uniforme y el segundo por la continuidad de π_{α_o}. Así, se ha ob-

tenido $\pi_{\alpha_{r_m}}(\hat{\theta}_{r_m}|x) \xrightarrow[m \to \infty]{} \pi_{\alpha_o}(\tau|x)$, con lo cual $\pi_{\alpha_o}(\tau|x) =$
$= \pi_{\alpha_o}(\hat{\theta}_o|x)$ lo que contradice la hipótesis de ser π_{α_o} unimodal.

Se concluye $\hat{\theta}_r \xrightarrow[r \to \infty]{} \hat{\theta}_o$ y, por tanto, la robustez del estimador.

Discusión y comentarios

(a) La convergencia uniforme de las densidades a posteriori desempeña un papel fundamental en la demostración del teorema. No obstante, esta condición, por sí sola, no es suficiente para asegurar la convergencia de las modas como prueba el siguiente contraejemplo:

Sea la función de densidad $f : \mathbb{R} \longrightarrow \mathbb{R}$ definida por

$$f(x) = \begin{cases} \frac{h_n}{a_n}(x-2n+a_n) & \text{si } x \in [2n-a_n, 2n) \quad (\forall n \in \mathbb{N}) \\ -\frac{h_n}{a_n}(x-2n-a_n) & \text{si } x \in [2n, 2n+a_n] \\ 2(x+1/2) & \text{si } x \in [-1/2, 0) \\ -2(x-1/2) & \text{si } x \in [0, 1/2] \end{cases}$$

siendo $h_n = 1 - (1/2)^n$, $a_n = 1/(2^n-1)$ $\forall n \in \mathbb{N}$.

Definamos ahora la sucesión de densidades $\{f_r\}_{r \in \mathbb{N}}$:

$$f_r(x) = \begin{cases} \dfrac{k_r}{b_r}\,(x-2r+b_r) & \text{si} \quad x \in [2r-b_r, 2r| \\ \dfrac{k_r}{b_r}\,(x-2r-b_r) & \text{si} \quad x \in [2r, 2r+b_r] \\ f(x) & \text{en el resto de los puntos} \end{cases}$$

siendo $k_r = 1 + (1/2)^r$, $b_r = 1/(2^r+1)$, $\forall r \in \mathbb{N}$.

Es inmediato comprobar que $\sup_{x \in R} |f_r(x) - f(x)| \xrightarrow[r \to \infty]{} 0$ y, por tanto, $f_r \xrightarrow[r \to \infty]{} f$ uniformemente en $\mathbb{R}$ y, sin embargo,

$$\text{Moda}\ (f_r) = 2r \not\longrightarrow \text{Moda}\ (f) = 0$$

La hipótesis $\lim_{\|\theta\| \to \infty} \pi_\alpha(\theta|x) = 0$ se introduce precisamente para evitar situaciones como la del contraejemplo anterior y se puede debilitar exigiendo simplemente que, para toda sucesión convergente de densidades a posteriori la correspondiente sucesión de modas esté contenida en un compacto.

(b) La hipótesis (ii) se introduce únicamente para asegurar la convergencia uniforme de las densidades a través de una condición relativamente sencilla y natural, aunque su verificación en un caso concreto puede resultar complicada. En este sentido, sería deseable asegurar la convergencia uniforme de las densidades a posteriori a través de hipótesis sobre el modelo (π, V) pero los enunciados resultantes serían demasiado complicados y artificales.

3.3. Mediana

La mediana es un estimador generalmente considerado como "robusto", si bien su empleo está más extendido en la metodología frecuentista que en la bayesiana.

A continuación establecemos un resultado auxiliar que nos permitirá demostrar de inmediato la robustez cualitativa de la mediana a posteriori.

LEMA III.2. Dada una sucesión $\{F_r\}_{r\in\mathbb{N}}$ de funciones de distribución (propias) en $\mathbb{R}$ sea μ_r una mediana de F_r $\forall r \in \mathbb{N}$ (es decir, $\lim_{x\to\mu_r^-} F_r(x) = F_r^-(\mu_r) \leq 1/2 \leq F(\mu_r)$).

Supongamos que $F_r \xrightarrow[r\to\infty]{d} F$ donde F es una función de distribución (propia) en $\mathbb{R}$.

Si μ es cualquier punto límite de la sucesión $\{\mu_r\}$ de medianas entonces μ es una mediana de F. En particular si μ es la única mediana de F entonces $\mu_r \xrightarrow[r\to\infty]{} \mu$.

Demostración:

Por ser μ punto límite de $\{\mu_r\}$, existe una subsucesión $\{\mu_{r_p}\}_{p\in\mathbb{N}}$ tal que $\mu_{r_p} \xrightarrow[p\to\infty]{} \mu$.

Sean x', x'' puntos de continuidad de F tales que $x' < \mu < x''$. Vamos a comprobar $F(x') \leq 1/2$, $F(x'') \geq 1/2$. En efecto $\mu_{r_p} \xrightarrow[p\to\infty]{} \mu \Longrightarrow \exists K_o \in \mathbb{N} \mid \forall p \geq K_o,\ x' < \mu_{r_p} < x''$.

Por la monotonía de las F_{r_p} se tiene

$$1/2 \geq F^-_{r_p}(\mu_{r_p}) \geq F_{r_p}(x')$$

Por tanto

$$\lim_{p\to\infty} F_{r_p}(x') = F(x') \leq 1/2$$

Análogamente

$$1/2 \leq F_{r_p}(\mu_{r_p}) \leq F_{r_p}(x'')$$

y

$$\lim_{p\to\infty} F_{r_p}(x'') = F(x'') \geq 1/2$$

Se sigue que

$$F^-(\mu) = \text{Sup}\,\{F(x) \mid x < \mu \ \text{ y } \ F \ \text{ es cont. en } x\} \leq 1/2 \leq$$
$$\leq F(\mu) = \text{Inf}\,\{F(x) \mid x > \mu \ \text{ y } \ F \ \text{ cont. en } x\}$$

Para probar la segunda parte del enunciado comprobaremos que, si μ es la única mediana de F, entonces μ es el único punto límite de la sucesión $\{\mu_r\}$, con lo cual se tendrá $\mu_r \xrightarrow[r\to\infty]{} \mu$.

En efecto, $\{\mu_r\}$ está acotada pues, en caso contrario, existiría una subsucesión $\{\mu_{r_m}\}$ tal que $\mu_{r_m} \xrightarrow[m\to\infty]{} \infty$ (resp. $-\infty$); eligiendo un punto de continuidad de F, x_o, tal que $F(x_o) > 1/2$ (resp. $F(x_o) < 1/2$) se tiene $F_r(x_o) \xrightarrow[r\to\infty]{} F(x_o)$, luego $\exists r_o \mid \forall r \geq r_o$ $F_r(x_o) > 1/2$ (resp. $< 1/2$).

Así, $\forall m \mid r_m \geq r_o$, $\mu_{r_m} \leq x_o$ (resp. $\mu_{r_m} \geq x_o$) en contradicción con $\mu_{r_m} \xrightarrow[m\to\infty]{} \infty$ (resp. $-\infty$).

Por tanto $\{\mu_r\}$ tiene un punto límite μ'. Repitiendo el razonamiento anterior se deduce que μ' es mediana de F y como, por hipótesis, la mediana es única, se tiene $\mu' = \mu$.

Se puede demostrar, con modificaciones triviales, un resultado análogo para cuantiles de orden arbitrario.

El resultado que nos interesa sobre la robustez de la mediana viene dado por el siguiente

TEOREMA III.5. Sea un modelo (π, V) con $\Omega = \mathbb{R}$, tal que la mediana de todas las distribuciones a posteriori es única.

Si (π, V) es estable en $f_{\alpha_o} \in V$ entonces la mediana a posteriori es un estimador cualitativamente robusto en f_{α_o}.

Demostración:

Es inmediata a partir del lema III.2.

Discusión y comentarios

(a) Este resultado coincide con lo que podría esperarse intuitivamente y viene a confirmar, en un nuevo sentido, la afirmación comúnmente admitida de que "la mediana es un estimador robusto".

(b) Introduciendo algunas modificaciones triviales se puede obtener un teorema análogo relativo a cualquier cuantil de la distribución a posteriori. Este resultado se utilizará en la sección 4 para estudiar la robustez de los intervalos de confianza.

3.4. Media

En este apartado establecemos dos teoremas referentes a la robustez cualitativa de la media a posteriori. El primero de ellos requiere el siguiente resultado auxiliar:

LEMA III.3. Sea $\{F_r\}_{r\in\mathbb{N}}$ sucesión de funciones de distribución (propias) y sea $\gamma_{pr} = \int_{\mathbb{R}} |x|^p dF_r(x)$. Supongamos que

(i) $F_r \xrightarrow[r\to\infty]{d} F$ siendo F f.d. propia.

(ii) Existe un $p > 0$ tal que la sucesión $\{\gamma_{pr}\}_{r\in\mathbb{N}}$ está acotada. Entonces

$$a_{sr} = \int_{\mathbb{R}} x^s d\, F_r(x) \xrightarrow[r\to\infty]{} \int_{\mathbb{R}} x^s dF(x) \quad \text{para } s < p$$

Este es un resultado ampliamente conocido. La demostración puede verse, por ejemplo, en Rao (1973) pág. 121.

La utilización de este lema es bastante obvia: En un modelo estable basta con imponer que las distribuciones a posteriori tengan momentos absolutos de orden $p > 1$ acotados por una constante, para asegurar que la media a posteriori es un estimador robusto.

Por tanto, puede enunciarse el siguiente

TEOREMA III.6. Sea un modelo (π, V) con $\Omega = \mathbb{R}$. Supongamos que

(i) (π, V) es estable en f_{α_o}

(ii) Para toda sucesión $\{f_{\alpha_r}\}_{r\in\mathbb{N}} \subset V$ tal que $F_{\alpha_r}(.|\theta) \xrightarrow[r\to\infty]{d} F_{\alpha_o}(.|\theta)$ $\forall\theta \in \Omega$ se verifica que $\forall x \in \bigcap_{r=1}^{\infty} X_{\alpha_r}$ $\exists\, K_x > 0$ y $p_x > 1$ cumpliendo

$$\int_{\Omega} |\theta|^{p_x}\, dF^{(\alpha_r)}(\theta|x) \leq K_x \quad \forall r \in \mathbb{N}$$

Entonces la media a posteriori es un estimador cualitativamente robusto en f_{α_o}.

Demostración:

Es inmediato a partir del lema III.3.

Discusión y comentarios

(a) La hipótesis (ii) está enunciada en su forma más general. Una condición suficiente para que se verifique es que π tenga soporte compacto y que para toda sucesión débilmente convergente de distribuciones a posteriori las correspondientes densidades estén uniformemente acotadas.

(b) Una idea natural es asegurar la acotación de los momentos a posteriori a través de condiciones impuestas directamente sobre el modelo pero nuevamente encontramos (análogamente al caso del teorema III.4) que nos veríamos obligados a imponer hipótesis exceisvamente complicadas y artificiosas.

(c) Se puede estudiar también la robustez de la media a posteriori sin necesidad de pasar por el lema III.3, mediante ideas similares a las utilizadas en el capítulo II, para garantizar la convergencia de las densidades. Este es el fundamento del siguiente resultado.

TEOREMA III.7. Sea un modelo definido por $V = \{f_\alpha(\cdot|\theta),\ \theta \in \Omega\}_{\alpha \in I}$ (con $I \subset R^m$) y la distribución a posteriori π en $\Omega \subset R^k$. Denotemos por $\hat{I}$ el cierre topológico de I en $\bar{R}^m$ siendo $\bar{R} = [-\infty, +\infty]$.

Supongamos que

(i) Para toda sucesión $\{\alpha_r\}_{r \in N}$ tal que $\lim_{r \to \infty} \alpha_r = c \in \hat{I} \setminus I$, se verifica que $\nexists \alpha \in I$ tal que $F_{\alpha_r}(.|\theta) \xrightarrow[r \to \infty]{d} F_\alpha(.|\theta) \quad \forall\theta \in \Omega$.

(ii) Para todo $x \in R^n$ $f_\alpha(x|\theta)$ es función contínua de (α,θ) en $I \times \Omega$.

(iii) Ω es compacto.

Entonces la media a posteriori es un estimador cualitativamente robusto en V.

<u>Demostración</u>:

Según se ha probado ya (ver teorema II.4, comentario (c)), las hipótesis (i), (ii) y (iii) implican la estabilidad del modelo y, más aún, permiten demostrar que

$$\lim_{r\to\infty} \alpha_r = \alpha_o \iff F_{\alpha_r}(.|\theta) \xrightarrow[r \to \infty]{d} F_{\alpha_o}(.|\theta) \quad \forall\theta \in \Omega \iff$$

$$\iff f_{\alpha_r}(x|\theta) \xrightarrow[r \to \infty]{} f_{\alpha_o}(x|\theta) \quad \forall x \in R^n,$$

$$\forall\theta \in \Omega \implies \pi_{\alpha_r}(\theta|x) \xrightarrow[r \to \infty]{} \pi_{\alpha_o}(\theta|x) \quad \forall\theta \in \Omega, \quad \forall x \in \bigcap_{i=o}^{\infty} X_{\alpha_i}.$$

Ahora, para demostrar la robustez de la media a posteriori, tomemos $\{\alpha_r\}$ cumpliendo $\alpha_r \xrightarrow[r \to \infty]{} \alpha_o \in I$. Entonces $J = \{\alpha_r\}_{r\in N} \cup \{\alpha_o\}$ es un compacto y $J \times \Omega$ también lo es.

Por otra parte, $\forall x \in \bigcap_{i=o}^{\infty} X_{\alpha_i}$, $\pi_\alpha(\theta|x) =$ $= [\int_\Omega f_\alpha(x|\theta)d\pi(\theta)]^{-1} f_\alpha(x|\theta)$ es función contínua de (α,θ) en virtud de la hipótesis (ii), con lo cual podemos asegurar que

$$\exists K_x > 0 \mid \pi_{\alpha_r}(\theta|x) \leq K_x \quad \forall r \in N, \quad \forall\theta \in \Omega$$

Por tanto $\theta\pi_{\alpha_r}(\theta|x) \leq \theta K_x$ que es función π-integrable en Ω (por ser compacto); aplicando el teorema de la convergencia dominada:

$$\int_\Omega \theta\pi_{\alpha_r}(\theta|x)d\pi(\theta) \xrightarrow[r \to \infty]{} \int_\Omega \theta\pi_{\alpha_o}(\theta|x)d\pi(\theta)$$

lo cual prueba la tesis del teorema.

3.5. Estimadores Bayes

Si nos situamos ahora en el contexto general, debemos introducir como elemento nuevo la función de pérdida:

$$L : \Omega \times \mathcal{D} \longrightarrow R$$

donde el espacio de acciones $\mathcal{D}$ es, en este caso, $\mathcal{D} = \Omega$.

Como es usual, diremos que $a^* = a^*(x)$ es el estimador Bayes de cuando

$$r_{\alpha,x}(a^*) = \min_{a \in \mathcal{D}} r_{\alpha,x}(a), \quad \text{siendo} \quad r_{\alpha,x}(a) = \int_\Omega L(\theta,a)\pi_\alpha(\theta|x)\,d\pi(\theta)$$

En la exposición que sigue supondremos, en todos los casos que este mínimo existe y es único.

Se trata de estudiar la robustez cualitativa de los estimadores así definidos.

Un ligero análisis del problema nos lleva a la conclusión de que el modo natural de proceder es el siguiente:

(a) Dar condiciones para asegurar

$$F_{\alpha_p}(.|\theta) \xrightarrow[r \to \infty]{d} F_{\alpha_o}(.|\theta) \quad \forall\theta \in \Omega \implies \pi_{\alpha_p}(\theta|x) \longrightarrow$$

$$\xrightarrow[r \to \infty]{} \pi_{\alpha_o}(\theta|x) \quad \forall x \in \bigcap_{i=o}^{\infty} X_{\alpha_i}.$$

Este problema ha aparecido reiteradamente en el desarrollo de la teoría y ya se han estudiado en el capítulo II las dos formas principales de abordarlo.

(b) A continuación, garantizar mediante hipótesis adecuadas que las $r_{\alpha,x}$ son contínuas y que

$$r_{\alpha_p,x}(a) = \int_\Omega L(\theta,a)\pi_\alpha(\theta|x)d\pi(\theta) \xrightarrow[p\to\infty]{} r_{\alpha_o,x}(a) =$$
$$= \int_\Omega L(\theta,a)\pi_{\alpha_o}(\theta|x)d\pi(\theta)$$

Se trata, en definitiva, de permutar el límite con la integral y ésto puede hacerse utilizando el teorema de la convergencia dominada, o bien, asegurando la convergencia uniforme de las $\pi_{\alpha_p}(.|x)$.

(c) Cuando se han conseguido los objetivos propuestos en (a) y (b), el problema se reduce a estudiar bajo qué condiciones la convergencia de una sucesión de funciones contínuas a otra función contínua implica la convergencia de los correspondientes mínimos al mínimo del límite.

Esta situación es completamente análoga a la que surgió en el estudio de la robustez de la moda a posteriori. Como se vió allí, interesa asegurar la convergencia uniforme de las $r_{\alpha_p,x}$ e imponer alguna condición suplementaria para evitar que los correspondientes mínimos "se vayan a infinito".

A partir de estas ideas obtenemos el siguiente

<u>TEOREMA III.8</u>. Sea un modelo $(\pi,\mathcal{V})$ con $\Omega = \mathbb{R}^k$ y una función de pérdida $L : \Omega \times \mathcal{D} \longrightarrow \mathbb{R}$ cumpliendo que, para toda distribución a

posteriori asociada al modelo, el correspondiente estimador Bayes existe y es único.

Supongamos que

(i) $(\pi, \mathcal{V})$ es estable en $f_{\alpha_o} \in \mathcal{V}$

(ii) Para toda sucesión $s = \{f_{\alpha_p}\}_{p \in N}$ $\mathcal{V}$ tal que $F_{\alpha_p}(.|\theta) \xrightarrow[p \to \infty]{d} F_{\alpha_o}(.|\theta)$ $\forall \theta \in \Omega$ existe, $\forall x \in \bigcap_{i=o}^{\infty} \chi_{\alpha_i}$, una función integrable $g_x^s : R^k \longrightarrow R$ cumpliendo $|\rho_{\alpha_p, x}| \leq g_x^s$ $\forall p \in N \cup \{o\}$, siendo $\rho_{\alpha,x}$ la función característica asociada a la distribución a posteriori $\pi_\alpha(.|x)$.

(iii) L es acotada y uniformemente contínua en $\Omega \times \Omega$.

(iv) Para toda sucesión $S = \{f_{\alpha_p}\}_{p \in N}$ tal que $F_{\alpha_p}(.|\theta) \xrightarrow[p \to \infty]{} F_{\alpha_o}(.|\theta)$ $\forall \theta \in \Omega$, existe, $\forall x \in \bigcap_{i=1}^{\infty} \chi_{\alpha_i}$, un conjunto compacto C_x^s Ω tal que el estimador Bayes $\hat{\theta}_p = a^*(\alpha_p, x) \in C_x^s$ $\forall p \in N$.

Entonces, el estimador Bayes es robusto en f_{α_o}.

<u>Demostración</u>:

$\forall \alpha \in I$ y $\forall x \in \chi_\alpha$, se verifica que $r_{\alpha,x}$ es función uniformemente contínua en $\mathcal{D} = \Omega$. En efecto, por la continuidad uniforme de L, dado $\varepsilon > 0$ $\exists \delta = \delta(\varepsilon)$ tal que $|L(\theta,a) - L(\theta,a')| \leq \varepsilon$ siempre que $\|a-a'\| \leq \delta$.

Así, se tiene:

$$|r_{\alpha,x}(a) - r_{\alpha,x}(a')| \leq \int_\Omega |L(\theta,a) - L(\theta,a')| \pi_\alpha(\theta|x) d\pi(\theta) \leq \varepsilon$$

siempre que $\|a-a'\| \leq \delta$.

Además, cuando $F_{\alpha_p}(.|\theta) \xrightarrow[p\to\infty]{d} F_{\alpha_o}(.|\theta)$ $\forall\theta \in \Omega$, se verifica

$$\forall x \in \bigcap_{i=o}^{\infty} X_{\alpha_i}, \quad r_{\alpha_p,x} \longrightarrow r_{\alpha_o,x} \quad \text{uniformemente en } \Omega.$$

Además, por (i) y (ii) (ver teorema III.4) tenemos que

$$\pi_{\alpha_p}(.|x) \xrightarrow[p\to\infty]{} \pi_{\alpha_o}(.|x) \quad \text{uniformemente en } \Omega, \text{ y}$$

$$\sup_{a\in\mathcal{D}} |r_{\alpha_p,x}(a) - r_{\alpha_o,x}(a)| \leq \sup_{a\in} \int_\Omega |L(\theta,a)|\,|\pi_{\alpha_p}(\theta|x) - \pi_{\alpha_o}(\theta|x)|\,d\pi(\theta) \leq$$

$$\leq \int_\Omega M \sup_{\theta\in\Omega} |\pi_{\alpha_p}(\theta|x) - \pi_{\alpha_o}(\theta|x)|d\pi(\theta) \xrightarrow[p\to\infty]{} 0$$

Sean ahora $\hat\theta_p$ y $\hat\theta_o$ los estimadores Bayes asociados a $\pi_{\alpha_p}(\theta|x)$ y $\pi_{\alpha_o}(\theta|x)$, respectivamente.

Supongamos que no se cumple $\hat\theta_p \xrightarrow[p\to\infty]{} \hat\theta_o$. Por la hipótesis (iv), $\exists C \subset \Omega$, C compacto, tal que $\hat\theta_p \in C$ $\forall p \in \mathbb{N}$. Entonces si $\hat\theta_p \not\longrightarrow \hat\theta_o$, $a_o \neq \hat\theta_o$, $a_o \in C$, y existe una subsucesión de $\{\hat\theta_p\}$, a la que designaremos por $\{\hat\theta_m\}$ tal que $\hat\theta_m \xrightarrow[m\to\infty]{} a_o$.

Esto implica $r_{\alpha_m,x}(\hat\theta_m) \xrightarrow[m\to\infty]{} r_{\alpha_o,x}(a_o)$ pues

$$|r_{\alpha_m,x}(\hat\theta_m) - r_{\alpha_o,x}(a)| \leq |r_{\alpha_m,x}(\hat\theta_m) - r_{\alpha_o,x}(\hat\theta_m)| +$$

$$+ |r_{\alpha_o,x}(\hat\theta_m) - r_{\alpha_o,x}(a_o)| \quad \text{donde} \quad |r_{\alpha_m,x}(\hat\theta_m) - r_{\alpha_o,x}(\hat\theta_m)| \leq$$

$$\leq \sup_{a\in\mathcal{D}} |r_{\alpha_m,x}(a) - r_{\alpha_o,x}(a)| \xrightarrow[m\to\infty]{} 0 \quad \text{y}$$

$$|r_{\alpha_o,x}(\hat\theta_m) - r_{\alpha_o,x}(a_o)| \xrightarrow[m\to\infty]{} 0 \quad \text{por la continuidad de } r_{\alpha_o,x}$$

Por otra parte, $r_{\alpha_p,x}(\hat\theta_p) \xrightarrow[p\to\infty]{} r_{\alpha_o,x}(\hat\theta_o)$ ya que

$$|r_{\alpha_p,x}(\hat\theta_p) - r_{\alpha_o,x}(\hat\theta_o)| \leq \max \{|r_{\alpha_p,x}(\hat\theta_p) - r_{\alpha_o,x}(\hat\theta_p)|,$$

$$|r_{\alpha_p,x}(\hat\theta_o) - r_{\alpha_o,x}(\hat\theta_o)|\} \xrightarrow[p \to \infty]{} 0$$

En definitiva, $r_{\alpha_o,x}(a_o) = r_{\alpha_o,x}(\hat\theta_o)$ con $a_o \neq \hat\theta_o$ lo cual contradice la unicidad del estimador Bayes.

Se concluye $\hat\theta_p \xrightarrow[p \to \infty]{} \hat\theta_o$ y, por tanto, la robustez cualitativa del estimador.

3.6. Conclusiones

Los resultados obtenidos coinciden en líneas generales con las ideas sugeridas por la intuición y la práctica estadística. Pueden concretarse en las siguientes consideraciones:

(a) La mediana es un estimador robusto siempre que el modelo sea estable.

(b) La robustez de la media a posteriori exige condiciones mucho más restrictivas.

En la teoría de la robustez "frecuentista" desarrollada por Hampel, la media muestral (que en muchos modelos coincide con la media a posteriori) aparece como ejemplo típico de estimador no robusto. Esto supone cierta discrepancia con la teoría presentada aquí, lo cual es perfectamente natural teniendo en cuanta que ambos desarrollos se basan en enfoques muy distintos.

(c) La moda a posteriori sólo es estimador robusto bajo condiciones muy fuertes.

(d) En suma, puede decirse, de una manera informal, que la ordenación de estos tres estimadores por orden creciente de robustez, es: moda, media, mediana.

(e) Para asegurar la robustez de un estimador Bayes genérico (ver teorema III.8) se exigen condiciones similares a las que se necesitan para la moda además de ciertas hipótesis de acotación y continuidad para la función de pérdida.

(f) En los teoremas relativos a la mediana, la moda y los estimadores Bayes, hemos impuesto como hipótesis previa la unicidad del estimador para todas las posibles distribuciones a posteriori. En modelos más complicados pueden surgir otros problemas suplementarios de "no robustez" debido al hecho de que puede desaparecer la unicidad al producirse pequeñas alteraciones en la verosimilitud.

(g) Finalmente, debe resaltarse el importante papel que desempeña la hipótesis de compacidad del espacio paramétrico Ω (o bien la condición análoga de que π tenga soporte compacto). La mayoría de los resultados se simplifican con ella y, de hecho, ha sido necesario introducirla en alguna ocasión. Puede decirse, "grosso modo", que la compacidad del espacio paramétrico es una cierta garantía de robustez.

4. ROBUSTEZ CUALITATIVA DE LAS REGIONES DE CONFIANZA Y TEST DE HIPOTESIS BAYESIANOS.

Después de analizar la robustez de los estimadores, parece natural estudiar la extensión de nuestro concepto de estabilidad a los otros dos problemas clásicos de inferencia: La estimación por regiones de confianza y el contraste de hipótesis.

Según señalábamos en I.2, el tratamiento de estos problemas por métodos frecuentistas encuentra serias dificultades cuando se trata de enunciar definiciones rigurosas de robustez que recojan las ideas intuitivas y, al mismo tiempo, resulten manejables.

Nuestro enfoque, aplicado a la metodología bayesiana, permite enunciar definiciones de robustez "naturales" para ambos problemas. Además, en el caso de los tests de hipótesis, el tratamiento matemático es sumamente sencillo si tomamos como punto de partida el concepto de estabilidad de un modelo analizado en el capítulo anterior. Naturalmente, esta mayor sencillez no puede considerarse como un argumento decisivo; en una situación concreta puede discutirse qué método ofrece una mayor riqueza de ideas, adaptación a la realidad y simplicidad de manejo. En gran parte, esta cuestión resulta inseparable de la polémica entre frecuentistas y bayesianos.

4.1. Tests de hipótesis

Con la notación usual, consideremos un modelo (π, V) y un contraste de hipótesis paramétrico planteado en la forma

$$H_o : \theta \in \Omega_o \subset R^k, \qquad H_1 : \theta \in \Omega_1 = \Omega \setminus \Omega_o$$

Siendo θ el parámetro de interés en la verosimilitud "teórica" $f_{\alpha_o}(.|\theta)$. Como es sabido, el tratamiento bayesiano de este problema (sin función de pérdida) consiste en calcular las probabilidades a posteriori

$$\pi_{\alpha_o}(H_o|x) = \int_{\Omega_o} \pi_{\alpha_o}(\theta|x)d\pi(\theta), \qquad \pi_{\alpha_o}(H_1|x) = \int_{\Omega_1} \pi_{\alpha_o}(\theta|x)d\pi(\theta)$$

que nos indican el "grado de creencia" a posteriori de cada una de las dos hipótesis.

Esto nos conduce de inmediato a la siguiente

<u>DEFINICION III.2</u>. Diremos que un test de hipótesis bayesiano es cualitativamente robusto respecto al modelo (π, V), en $f_{\alpha_o} \in V$ cuando, para toda sucesión $\{f_{\alpha_r}\}_{r\in\mathbb{N}}$ de verosimilitudes de V tal que

$$F_{\alpha_r}(.|\theta) \xrightarrow[r\to\infty]{d} F_{\alpha_o}(.|\theta) \qquad \forall\theta \in \Omega$$

se cumple

$$\pi_{\alpha_r}(H_o|x) \xrightarrow[r\to\infty]{} \pi_{\alpha_o}(H_o|x) \qquad \forall x \in \bigcap_{i=o}^{\infty} X_{\alpha_i}$$

siendo $X_\alpha = \{x \in \mathbb{R}^n \mid 0 < \int_\Omega f_\alpha(x|\theta)d\pi(\theta) < \infty\}$.

Siempre que la frontera de Ω_o, $\partial\Omega_o$, tenga probabilidad cero respecto a $\pi_{\alpha_o}(.|x)$, (lo cual sucederá, en particular, siempre que $\pi(\partial\Omega_o) = 0$), la estabilidad del modelo en f_{α_o} implica la robustez del test. Este hecho es consecuencia directa de la siguiente caracterización de la convergencia débil:

$F_r \xrightarrow[r\to\infty]{d} F$ si y sólo si para todo conjunto medible A tal que

$\int_{\partial A} dF = 0$ se verifica $\int_A dF_r \xrightarrow[r \to \infty]{} \int_A dF$.

Por tanto, puede decirse que la robustez cualitativa de los test Bayesianos está implícita en la estabilidad del modelo, a diferencia de lo que ocurre con la robustez de los estimadores que requiere, en general, condiciones adicionales.

4.2. Regiones de confianza

Dado un modelo (π, V), supongamos que se dispone de un criterio que nos permite determinar, para cada densidad a posteriori $\pi_\alpha(\cdot|x)$, un conjunto $S_{\alpha,x} \subset \Omega$ tal que $\int_{S_{\alpha,x}} \pi_\alpha(\theta|x) = d\pi(\theta) = 1-\beta$ siendo $\beta \in (0,1)$ una constante prefijada.

$S_{\alpha,x}$ es, por tanto, una región de confianza para θ asociada a $\pi_\alpha(\cdot|x)$ al nivel de probabilidad $1-\beta$.

Cuando se intenta adaptar el concepto de robustez cualitativa a la estimación por regiones de confianza, el primer problema que surge es definir adecuadamente la "proximidad" entre regiones de R^k. Una idea que parece plausible en este sentido, es considerar que dos regiones A y B están "próximas" cuando su diferencia simétrica $A \Delta B = (A \setminus B) \cup (B \setminus A)$ es un conjunto "pequeño".

DEFINICION III.3. Diremos que el método de estimación definido por las regiones de confianza $S_{\alpha,x}$ es cualitativamente robusto en $f_{\alpha_o} \in V$ cuando, para toda sucesión de verosimilitudes $\{f_{\alpha_r}\}_{r \in N} \subset V$ tal que $F_{\alpha_r}(\cdot|\theta) \xrightarrow[r \to \infty]{d} F_{\alpha_o}(\cdot|\theta)$ $\forall \theta \in \Omega$, se verifica

$$\pi(S_{\alpha_r,x} \,\Delta\, S_{\alpha_o,x}) \xrightarrow[r \to \infty]{} 0 \qquad \forall x \in \bigcap_{i=o}^{\infty} X_{\alpha_i}.$$

Discusión y comentarios.

(a) Obviamente, la robustez así definida depende del modelo (π, V) y del criterio adoptado para elegir las regiones $S_{\alpha,x}$. En inferencia bayesiana es frecuente utilizar regiones de confianza a las que se impone la condición de tener "máxima densidad a posteriori" (M.D.P.), es decir, que $S_{\alpha,x}$ se determina exigiendo que cumpla

$$\int_{S_{\alpha,x}} \pi_\alpha(\theta|x)\,d\pi(\theta) = 1-\beta \quad \text{y} \quad \pi_\alpha(\theta_1|x) \geq \pi_\alpha(\theta_2|x), \quad \forall\theta_1 \in S_{\alpha,x}, \quad \forall\theta_2 \notin S_{\alpha,x}$$

En general esta condición no es suficiente para determinar de forma única las $S_{\alpha,x}$, lo cual obliga a introducir criterios adicionales.

(b) Si definimos, para cualquier par A,B de subconjuntos de Borel de Ω, $d_\pi(A,B) = \pi(A \Delta B)$ se tiene que $d_\pi(A,B) = d_\pi(B,A)$ y también para todo conjunto C de Borel, $d_\pi(A,B) \leq d_\pi(A,C) + d_\pi(C,B)$ ya que $(A \Delta B) \subset (A \Delta C) \cup (B \Delta C)$.

Sin embargo, d_π no es una distancia pues es claro que no cumple en general $d(A,B) = 0 \iff A = B$. Este inconveniente puede obviarse identificando las regiones que difieran en conjuntos de π-medida nula. Formalmente ésto supone definir en la familia $\mathcal{B}$ de subconjuntos de Borel de Ω, la relación $A \sim B \iff \pi(A \Delta B) = 0$.

Es trivial comprobar que $\sim$ es relación de equivalencia. Por tanto, d_π induce en el conjunto cociente $\mathcal{B}/\sim$ una distancia, lo cual nos permite dar una interpretación matemática y estadística de la definición III.3 en términos de continuidad. La formalización de esta idea es inmediata y no se dará aquí.

El siguiente teorema prueba la robustez cualitativa en el caso de intervalos de confianza unidimensionales definidos mediante cuantiles (prescindiendo del criterio M.D.P.).

TEOREMA III.9. Sea un modelo (π,V) en el que $\Omega = \mathbb{R}$, π es μ_L-contínua y todas las distribuciones a posteriori son μ_L-contínuas.

Para todo $\beta \in (0,1)$ sea $\theta^{(\alpha)}_{\beta,x}$ el cuantil de orden β de la distribución $F^{(\alpha)}_{\pi}(.|x)$ (e.d. $F^{(\alpha)}_{\pi}(\theta^{(\alpha)}_{\beta,x}|x) = \beta$), y definamos como regiones de confianza con nivel $1-\beta_o$ a los intervalos $S_{\alpha,x} = [\theta^{(\alpha)}_{\beta_1,x}, \theta^{(\alpha)}_{1-\beta_2,x}]$ donde $\beta_1,\beta_2 \in (0,1)$ son constantes prefijadas tales que $\beta_o = \beta_1 + \beta_2$.

Si (π,V) es estable en f_{α_o} entonces el método de estimación así definido es cualitativamente robusto en f_{α_o}.

Demostración:

Sea $\{f_{\alpha_r}\}_{r\in\mathbb{N}} \subset V$ tal que $F_{\alpha_r}(.|\theta) \xrightarrow[r\to\infty]{d} F_{\alpha_o}(.|\theta)$ $\forall\theta \in \Omega$. Entonces por la estabilidad en f_{α_o}, se tiene

$$F^{(\alpha_r)}_{\pi}(.|x) \xrightarrow[r\to\infty]{d} F^{(\alpha_o)}_{\pi}(.|x) \quad \forall x \in \bigcap_{r=o}^{\infty} X_{\alpha_r}, \quad \text{siendo}$$

$$X_{\alpha} = \{x \in \mathbb{R}^n \mid 0 < \int_{\Omega} f_{\alpha}(x|\theta)d\pi(\theta) < \infty\}$$

Teniendo en cuenta la robustez de los cuantiles a posteriori (ver apartado III.3.3), esto implica:

$$\theta_{\beta_1,x}^{(\alpha_r)} \xrightarrow[r\to\infty]{} \theta_{\beta_1,x}^{(\alpha_o)} \qquad y \qquad \theta_{1-\beta_2,x}^{(\alpha_r)} \xrightarrow[r\to\infty]{} \theta_{1-\beta_2,x}^{(\alpha_o)}$$

Denotando por $I_1^{(r)}$ e $I_2^{(r)}$ los intervalos cerrados de extramos $\theta_{\beta_1,x}^{(\alpha_r)}$, $\theta_{\beta_1,x}^{(\alpha_o)}$ y $\theta_{1-\beta_2,x}^{(\alpha_r)}$, $\theta_{1-\beta_2,x}^{(\alpha_o)}$ respectivamente, y haciendo $T_r = S_{\alpha_r,x} \Delta S_{\alpha_o,x}$ se sigue que

$$\mu_L(T_r) = \mu_L(I_1^{(r)} \cup I_2^{(r)}) \xrightarrow[r\to\infty]{} 0$$

A partir de aquí demostraremos que $\pi(T_r) \xrightarrow[r\to\infty]{} 0$. En efecto, si no fuese así existirían $\varepsilon > 0$ y una subsucesión $\{T_{r_m}\}_{m\in\mathbb{N}}$ tales que

$$\mu_L(T_{r_m}) < \frac{1}{2^m} \qquad y \qquad \pi(T_{r_m}) > \varepsilon$$

Entonces, el conjunto $A = \lim\sup T_{r_m} = \bigcap_{p=1}^{\infty} B_p$ (siendo $B_p = \bigcup_{m=p}^{\infty} T_{r_m}$), verificaría

$$\mu_L(A) = \mu_L(\lim B_p) = \lim \mu_L(B_p) \leq \sum_{m=o}^{\infty} \mu_L(T_{r_m}) = \frac{1}{2^{p-1}} \xrightarrow[p\to\infty]{} 0$$

y además $\pi(A) = \pi(\lim B_p) = \lim \pi(B_p) \geq \varepsilon$.

Sin embargo, $\mu_L(A) = 0$ y $\pi(A) \geq \varepsilon$ contradice la hipótesis de ser π μ_L-contínua.

En definitiva, se concluye $\pi(T_r) \xrightarrow[r\to\infty]{} 0$ y, por tanto, la robustez cualitativa del método.

5. ESTABILIDAD CONJUNTA

El concepto de estabilidad propuesto y analizado en esta memoria sugiere la posibilidad de esbozar lo que podría llamarse la "teoría ge neralizada de la estabilidad en inferencia bayesiana" cuya principal línea de trabajo consistiría en estudiar la estabilidad (en el sentido de "continuidad") de los métodos de inferencia cuando todos los elementos del problema (distribución a priori, verosimilitud y función de pérdida) varían simultáneamente;la teoría desarrollada en el presente trabajo puede considerarse como un primer paso en esta dirección. Una segunda etapa sería analizar la "estabilidad conjunta", respecto a la verosimilitud y a la distribución a priori, del "operador Bayes". En otras palabras, se trataría de estudiar como afectan a la distribución a posteriori los "pequeños cambios" simultáneos en π y en f. Este es precisamente el objetivo de esta sección.

El concepto fundamental se establece en la siguiente

DEFINICION III.4. Consideremos el "modelo generalizado" (P,V) donde, con la notación usual, V es una familia de verosimilitudes y P es un conjunto de distribuciones a priori sobre Ω.

Diremos que (P,V) es estable en $(\pi, f_{\alpha_o}) \in P \times V$ cuando, para toda sucesión $S=\{(\pi_r, f_{\alpha_r})\}_{r\in\mathbb{N}} \subset P \times V$ tal que $\pi_r \xrightarrow[r\to\infty]{d} \pi$ y $F_{\alpha_r}(.|\theta) \xrightarrow[r\to\infty]{d} F_{\alpha_o}(.|\theta)$ $\forall\theta \in \Omega$, se verifica

$$F_{\pi_r}^{(\alpha_r)}(.|x) \xrightarrow[r\to\infty]{d} F_{\pi}^{(\alpha_o)}(.|x), \quad \forall x \in \bigcap_{r=o}^{\infty} \chi_r^s$$

siendo $\chi_r^s = \{x \in \mathbb{R}^n \mid 0 < \int_\Omega f_{\alpha_r}(x|\theta)d\pi_r(\theta) < \infty\}$.

El teorema que demostramos a continuación proporciona una condición suficiente para asegurar la estabilidad conjunta de un modelo generalizado.

<u>TEOREMA III.10</u>. Sea un modelo generalizado (P,V) donde P es un conjunto arbitrario de distribuciones sobre $\Omega \subset R^k$.

Supongamos que la familia V verifica

(i) Para toda sucesión $\{f_{\alpha_r}\}_{r\in N} \subset V$ tal que $F_{\alpha_r}(.|\theta) \xrightarrow[r\to\infty]{d} F_{\alpha_o}(.|\theta)$ $\forall\theta \in \Omega$ se cumple $f_{\alpha_r}(x|.) \xrightarrow[r\to\infty]{} f_{\alpha_o}(x|.)$ uniformemente en Ω $\forall x \in R^n$.

(ii) $f_{\alpha_o}(x|.)$ es función contínua y acotada en Ω $\forall x \in R^n$.

Entonces, (P,V) es estable en (π, f_{α_o}) $\forall \pi \in P$.

<u>Demostración</u>:

Sea $\{(\pi_r, f_{\alpha_r})\}_{r\in N} \subset V$ tal que $\pi_r \xrightarrow[r\to\infty]{d} \pi \in P$ y $F_{\alpha_r}(.|\theta) \xrightarrow[r\to\infty]{d} F_{\alpha_o}(.|\theta)$ $\forall\theta \in \Omega$.

Por la hipótesis (i) tenemos, para un $x \in R^n$ fijo:

$\forall\varepsilon > 0$, $\exists r_o = r_o(\varepsilon) \in N$ tal que $\forall r \geq r_o$,

$$|f_{\alpha_r}(x|\theta) - f_{\alpha_o}(x|\theta)| < \varepsilon$$

Esto implica

$$\int_\Omega (f_{\alpha_o}(x|\theta)-\varepsilon)d\pi_r(\theta) \leq \int_\Omega f_{\alpha_r}(x|\theta)d\pi_r(\theta) \leq \int_\Omega (f_{\alpha_o}(x|\theta)+\varepsilon)d\pi_r(\theta)$$

Pero $\int_\Omega f_{\alpha_o}(x|\theta)d\pi_r(\theta) \xrightarrow[r\to\infty]{} \int_\Omega f_{\alpha_o}(x|\theta)d\pi(\theta)$ ya que $f_{\alpha_o}(x|\theta)$ es función contínua y acotada de θ y $\pi_r \xrightarrow[r\to\infty]{d} \pi$. Así pues

$$\int_\Omega f_{\alpha_o}(x|\theta)d\pi(\theta)-\varepsilon \leq \liminf \int_\Omega f_{\alpha_r}(x|\theta)d\pi_r(\theta) \leq$$

$$\leq \limsup \int_\Omega f_{\alpha_r}(x|\theta)d\pi_r(\theta) \leq \int_\Omega f_{\alpha_o}(x|\theta)d\pi(\theta) + \varepsilon, \quad \forall\varepsilon > 0$$

Se sigue

$$\int_\Omega f_{\alpha_r}(x|\theta)d\pi_r(\theta) \xrightarrow[r \to \infty]{} \int_\Omega f_{\alpha_o}(x|\theta)d\pi(\theta) \qquad (1)$$

Por otra parte, para toda función $u : \Omega \longrightarrow R$ contínua y acotada tenemos también

$$\int_\Omega f_{\alpha_r}(x|\theta)u(\theta)d\pi_r(\theta) \xrightarrow[r \to \infty]{} \int_\Omega f_{\alpha_o}(x|\theta)u(\theta)d\pi(\theta) \qquad (2)$$

El razonamiento es idéntico al utilizado para demostrar (1) ya que $g_{\alpha_r}(x|.) = f_{\alpha_r}(x|.)u(.) \xrightarrow[r \to \infty]{} g_{\alpha_o}(x|.) = f_{\alpha_o}(x|\theta)u(\theta)$ uniformemente en Ω (por ser u acotada) y $g_{\alpha_o}(x|.)$ es también función contínua y acotada en Ω.

En definitiva, de (1) y (2) resulta que para toda función $u : \Omega \longrightarrow R$ contínua y acotada:

$$\int_\Omega u(\theta)dF_{\pi_r}^{(\alpha_r)}(.|x) = \left[\int_\Omega f_{\alpha_r}(x|\theta)d\pi_r(\theta)\right]^{-1} \int_\Omega u(\theta)f_{\alpha_r}(x|\theta)d\pi_r(\theta) \rightarrow$$

$$\xrightarrow[r \to \infty]{} \left[\int_\Omega f_{\alpha_o}(x|\theta)d\pi(\theta)\right]^{-1} \int_\Omega u(\theta)f_{\alpha_o}(x|\theta)d\pi(\theta) =$$

$$= \int_\Omega u(\theta)dF_{\pi_o}^{(\alpha_o)}(.|x)$$

$\forall x \in R^n \mid 0 < \int_\Omega f_{\alpha_o}(x|\theta)d\pi(\theta) < \infty$ y $0 < \int_\Omega f_{\alpha_r}(x|\theta)d\pi_r(\theta) < \infty$,

$\forall r \in N$

quedando probada la estabilidad en (π, f_{α_o}).

<u>Discusión y comentarios</u>

(a) Como puede observarse, la clave de la demostración está en la convergencia uniforme de las densidades. El teorema II.6 da una condición suficiente, bastante natural y amplia, para que se cumpla esta hipótesis.

(b) Se pueden obtener otros teoremas análogos en los que se utilice un esquema de demostración distinto (el teorema de Scheffé en lugar de la definición de convergencia débil). Para ello, sería necesario imponer hipótesis más fuertes sobre P con objeto de asegurar la convergencia puntual de las densidades a priori aunque, paralelamente, podrían rebajarse las restricciones sobre V.

REFERENCIAS

ANDREWS, D.F., BICKEL, P.J., HAMPEL, F.R., HUBER,P.J., ROGERS, W.H., and TUKEY, J.W. (1972). "Robust estimates of location. Survey and advances". Princeton University Press, Princeton New Jersey.

BILLINGSLEY, P. (1971). "Weak convergence of measures: Applications in Probability". Regional Conference Series in Applied Mathematics n° 5. SIAM. Philadelphia.

BOX, G.E.P. and TIAO, G.C. (1962). "A further look at robustness via Baye's theorem". Biometrika 49, pp. 419-432.

BOX, G.E.P. and TIAO, G.C. (1964). "A Bayesian approach to the importance of assumptions applied to the comparison of variances". Biometrika 51, pp. 153-167.

BOX, G.E.P. and TIAO, G.C. (1968). "A Bayesian approach to some outlier problems". Biometrika 55, pp. 119-129.

BOX, G.E.P. and TIAO, G.C. (1973). "Bayesian Inference in Statistical Analysis". Addison-Wesley, Reading Massachusetts.

DE GROOT, M.F. (1970). "Optimal Statistical Decisions". Mc.Graw Hill. New York.

DE ROBERTIS, L. and HARTIGAN, J.A. (1979). "Ranges of piror measures" Tech. Rep. Yale University.

DENIAU, C., OPPENHEIM, G. et VIANO, C. (1977). "Robustesse: π-robustesse et minimax-robustesse". Astérisque 43-44, pp. 151-166. Societé Mathématique de France.

DICKEY, J.M. (1976). "Approximate posterior distributions". J. Amer. Statist. Ass., 71, 680-689.

FELLER, W. (1971). "An introduction to probability theory and its applications". John Wiley & Sons. New York.

GIRON, F.J. (1981). Comunicación personal sobre el tema "Mixturas infinitas de distribuciones".

HAMPEL, F.R. (1968). "Contributions to the theory of robust estimation". Ph. D. Thesis. University of California. Berkeley.

HAMPEL, F.R. (1971). "A general qualitative definition of robustness" Ann. Math. Statist. 42, pp. 1887-1896.

HAMPEL, F.R. (1973). "Robust estimation: A condensed partial survey" Z. Wahrscheinlichkeitstheorie Kerw. Gebiete, 27, pp. 87-104.

HAMPEL, F.R. (1974). "The influence curve and its role in robust estimation". J. Amer. Statist. Ass., 62, pp. 1179-1186.

HOGG, R.V. (1979). "An introduction to robust estimation" in "Robutness in Statistics", Launer, R.L. an Wilkinson G.N., Ed. Academic Press. New York.

HUBER, P.J. (1964). "Robust estimation of a location parameter". Am. Math. Statist., 35, pp. 73-101.

HUBER, P.J. (1972). "Robust Statistics: A Review". Ann. Math. Statist. 43, pp. 1041-1067.

HUBERT, P.J. (1977). "Robust Statistical Procedures". Regional Conference Series in Applied Mathematics n° 27. SIAM. Philadelphia.

HUBER, P.J. (1981). "Robust Statistics". John Wiley & Sons. New York.

KADANE, J.B. and CHUANG, D.T. (1978). "Stable decision problems". Ann. of Statist., 6, n° 5, pp. 1095-1110.

KOLMOGOROV, A.N. y FOMIN, S.V. (1975). "Elementos de la Teoría de Funciones y del Análisis Funcional". Mir. Moscú.

LINDLEY, D.V. (1971). "Bayesian Statistics: A Review". Regional Conference Series in Applied Mathematics. SIAM. Philadelphia.

RAO, C.R. (1973). "Linear Statistical Inference and its applicatións" John Wiley & Sons. New York.

RIOS, M.J. (1981) "Problemas de decisión con información parcial a priori". Tesis Doctoral. Universidad Complutense. Madrid.

SCHEFFE, H. (1947). "A useful convergence theorem for Probability Theory". Ann. Math. Statist. 18, 434-438 .

SERFLING, R.J. (1980). "Approximation Theorems of Mathematical Statistics". John Wiley & Sons. New York.

STRASSEN, V. (1965). "The existence of probability measures with given marginals". Ann. Math. Statist., 36, pp. 423-439.

ZELLNER, A. (1971). "An introduction to Bayesian inference in econometrics". John Wiley & Sons. New York.