

**UNIVERSIDAD DE SEVILLA**

**Departamento de Economía Aplicada I**

OBTENCIÓN DE DISTRIBUCIONES A PRIORI NO INFORMATIVAS USANDO MEDIDAS DE INFORMACIÓN. APLICACIÓN A LA EVALUACIÓN DE LAS REVISTAS CIENTÍFICAS.

Trabajo presentado por Fco. Javier Ortega Irizo para la obtención del título de doctor, dirigido por el profesor Dr. Jesús Basulto Santos.

Sevilla, Marzo de 2001.

*A mi hija María.*

## **AGRADECIMIENTOS.**

Quisiera expresar mi más sincero agradecimiento al Dr. D. Jesús Basulto, director del presente trabajo, por todo el esfuerzo y tiempo dedicado y por sus consejos, que en todo momento, han actuado de guía en la elaboración del mismo.

Asimismo, expresar también mi agradecimiento a los compañeros del Departamento de Economía Aplicada I que, de una u otra forma, me han prestado su apoyo.



---

## ÍNDICE.

•	<b>INTRODUCCIÓN.</b>	<b>1</b>
	<b>CAPÍTULO 1: MEDIDAS DE INFORMACIÓN.</b>	
	<b>DEFINICIÓN DE MODELO REGULAR.</b>	<b>11</b>
1.1	Definición clásica de modelo regular.	11
1.2	Información de Fisher. Propiedades.	15
1.3	Medidas de información generales.	21
	1.3.1. Propiedades de la medida de Información general.	28
	1.3.2. Relación entre las medidas de información generales y la Información de Fisher.	31
1.4	Definición alternativa de modelo regular.	37
1.5.	La Inferencia Condicional.	41
	<b>CAPÍTULO 2: DISTRIBUCIONES A PRIORI EN MODELOS REGULARES: LA REGLA DE JEFFREYS.</b>	<b>57</b>
2.1	Invarianza ante reparametrizaciones.	58
2.2	La regla de Jeffreys. Caso uniparamétrico.	59
	2.2.1 Interpretación geométrica de la regla de Jeffreys.	67
	2.2.2 Coincidencia de resultados con la Inferencia Fiducial y Clásica.	77
2.3	La regla de Jeffreys. Caso multiparamétrico.	81
	2.3.1 La regla general de Jeffreys y la regla modificada.	81

2.3.2	Coincidencia de intervalos bayesianos y clásicos para más de un parámetro.	97
<b>CAPÍTULO 3:</b>	<b>MODELOS NO REGULARES. GENERALIZACIÓN DE LA REGLA DE JEFFREYS.</b>	<b>107</b>
3.1	Elección de la distribución a priori. Caso uniparamétrico.	108
3.1.1	Invarianza ante reparametrizaciones.	115
3.1.2	Propiedades frecuentistas de los intervalos bayesianos.	124
3.1.3	Subfamilias uniparamétricas de la familia $U(\alpha, \beta)$ .	125
3.2	Elección de la distribución a priori. Caso multiparamétrico.	134
3.2.1	Propiedades frecuentistas de los intervalos bayesianos.	144
<b>CAPÍTULO 4:</b>	<b>ANÁLISIS DE ALGUNOS MODELOS DE DURACIÓN.</b>	<b>161</b>
4.1	Algunos conceptos básicos de los modelos de duración.	161
4.1.1	La función de supervivencia y la función de riesgo.	163
4.1.2	Principales tipos de censura.	165
4.1.3	Verosimilitudes asociadas a datos con censura.	167
4.2	El modelo Exponencial.	171
4.2.1	El modelo Exponencial sin censura.	171
4.2.2	El modelo Exponencial con censura tipo I.	172
4.2.3	El modelo exponencial con parámetro umbral y censura tipo I.	175
4.3	El modelo Weibull.	195
4.3.1	El modelo Weibull sin censura.	196
4.3.2	El modelo Weibull con censura tipo I.	198

<b>CAPÍTULO 5:</b>	<b>APLICACIÓN A LA EVALUACIÓN DE LAS REVISTAS CIENTÍFICAS</b>	<b>203</b>
5.1	El uso de las citas para la evaluación de las revistas científicas.	204
5.2	El Factor de Impacto y el proceso de obsolescencia.	207
5.3	Datos utilizados y procedimiento seguido.	211
5.4	Resultados y conclusiones.	217
5.4.1	Tres temáticas distintas.	217
5.4.2	Doce revistas del ámbito de Economía Aplicada.	218
•	<b>CUESTIONES ABIERTAS Y CONCLUSIONES GENERALES.</b>	<b>227</b>
•	<b>APÉNDICE A:</b> Datos y gráficas de antigüedad de las citas.	<b>233</b>
•	<b>APÉNDICE B:</b> Listados de los programas con Mathematica 3.0.	<b>249</b>
•	<b>BIBLIOGRAFÍA.</b>	<b>263</b>
1.	Bibliografía global.	263
2.	Bibliografía por capítulos.	270





## INTRODUCCIÓN.

Una de las objeciones principales que históricamente se ha puesto al enfoque bayesiano de la inferencia paramétrica es la dificultad existente a la hora de elegir una distribución a priori para el (los) parámetro(s) desconocido(s).

Fisher, R.A. (1930) indica que se incurre en un error importante al pensar que nuestras creencias racionales acerca de un parámetro pueden expresarse siempre en términos de una distribución de probabilidad (Pedersen, J.G. (1978)), lo que le llevó a buscar procedimientos alternativos de inferencia que pudiesen ser aplicados cuando no dispusiésemos de una distribución a priori, construyendo así el argumento fiducial (Fisher, R.A. (1922,1930,1959)).

No obstante, y quizás paradójicamente, estos trabajos pueden ser considerados como el origen “contemporáneo” del problema de elegir una distribución a priori cuando no disponemos de una opinión experta subjetiva para obtenerla, es decir, del problema de elegir lo que en un principio se llamó distribución a priori no informativa (posteriormente, se han introducido otros términos como referencia a priori (Kass, R.E. and Washerman, L. (1995,b)), Berger, J. and Bernardo, J.M. (1992)), distribución a priori imparcial, (Basulto, J. (1997)), etc.) Savage, L.J. (1961), dice que la Inferencia Fiducial es “*un intento de hacer la tortilla bayesiana sin romper el huevo bayesiano*”<sup>1</sup>. Podemos encontrar también en Box, G.E. and Tiao, G.C. (1973), las siguientes palabras: “*Aunque este enfoque no usa probabilidades a priori, Fisher deja claro que la inferencia fiducial intenta cubrir la situación en la que no se conoce nada a priori sobre los parámetros, y las soluciones que proporciona este método son muy parecidas a las obtenidas a partir del teorema de Bayes con distribuciones a*

---

<sup>1</sup> Savage, L.J. (1961). El texto original es: “*...an attempt to make the Bayesian omelette without braking the Bayesian eggs*”. La cita puede encontrarse en Box, G.E. and Tiao, G.P. (1973), pág. 73.

*priori no informativas*”<sup>2</sup>. Una visión global y bastante completa de los trabajos de Fisher sobre inferencia fiducial puede verse en Zabell, S.L. (1992).

Uno de los objetivos fundamentales del presente trabajo es abordar el problema de la elección de una distribución a priori no informativa que sea adecuada para poder realizar un análisis bayesiano paramétrico, una vez elegidos los modelos que describen el comportamiento de los datos. En este aspecto, el trabajo es una continuación de un trabajo de investigación anterior titulado “*La regla de Jeffreys: un estudio de su significado y propiedades*” bajo la dirección del Dr. Basulto, J. Históricamente, el problema surge con la propia inferencia bayesiana, pues el propio Bayes en 1763 ya lo trata. La solución aportada es la del uso de una distribución a priori uniforme. Esta idea fue defendida posteriormente por varios autores entre los que cabe destacar a Laplace. Hay autores, como Stigler, S.M. (1986), que defienden que Bayes no propuso como solución general el uso de la distribución uniforme, sino que ante un problema concreto (un modelo binomial de probabilidad de éxito  $\theta$  desconocida) y bajo una premisa concreta (obtener una distribución predictiva uniforme en el conjunto  $\{0, 1, \dots, n\}$ ), llegó a la conclusión de utilizar una distribución a priori uniforme en este caso.

Como es conocido, la propuesta de usar una distribución uniforme puede resultar incoherente dado que no posee la propiedad de invarianza ante reparametrizaciones (ver el capítulo 2). Así, por ejemplo, en un modelo Normal donde la media es conocida, ¿hemos de elegir una distribución uniforme para la desviación típica o para la varianza del modelo?. Evidentemente, ambas distribuciones no pueden ser uniformes, pues siendo  $X$  una variable aleatoria positiva con distribución uniforme, ni  $X^2$  ni  $\sqrt{X}$  tienen distribuciones planas.

---

<sup>2</sup> Box, G.E. and Tiao, G.P. (1973), pág. 73. El texto original es: “*Although this approach does not employ prior probability, Fisher made it clear that fiducial inference was intended to cover the situation where nothing was known about the parameters a priori, and the solutions which are accessible to this method closely parallel those obtained from Bayes theorem with noninformative prior distributions.*”

---

El argumento de invarianza condujo a Jeffreys, H. (1961) a la propuesta que finalmente ha sido más aceptada (al menos en el caso uniparamétrico) de tomar la distribución a priori proporcional a la raíz cuadrada de la Información de Fisher del modelo, que es la conocida Regla de Jeffreys, aplicable sólo a los llamados modelos regulares.

Desde el trabajo de Welch, B.L. and Peers, H.W. (1963), se ha generalizado la búsqueda de distribuciones a priori no informativas con buenas propiedades “desde el punto de vista frecuentista”, es decir, distribuciones que conduzcan a intervalos bayesianos de probabilidad  $1-\alpha$  cuyo nivel de confianza desde el punto de vista frecuentista sea también  $1-\alpha$  (o al menos se tenga esta propiedad para grandes muestras). Esta propiedad que en un principio indicaría que la distribución a priori usada es no informativa, a la vez sirve también para reforzar los resultados de la inferencia clásica dada la difícil interpretación teórica que tienen los intervalos de confianza con ella obtenidos.

Podríamos preguntarnos por qué en una situación en la que no tenemos una opinión subjetiva a priori sobre el parámetro, nos planteamos buscar una distribución a priori en vez de usar directamente el enfoque clásico de la inferencia. Esta cuestión entraría de lleno en la tan debatida polémica de la idoneidad de utilizar el enfoque clásico o el enfoque bayesiano. Hemos de aclarar que no entra en los objetivos del presente trabajo dar respuesta a la anterior cuestión. De hecho, pensamos que lo ideal sería abandonar el tradicional enfoque de enfrentamiento entre ambas perspectivas de la inferencia y sustituirlo por otro de complementariedad. Como ha quedado claro a lo largo de la historia de la estadística, ambos enfoques tienen sus virtudes y sus inconvenientes y por ello creemos que cualquier estadístico aplicado, independientemente de que pueda mostrar sus preferencias por una u otra forma de trabajar, ha de tener unos mínimos conocimientos de ambas técnicas que le permitan poder usarlas según crea más conveniente en función del problema que se pretenda resolver.

Desde el punto de vista teórico, es de sobra conocida la diferencia fundamental entre los enfoques bayesiano y clásico de la inferencia paramétrica. En las técnicas clásicas, el parámetro se considera una cantidad desconocida pero fija. De esta forma, la inferencia se realiza basándonos en la distribución de estadísticos en el muestreo, es decir, teniendo en cuenta todas las posibles muestras que podríamos obtener (y las probabilidades de cada una de ellas) obtenemos probabilidades relacionadas con el parámetro. Estas probabilidades se entienden desde el punto de vista frecuentista, es decir, un intervalo de confianza para un parámetro  $\theta$  al 95% significaría que si repetimos un número grande de veces el experimento, el intervalo obtenido en cada uno de ellos contendría al verdadero valor del parámetro en un 95% de los casos. Sin embargo, si analizamos cómo se llevan a cabo los experimentos para hacer inferencia, en realidad vemos que obtenemos una única muestra y con ella obtenemos todo lo posible acerca del parámetro. Parece quizás más acertado buscar técnicas en las que se tenga en cuenta una única muestra: la que se obtiene al realizar el experimento. Por ello, el enfoque bayesiano consiste en considerar que el parámetro va a ser, no una cantidad fija, sino una variable aleatoria, que como tal tendrá una distribución de probabilidad que es la llamada distribución a priori. Esta distribución es escogida subjetivamente por el experimentador sobre la base de sus conocimientos a priori y la obtención de una muestra hará que sus creencias cambien, obteniendo la distribución a posteriori para el parámetro como una combinación de la distribución a priori y la información obtenida a partir de la muestra. De esta forma, las inferencias que se realicen posteriormente se basarán en la distribución a posteriori del parámetro, de forma que, por ejemplo, un intervalo al 95% será un intervalo que contenga un 95% de la masa de probabilidad de la distribución a posteriori de  $\theta$ .

Otra propiedad teórica interesante del enfoque bayesiano es la facilidad con que puede generalizarse de uno a varios parámetros, ya que basta obtener la distribución a posteriori conjunta de los parámetros para, teóricamente, poder estimar cualquier característica que deseemos conocer. Además, el desarrollo de las técnicas de simulación (especialmente del muestreo de Gibbs (Geman, S. And Geman, D.

(1984), Gelfand, A.E. and Smith, A.F. (1990), etc.) ha hecho posible el obtener resultados prácticos a partir de densidades teóricamente intratables.

Queremos también destacar que el concepto de distribución a priori no informativa ha ido evolucionando con el paso del tiempo. Inicialmente, se planteó la búsqueda de distribuciones a priori a través de procedimientos que representaran formalmente un estado de ignorancia. Sin embargo, se ha ido produciendo un giro hacia el enfoque de disponer de distribuciones con “buenas propiedades” y que puedan ser utilizadas “por defecto” en el caso de que no dispongamos de una distribución subjetiva. En Kass, R.E. and Wasserman, L. (1995,b), podemos leer: *“Para simplificar, parece útil identificar dos interpretaciones de las referencias a priori. La primera dice que las referencias a priori son representaciones formales de ignorancia, la segunda dice que no podemos decidirnos objetivamente por una única distribución a priori que represente ignorancia. Así, las referencias a priori son elegidas por “acuerdo público”, tales como los de los pesos y medidas. En esta interpretación, las referencias a priori serían algo así como la opción por defecto en un paquete estadístico para ordenadores. (...). La primera interpretación era, en un principio, la interpretación dominante (...). En gran medida, se está produciendo un giro hacia la segunda interpretación.”*<sup>3</sup>

De esta forma, el objetivo primordial de la primera parte metodológica del trabajo (que comprende los capítulos 1, 2 y 3), es que, una vez que decidamos utilizar el enfoque bayesiano para resolver nuestro problema (por motivos teóricos, prácticos o de ambos tipos), la elección de la distribución a priori no sea un obstáculo si no disponemos de la información de un experto que nos permita elegir una distribución subjetiva. Este problema, ha sido básicamente resuelto en los modelos regulares, aunque presentando mayores inconvenientes y desacuerdos en el caso

---

<sup>3</sup> Kass, R.E. and Wassermann, L. (1995,b), pág. 33. El texto original es: *“At the risk of oversimplification, it seems useful to identify two interpretations of reference priors. The first assert that reference priors are formal representations of ignorance. Instead, reference priors are chosen by public agreement, much like units of length and weight. In this interpretation, reference priors are akin to a default option in a computer package. (...). The first interpretation was, at one time, the dominant interpretation (...). For the most part, the mood has shifted towards the second interpretation.”*

multiparamétrico. No obstante, el problema permanece abierto cuando los modelos se apartan de las condiciones de regularidad habituales, dada la imposibilidad de aplicar la regla de Jeffreys, siendo algunos de estos modelos de interés tanto teórico como práctico (modelos uniformes, modelos truncados en un valor desconocido, modelos con parámetro de garantía (ver capítulo 4), etc.).

En el capítulo 1 exponemos una serie de conceptos y procedimientos generales que aparecerán a lo largo del resto del trabajo por su relación con el problema de elegir una distribución a priori no informativa. Así, trataremos los conceptos y propiedades de la Información de Fisher (aplicable a los modelos regulares) y de otra medida de información aplicable a todos los modelos (que hemos denominado medida de información general). La relación entre ambas medidas nos llevará también a adoptar una definición de modelo regular distinta a la usada habitualmente. El concepto de información va a jugar un papel clave en la obtención de las distribuciones a priori. También haremos una exposición del principio de condicionar a los estadísticos auxiliares, ya que, cuando en un modelo el estadístico máximo-verosímil no es suficiente y existe un estadístico auxiliar, según el anterior principio de Fisher, hemos de hacer nuestra inferencia condicionando al estadístico auxiliar, obteniendo resultados distintos a los obtenidos con la inferencia no condicional. En multitud de ejemplos pondremos de manifiesto en el trabajo que con las distribuciones a priori que proponemos los intervalos bayesianos coinciden numéricamente con los intervalos obtenidos a través de la inferencia condicional.

En el capítulo 2 hacemos una recapitulación de los resultados que hemos considerado más importantes concernientes a la obtención de una distribución a priori no informativa en el caso de modelos regulares, que es la situación más habitual en la literatura. Este fue el tema de un trabajo de investigación anterior titulado “La Regla de Jeffreys: un Estudio de su Significado y Propiedades” y en este sentido este capítulo constituye un resumen de los resultados en él presentados con un interés especial en mostrar los métodos de obtención de distribuciones a priori que han tenido una mayor aceptación. En el caso de modelos regulares con un único parámetro

desconocido existe acuerdo total en elegir la distribución a priori no informativa que se obtiene a través de la Regla de Jeffreys, dadas las buenas propiedades que presenta esta propuesta con respecto a distintos criterios, sobre todo por sus propiedades de invarianza y por las propiedades frecuentistas de los intervalos bayesianos que se obtienen. Cuando el modelo tiene más de un parámetro desconocido, la generalización directa de la regla mantiene la invarianza ante reparametrizaciones pero muestra serias deficiencias, lo que ha llevado a que sea más utilizada la llamada Regla de Jeffreys modificada, aplicable a modelos de localización-escala (y a modelos ortogonales).

En el capítulo 3, intentamos generalizar las ideas vistas con respecto a la elección de distribuciones a priori en modelos regulares, proporcionando un criterio de obtención de tales distribuciones que es aplicable también a modelos no regulares y que conduce en el caso regular a la misma distribución que se obtiene al aplicar la Regla de Jeffreys y la Regla de Jeffreys modificada en los casos de uno o más parámetros respectivamente. La base fundamental de la propuesta que haremos reside en los comentarios hechos por el propio Jeffreys, H. (1946). Partiendo de una medida de información basada en la discrepancia o distancia entre las funciones de densidad correspondientes a dos valores distintos del parámetro, Jeffreys, H. (1946) observa que, bajo condiciones de regularidad, para cada valor  $\theta$  del espacio paramétrico, la información o discrepancia entre las densidades correspondientes a  $\theta$  y a  $\theta + \Delta\theta$  es del orden de  $(\Delta\theta)^2$ , con constante de proporcionalidad igual a la información de Fisher del modelo en  $\theta$ , que es la función base para definir la distribución a priori no informativa. También señala que esto no siempre es correcto, poniendo como ejemplo el modelo uniforme con recorrido en  $[0, \theta]$ . Pondremos de manifiesto que tanto en este modelo uniforme como en todos los modelos no regulares usuales ocurre que la información correspondiente a  $\theta$  y a  $\theta + \Delta\theta$  es del orden de  $|\Delta\theta|$  y que la función del parámetro que actúa como constante de proporcionalidad es la que debe darnos la clave para definir la distribución a priori en el caso no regular. En el caso multiparamétrico, buscaremos reparametrizaciones ortogonales para buscar las distribuciones a priori por separado.

---

La segunda parte del trabajo (que corresponde a los capítulos 4 y 5) es una aplicación práctica de los métodos bayesianos usando distribuciones a priori no informativas a datos bibliométricos utilizando modelos de duración. Esta aplicación surge dentro del ámbito del grupo de investigación dirigido por el Dr. Basulto, J. que viene desarrollando su trabajo en el campo de la bibliometría desde su creación.

En el capítulo 4 pretendemos ilustrar cómo resolver de forma práctica distintos modelos al usar la metodología descrita en los capítulos anteriores, presentando los distintos problemas que surgen y abordando sus soluciones por medio de las herramientas disponibles, todo ello desde una perspectiva eminentemente práctica. Nos centraremos en los modelos de duración, de gran utilidad práctica en diversos campos de la ciencia, pero a la vez bastante complicados de estudiar debido a que la presencia de censura es bastante habitual. El uso de métodos de simulación tipo Montecarlo junto con el algoritmo de Gibbs (que permite simular muestras de distribuciones multidimensionales a partir de las densidades condicionadas unidimensionales) hace que dispongamos de procedimientos bastante mecánicos para resolver modelos de gran complejidad teórica. Hemos analizado los modelos Exponencial, Exponencial con parámetro umbral y Weibull, ofreciendo en el apéndice B los listados de los programas utilizados para su resolución.

En este capítulo 5 vamos a aplicar la metodología descrita con anterioridad al campo de la Bibliometría, centrándonos en el estudio de la distribución de la antigüedad de las citas en la literatura científica. El indicador más comúnmente utilizado para evaluar y clasificar a las revistas científicas es el Factor de Impacto, que para cada año sólo tiene en cuenta las citas hechas a artículos publicados en los dos años anteriores. En general, existe tendencia a considerar sólo citas a artículos recientes. Esto se pone de manifiesto, por ejemplo, en el hecho de que en las bases de datos del ISI (“Institute Scientific Information”) las citas con 10 o más años de antigüedad aparecen agregadas. En nuestra aplicación, uno de los objetivos es efectuar una evaluación o comparación entre las revistas que tenga en cuenta “toda la historia” en cuanto a las citas recibidas y no sólo “la historia reciente”. Para ello hemos buscado



modelos probabilísticos adecuados para describir la variable que nos proporciona la antigüedad de las citas recibidas por una determinada revista científica. Si utilizamos los datos que proporciona directamente el ISI, necesitamos introducir modelos censurados. Dado el alto porcentaje de censura que presentan las distribuciones, hemos elaborado estrategias de búsqueda para poder disponer de la distribución completa correspondiente a las revistas que hemos analizado. Estos últimos datos proceden de un número menor de revistas fuente pero que permiten utilizar modelos sin censura.

Finalmente, tras indicar algunas conclusiones generales y problemas abiertos que constituyen las posibles líneas de investigación futuras, en sendos apéndices ofrecemos los datos que hemos utilizado para nuestra aplicación práctica (junto con algunos gráficos) así como los listados de todos los programas que hemos confeccionado para poder llevar a la práctica los procesos de simulación que permiten resolver los modelos planteados.



**CAPÍTULO 1: MEDIDAS DE INFORMACIÓN.**  
**DEFINICIÓN DE MODELO REGULAR.**

En este capítulo vamos a exponer una serie de conceptos y procedimientos generales que aparecerán a lo largo del resto del trabajo por su relación con el problema de elegir una distribución a priori no informativa. Así, trataremos los conceptos y propiedades de la Información de Fisher (aplicable a los modelos regulares) y de otra medida de información aplicable a todos los modelos (que hemos denominado medida de información general). La relación entre ambas medidas nos llevará también a adoptar una definición de modelo regular distinta a la usada habitualmente. El concepto de información va a jugar un papel clave en la obtención de las distribuciones a priori. También haremos una exposición del principio de condicionar a los estadísticos auxiliares, ya que, cuando en un modelo, el estadístico máximo-verosímil no es suficiente y existe un estadístico auxiliar, según el anterior principio de Fisher, hemos de hacer nuestra inferencia condicionando al estadístico auxiliar, obteniendo resultados distintos a los obtenidos con la inferencia no condicional. En multitud de ejemplos pondremos de manifiesto en el trabajo que con las distribuciones a priori que proponemos los intervalos bayesianos coinciden numéricamente con los intervalos obtenidos a través de la inferencia condicional.

**1.1 Definición clásica de modelo regular.**

Podemos decir que el objetivo fundamental de la estadística es extraer la **información relevante** que hay contenida en un conjunto de datos acerca de alguna característica desconocida. A la vista de ello, es fundamental disponer de técnicas de **reducción de la dimensionalidad**, es decir, métodos que permitan disponer de la misma **cantidad** de información relevante pero con datos encuadrados en un espacio

de dimensión más pequeña (o en definitiva, trabajando con **menos datos**). En este sentido, juegan un papel fundamental los estadísticos, que son funciones del espacio muestral (de dimensión igual al número de observaciones  $n$ ) en otros espacios que pueden ser, y generalmente son, de menor dimensión; dentro de ellos, son de especial importancia los estadísticos suficientes, que, en esencia, son aquellos que contienen la misma información relevante que la muestra de tamaño  $n$ .

Se hace necesario así introducir medidas que cuantifiquen la cantidad de información relevante que obtenemos sobre una característica o parámetro desconocido cuando trabajamos con un estadístico, ya que no sólo es importante el aspecto cualitativo de la suficiencia frente a la insuficiencia, sino que ante modelos en los que nos vemos obligados a trabajar con estadísticos no suficientes necesitamos cuantificar “la cantidad de insuficiencia” ante la que nos encontramos y cómo varía ésta a medida que aumentamos el tamaño muestral, puesto que si en el límite dicha cantidad es cero, sabremos que para grandes muestras estamos trabajando con un estadístico “prácticamente suficiente”.

Gran parte de la literatura estadística se ha dedicado al estudio de los llamados “modelos de probabilidad regulares”, que cumplen una serie de condiciones que los hacen más manejables. En ellos, la medida de información más conocida es la llamada Información de Fisher, que no tiene sentido como tal en modelos no regulares.

En un principio, vamos a definir los modelos regulares y la Información de Fisher en la forma en que es habitual en los libros básicos de inferencia. Posteriormente, vamos a ver cómo, a partir de medidas de información generales (aplicables a modelos regulares y no regulares), puede obtenerse la Información de Fisher y una definición alternativa de modelo regular menos restrictiva que la usual.

En la definición de modelo regular, las condiciones pueden variar ligeramente de unos autores a otros. Vamos a recoger la definición que aparece en Azzalini, A. (1996).

**Definición:** Dada una familia de distribuciones  $P_\theta$ ,  $\theta \in \Theta \subseteq \mathbb{R}^k$  con funciones de densidad  $f(x, \theta) \equiv f_\theta(x)$ , decimos que constituyen un modelo regular si se verifican las siguientes condiciones:

- i) El modelo es identificable, en el sentido de que  $\forall \theta_1 \neq \theta_2 \in \Theta$ , ha de existir al menos un conjunto B del espacio muestral tal que  $P_{\theta_1}(B) \neq P_{\theta_2}(B)$ .
- ii) El espacio paramétrico  $\Theta$  es un intervalo abierto de  $\mathbb{R}^k$ .
- iii) Todas las funciones de densidad especificadas por el modelo tienen el mismo soporte.
- iv) Para la función f, la derivación con respecto a  $\theta$  y la integración con respecto a x pueden intercambiarse hasta orden dos. Concretamente:

$$\text{a) } \int \frac{\partial}{\partial \theta} f(x, \theta) dx = \frac{\partial}{\partial \theta} \int f(x, \theta) dx$$

$$\text{b) } \int \frac{\partial^2}{\partial \theta^2} f(x, \theta) dx = \frac{\partial^2}{\partial \theta^2} \int f(x, \theta) dx$$

donde debe entenderse que si  $k > 1$ , entonces  $\frac{\partial^2}{\partial \theta^2}$  representa  $\frac{\partial^2}{\partial \theta^i \partial \theta^i}$ .

Las condiciones exigidas en la definición llevan a que se cumplan las siguientes propiedades:

$$1. E_\theta \left[ \frac{\partial \log f(x, \theta)}{\partial \theta} \right] = 0 \quad \forall \theta$$

$$2. E_\theta \left[ \left( \frac{\partial \log f(x, \theta)}{\partial \theta} \right)^2 \right] = -E_\theta \left[ \frac{\partial^2 \log f(x, \theta)}{\partial \theta^2} \right] \quad \forall \theta$$

**Dmt:**

Propiedad 1:

$$\begin{aligned} E_0 \left[ \frac{\partial \log f(x, \theta)}{\partial \theta} \right] &= \int \left( \frac{\partial \log f(x, \theta)}{\partial \theta} \right) f(x, \theta) dx = \int \frac{\frac{\partial f(x, \theta)}{\partial \theta}}{f(x, \theta)} f(x, \theta) dx = \\ &= \int \frac{\partial f(x, \theta)}{\partial \theta} dx = \frac{\partial}{\partial \theta} \left( \int f(x, \theta) dx \right) = \frac{\partial}{\partial \theta} (1) = 0 \end{aligned}$$

♣

Propiedad 2:

$$\begin{aligned} 0 &= E_0 \left[ \frac{\partial \log f(x, \theta)}{\partial \theta} \right] = \int \left( \frac{\partial \log f(x, \theta)}{\partial \theta} \right) f(x, \theta) dx \Rightarrow \\ \Rightarrow 0 &= \frac{\partial}{\partial \theta} \left( \int \left( \frac{\partial \log f(x, \theta)}{\partial \theta} \right) f(x, \theta) dx \right) = \int \left( \frac{\partial}{\partial \theta} \left( \frac{\partial \log f(x, \theta)}{\partial \theta} \right) f(x, \theta) \right) dx = \\ &= \int \frac{\partial}{\partial \theta} \left( \frac{\partial \log f(x, \theta)}{\partial \theta} \right) f(x, \theta) dx + \int \frac{\partial \log f(x, \theta)}{\partial \theta} \frac{\partial f(x, \theta)}{\partial \theta} dx = \\ &= E_0 \left[ \frac{\partial^2 \log f(x, \theta)}{\partial \theta^2} \right] + \int \frac{\partial \log f(x, \theta)}{\partial \theta} \frac{\frac{\partial f(x, \theta)}{\partial \theta}}{f(x, \theta)} f(x, \theta) dx = \\ &= E_0 \left[ \frac{\partial^2 \log f(x, \theta)}{\partial \theta^2} \right] + \int \left( \frac{\partial \log f(x, \theta)}{\partial \theta} \right)^2 f(x, \theta) dx = \\ &= E_0 \left[ \frac{\partial^2 \log f(x, \theta)}{\partial \theta^2} \right] + E_0 \left[ \left( \frac{\partial \log f(x, \theta)}{\partial \theta} \right)^2 \right] = 0 \end{aligned}$$

♣

**Nota:** En las condiciones que se exigen a un modelo para ser regular, aparece en la mayoría de los textos iii), es decir, que el recorrido de la variable no dependa del

parámetro. Sin embargo, veremos más adelante que en realidad esta condición es suficiente, supuesto que  $f(x, \theta)$  es una función regular, para asegurar iv) pero no es una condición necesaria. Además, un modelo que no verifique iii) pero sí iv) debe considerarse regular, pues tendrá las propiedades de éstos. Observemos además que la condición iv).a) es necesaria y suficiente para que se tenga  $E[(\partial \log f(x, \theta)) / (\partial \theta)] = 0$ , independientemente de que se verifique o no iii).

## 1.2 Información de Fisher. Propiedades.

**Definición:** Se llama Información de Fisher que la variable  $X$  proporciona sobre el parámetro  $\theta \in \Theta \subseteq \mathbb{R}$  a:

$$I_x(\theta) = E_{\theta} \left[ \left( \frac{\partial \log f(x, \theta)}{\partial \theta} \right)^2 \right]$$

Observemos entonces que en los modelos regulares, se tendrá:

$$I_x(\theta) = -E_{\theta} \left[ \frac{\partial^2 \log f(x, \theta)}{\partial \theta^2} \right]$$

Una interpretación muy extendida de la Información de Fisher como la cantidad de información que proporciona un dato acerca del parámetro y que puede verse, por ejemplo, en Peña, D. (1986) es la siguiente.

Dada una muestra  $x_1, \dots, x_n$  y dos valores del parámetro  $\theta_1, \theta_2$ , si se tiene

$$L(\theta_1 | x) = f(x_1, \dots, x_n; \theta_1) > f(x_1, \dots, x_n; \theta_2) = L(\theta_2 | x)$$

decimos que  $\theta_1$  es más verosímil que  $\theta_2$  para la muestra dada.

Ahora bien, al tener la verosimilitud unidades, no podemos comparar restando. Así, es más indicado estudiar estas diferencias a través de cocientes del tipo

$$\frac{L(\theta_1 | x)}{L(\theta_2 | x)}$$

En lugar del cociente, podemos usar la diferencia en logaritmos para comparar valores de la verosimilitud. Consideremos por ello la función  $\ell(\theta) = \ell(\theta | x) = \log L(\theta | x)$ .

Llamaremos discriminación contenida en la muestra  $x_1, \dots, x_n$  entre  $\theta_2$  y  $\theta_1$  a la diferencia entre los logaritmos de las verosimilitudes en ambos valores. Si el espacio paramétrico  $\Theta$  es un intervalo abierto, entonces llamamos discriminación relativa entre  $\theta_2$  y  $\theta_1$  a:

$$\frac{\ell(\theta_2) - \ell(\theta_1)}{\theta_2 - \theta_1}$$

Tomando límite cuando  $\theta_2$  tiende a  $\theta_1$ , obtenemos la tasa de discriminación de la muestra  $x_1, \dots, x_n$  respecto al parámetro  $\theta$  en el punto  $\theta_1$  como:

$$d(\theta_1, x) = \lim_{\theta_2 \rightarrow \theta_1} \frac{\ell(\theta_2) - \ell(\theta_1)}{\theta_2 - \theta_1} = \left. \frac{\partial \ell(\theta | x)}{\partial \theta} \right|_{\theta = \theta_1}$$

La tasa definida hasta ahora depende de  $x_1, \dots, x_n$ . Podemos hallar ahora un valor representativo de dicha tasa para todo el espacio muestral. En principio, podemos pensar en calcular el valor medio, pero, como es conocido, dicha media es nula. Así, para evitar el problema de compensación de signos, calculamos



$$E_{x|\theta} \left[ d(\theta, x)^2 \right] = E_{x|\theta} \left[ \left( \frac{\partial \ell(\theta | x)}{\partial \theta} \right)^2 \right] = I(\theta)$$

En conclusión, vemos que  $I(\theta)$  nos proporciona una medida de la capacidad de discriminación entre valores cercanos del parámetro que tiene un dato en un modelo dado. Es decir, dados  $\theta_1$  y  $\theta_2$  tales que  $I(\theta_1) > I(\theta_2)$ , entonces un dato “proporciona más información” cuando el verdadero valor del parámetro es  $\theta_1$  que cuando es  $\theta_2$ . Por tanto, la Información de Fisher puede interpretarse como la tasa de discriminación puntual entre distintos valores del parámetro. Observemos además, que esta interpretación es válida siempre que se cumpla  $E_{\theta} \left[ (\partial \ell) / (\partial \theta) \right] = 0$  independientemente de que se verifiquen otras condiciones de regularidad. Así,  $I(\theta) = E_{\theta} \left[ \left( (\partial \ell) / (\partial \theta) \right)^2 \right]$  es una medida de discriminación puntual siempre que  $E_{\theta} \left[ (\partial \ell) / (\partial \theta) \right] = 0$ .

Si el espacio paramétrico tiene dimensión  $k$ , entonces definimos la llamada matriz de Información de Fisher como

$$I(\theta) = E_{\theta} \left[ \begin{pmatrix} \frac{\partial \ell}{\partial \theta} \\ \left( \frac{\partial \ell}{\partial \theta} \right)^t \end{pmatrix} \right]$$

Bajo las condiciones de regularidad señaladas aplicadas a cada parámetro por separado, se verifican

1.  $E_{\theta} \left[ \frac{\partial \ell}{\partial \theta} \right] = 0 \quad \forall \theta$

2.  $E_{\theta} \left[ \begin{pmatrix} \frac{\partial \ell}{\partial \theta} \\ \left( \frac{\partial \ell}{\partial \theta} \right)^t \end{pmatrix} \right] = -E_{\theta} \left[ \frac{\partial^2 \ell}{\partial \theta \partial \theta^t} \right] \quad \forall \theta$

A partir de la propiedad 1, obtenemos que la matriz de Información de Fisher es una matriz de varianzas-covarianzas, más concretamente, la matriz correspondiente al vector gradiente del logaritmo de la verosimilitud.

Usando la propiedad 2, la matriz de Información es aquella cuyos elementos vienen dados por

$$(I(\theta))_{ij} = -E_{\theta} \left[ \frac{\partial^2 \ell}{\partial \theta_i \partial \theta_j} \right]$$

Al igual que en el caso unidimensional, esta matriz tiene sentido siempre que se verifique  $E_{\theta} \left[ \frac{\partial \ell}{\partial \theta} \right] = 0$ .

### **Propiedades de la Información de Fisher.**

Dada una muestra aleatoria simple  $X_1, \dots, X_n$  y un estadístico  $T(X_1, \dots, X_n)$ , a partir de ahora la notación será:

$I_{X_1}(\theta) = I(\theta) \equiv$  Información proporcionada por una muestra de tamaño 1.

$I_{\bar{X}}(\theta) \equiv$  Información proporcionada por la muestra de tamaño n.

$I_T(\theta) \equiv$  Información proporcionada por el estadístico.

**Proposición:** La Información de Fisher verifica las siguientes propiedades:

i) Si T y S son estadísticos independientes, entonces:

$$I_{(T,S)}(\theta) = I_T(\theta) + I_S(\theta)$$

Como consecuencia inmediata,  $I_{\bar{X}}(\theta) = nI_{X_1}(\theta)$

ii) T es auxiliar sii  $I_T(\theta) = 0 \quad \forall \theta$ .

(para una definición de estadístico auxiliar, véase la sección 1.5)

iii)  $I_{(T,S)}(\theta) = E_S [I_{T|S}(\theta)] + I_S(\theta) \quad \forall \theta$

(donde  $E_S$  significa esperanza en la distribución de  $S|\theta$ ).

iv)  $I_T(\theta) \leq I_{\bar{X}}(\theta)$ , dándose la igualdad  $\forall \theta$  sii T es suficiente.

Observemos que entonces de iii) y iv) se obtiene que (T,S) suficiente y S auxiliar implican que  $E_S [I_{T|S}(\theta)] = I_{\bar{X}}(\theta)$ .

**Dmt:**

**Nota:** A lo largo de las demostraciones se sobreentenderá que muchas de las igualdades proporcionadas son ciertas salvo un conjunto de medida nula.

$$\begin{aligned} \text{i) } f_0(t,s) = f_0(t)f_0(s) &\Rightarrow \frac{\partial^2 \log f_0(t,s)}{\partial \theta^2} = \frac{\partial^2 \log f_0(t)}{\partial \theta^2} + \frac{\partial^2 \log f_0(s)}{\partial \theta^2} \Rightarrow \\ &\Rightarrow -E_0 \left[ \frac{\partial^2 \log f_0(t,s)}{\partial \theta^2} \right] = -E_0 \left[ \frac{\partial^2 \log f_0(t)}{\partial \theta^2} \right] - E_0 \left[ \frac{\partial^2 \log f_0(s)}{\partial \theta^2} \right] \Rightarrow \\ &\Rightarrow I_{(T,S)}(\theta) = I_T(\theta) + I_S(\theta). \end{aligned}$$

♣

$$\text{ii) } I_T(\theta) = E_0 \left[ \left( \frac{\partial \log f_0(t)}{\partial \theta} \right)^2 \right] = 0 \Leftrightarrow \frac{\partial \log f_0(t)}{\partial \theta} = 0 \Leftrightarrow f_0(t) \text{ es independiente de } \theta$$

y la última condición es la definición de estadístico auxiliar (es decir, su distribución no depende del parámetro).

♣

$$\begin{aligned}
 \text{iii) } f_{\theta}(t, s) = f_{\theta}(t|s)f_{\theta}(s) &\Rightarrow -\frac{\partial^2 \log f_{\theta}(t, s)}{\partial \theta^2} = -\frac{\partial^2 \log f_{\theta}(t|s)}{\partial \theta^2} - \frac{\partial^2 \log f_{\theta}(s)}{\partial \theta^2} \\
 \iint_{\mathcal{X} \times \mathcal{S}} \left( -\frac{\partial^2 \log f_{\theta}(t, s)}{\partial \theta^2} \right) f_{\theta}(t, s) dt ds &= \\
 = \iint_{\mathcal{X} \times \mathcal{S}} \left( -\frac{\partial^2 \log f_{\theta}(t|s)}{\partial \theta^2} \right) f_{\theta}(t, s) dt ds &+ \iint_{\mathcal{X} \times \mathcal{S}} \left( -\frac{\partial^2 \log f_{\theta}(s)}{\partial \theta^2} \right) f_{\theta}(t, s) dt ds \Rightarrow \\
 \Rightarrow I_{(T, S)}(\theta) = \int_{\mathcal{S}} \left[ \int_{\mathcal{X}} \left( -\frac{\partial^2 \log f_{\theta}(t|s)}{\partial \theta^2} \right) f_{\theta}(t|s) dt \right] f_{\theta}(s) ds &+ \\
 + \int_{\mathcal{S}} \left( -\frac{\partial^2 \log f_{\theta}(s)}{\partial \theta^2} \right) f_{\theta}(s) ds \int_{\mathcal{X}} f_{\theta}(t|s) dt &= \\
 = E_S [I_{T|S}(\theta)] + I_S(\theta) &
 \end{aligned}$$

♣

$$\text{iv) } f_{\theta}(\bar{x}) = f_{\theta}(\bar{x}|t) f_{\theta}(t) \Rightarrow \frac{\partial^2 \log f_{\theta}(\bar{x})}{\partial \theta^2} = \frac{\partial^2 \log f_{\theta}(\bar{x}|t)}{\partial \theta^2} + \frac{\partial^2 \log f_{\theta}(t)}{\partial \theta^2}$$

Ahora, tomando esperanza y cambiando el signo, obtenemos que:

$$\begin{aligned}
 I_{\bar{X}}(\theta) &= -E_{\theta} \left[ \frac{\partial^2 \log f_{\theta}(\bar{x}|t)}{\partial \theta^2} \right] + I_T(\theta) \Rightarrow \\
 \Rightarrow I_{\bar{X}}(\theta) &= E_{\theta} \left[ \left( \frac{\partial \log f_{\theta}(\bar{x}|t)}{\partial \theta} \right)^2 \right] + I_T(\theta) \Rightarrow \\
 I_{\bar{X}}(\theta) &\geq I_T(\theta) \quad \forall \theta
 \end{aligned}$$

Además, la igualdad entre ambas informaciones se dará sii:

$$E_{\theta} \left[ \left( \frac{\partial \log f_{\theta}(\bar{x} | t)}{\partial \theta} \right)^2 \right] = 0 \Leftrightarrow \frac{\partial \log f_{\theta}(\bar{x} | t)}{\partial \theta} = 0 \Leftrightarrow f_{\theta}(\bar{x} | t) \text{ no depende de } \theta$$

y la última condición expresada es una de las posibles definiciones de suficiencia.



### 1.3 Medidas de Información generales.

La interpretación de la medida de Información de Fisher puede hacerse desde otra perspectiva más atrayente y clarificadora.

Consideremos una familia de densidades  $\{ f(x, \theta), \theta \in \Theta \subseteq \mathbb{R}^k \}$ . Dados los valores  $\theta_1, \theta_2 \in \Theta$ , vamos a plantear la cantidad de información entre  $f(\cdot, \theta_1)$  y  $f(\cdot, \theta_2)$  como una medida de “lo diferente” que son ambas funciones. Así, si la información es elevada, un solo dato nos permitirá discriminar bien entre  $\theta_1$  y  $\theta_2$  y si la información es pequeña será porque ambas densidades serán parecidas y no nos será fácil discernir entre los valores  $\theta_1$  y  $\theta_2$ . Una posibilidad evidente sería tomar como cantidad de información una medida basada en una distancia entre las dos distribuciones de probabilidad.

En Pitman, E.J. (1979) podemos ver que no es buena solución usar  $\int |f(x, \theta_1) - f(x, \theta_2)| dx$  debido a que el valor absoluto es una función analíticamente complicada de manejar y que además la norma  $L^1$  da el mismo peso a diferencias iguales entre  $f_1$  y  $f_2$  independientemente de que éstas tomen valores pequeños o grandes. En ambos aspectos, tiene mejor comportamiento la norma  $L^2$  de  $\sqrt{f_1} - \sqrt{f_2}$ , por lo que Pitman toma la siguiente definición:

**Definición:** Sean  $P_1$  y  $P_2$  medidas de probabilidad definidas en un mismo  $\sigma$ -álgebra de un espacio  $\chi$ . Sean  $f_1$  y  $f_2$  las funciones de densidad respecto a una medida dominante  $\mu$ . Se define la cantidad de información entre  $P_1$  y  $P_2$  como:

$$J(P_1, P_2) = \int (\sqrt{f_1} - \sqrt{f_2})^2 d\mu$$

Aplicando esta definición a la situación de inferencia paramétrica en la que trabajaremos habitualmente, la medida de información quedaría

$$J(\theta_1, \theta_2) = \int (\sqrt{f(x, \theta_1)} - \sqrt{f(x, \theta_2)})^2 dx$$

Esta medida de distancia puede encontrarse también en Jeffreys, H. (1946), Matusita, K. (1955), etc. En este último trabajo, aparece también el concepto de afinidad.

**Definición:** En la situación de la definición anterior, se define la afinidad entre  $P_1$  y  $P_2$  como

$$A(P_1, P_2) = \int \sqrt{f_1 f_2} d\mu$$

Por tanto, es inmediato ver que se verifica

$$J(\theta_1, \theta_2) = 2 (1 - A(\theta_1, \theta_2))$$

Por otra parte,  $0 \leq A^2(\theta_1, \theta_2) = \left( \int \sqrt{f_1} \sqrt{f_2} d\mu \right)^2 \leq \left( \int f_1 d\mu \right) \left( \int f_2 d\mu \right) = 1$  y así, también se verifica  $0 \leq J(\theta_1, \theta_2) \leq 2$ .

Hay dos casos extremos que ayudan a interpretar estas medidas.

- Si  $\theta_1 = \theta_2$ , entonces  $A(\theta_1, \theta_2) = 1$  y  $J(\theta_1, \theta_2) = 0$ .
- Si  $\text{rg}(f(x, \theta_1)) \cap \text{rg}(f(x, \theta_2)) = \emptyset$ , entonces  $A(\theta_1, \theta_2) = 0$  y  $J(\theta_1, \theta_2) = 2$ .

Es decir, si  $\theta_1 = \theta_2$ , con una observación nos sería totalmente imposible discernir entre las dos distribuciones (por ser iguales); con dos distribuciones de rangos disjuntos, una única observación nos permite decidir con total seguridad qué valor del parámetro ha generado el dato.

Esta medida de información ha sido bastante usada por diversos autores. No obstante, tiene el inconveniente de no ser una medida aditiva, es decir, la información que proporcionan dos muestras o estadísticos independientes no es la suma de las informaciones correspondientes a cada uno de ellos. En efecto, es fácil ver que la afinidad es multiplicativa; dados dos estadísticos para el mismo modelo generador de datos S y T independientes, entonces

$$A_{(S,T)}(\theta_1, \theta_2) = A_S(\theta_1, \theta_2) A_T(\theta_1, \theta_2)$$

y de esta relación se deduce que  $J_{(S,T)}(\theta_1, \theta_2) \neq J_S(\theta_1, \theta_2) + J_T(\theta_1, \theta_2)$ . Por ello, en el presente trabajo vamos a usar fundamentalmente la definición adoptada por Akahira, M. and Takeuchi, K. (1991), que se basa en el logaritmo de la afinidad y por tanto resulta ser una medida aditiva. Cabe destacar la medida de información dada por Jeffreys, H. (1946) y que utiliza, entre otros, Kullback, S. (1968) y que viene dada por

$$\tilde{J}(P_1, P_2) = \int (f_1 - f_2) \log \frac{f_2}{f_1} d\mu$$

La definición que vamos a usar principalmente es la que sigue:

**Definición:** Sean  $P_1$  y  $P_2$  medidas de probabilidad definidas en un mismo  $\sigma$ -álgebra de un espacio  $\chi$ . Sean  $f_1$  y  $f_2$  las funciones de densidad respecto a una medida dominante  $\mu$ . Se define la cantidad de información entre  $P_1$  y  $P_2$  como:

$$I(P_1, P_2) = -8 \log \int \sqrt{f_1} \sqrt{f_2} \, d\mu = -8 \log A(P_1, P_2)$$

En el caso en que trabajaremos habitualmente, es decir, en una familia  $\{f(x, \theta), \theta \in \Theta \subseteq \mathbb{R}^k\}$  donde las densidades son con respecto a la medida de Lebesgue, tendremos:

$$I(\theta_1, \theta_2) = -8 \log \int \sqrt{f(x, \theta_1)} \sqrt{f(x, \theta_2)} \, dx = -8 \log A(\theta_1, \theta_2)$$

Cabe quizás destacar que con esta definición, la cantidad de información está en el intervalo  $[0, +\infty]$ .

**Ejemplo 1.1:** Consideremos el modelo  $\text{Exp}(\theta)$ ,  $\theta > 0$ .

Vamos a calcular la afinidad y la información para dos valores cualesquiera  $\theta_1, \theta_2 > 0$ .

$$A(\theta_1, \theta_2) = \int_0^{+\infty} \sqrt{\theta_1 \theta_2} \, e^{-\theta_1 x} \, e^{-\theta_2 x} \, dx = \frac{2\sqrt{\theta_1 \theta_2}}{\theta_1 + \theta_2}$$

$$I(\theta_1, \theta_2) = -8 \log A(\theta_1, \theta_2) = -8 \log \left( \frac{2\sqrt{\theta_1 \theta_2}}{\theta_1 + \theta_2} \right)$$

Para entender mejor el significado, vamos a considerar uno de los valores del parámetro fijo (por ejemplo igual a 1), quedando entonces:



$$A(1, \theta) = \frac{2\sqrt{\theta}}{1+\theta}$$

$$I(1, \theta) = -8 \log\left(\frac{2\sqrt{\theta}}{1+\theta}\right)$$

Las representaciones gráficas de ambas funciones son las que se ofrecen a continuación:

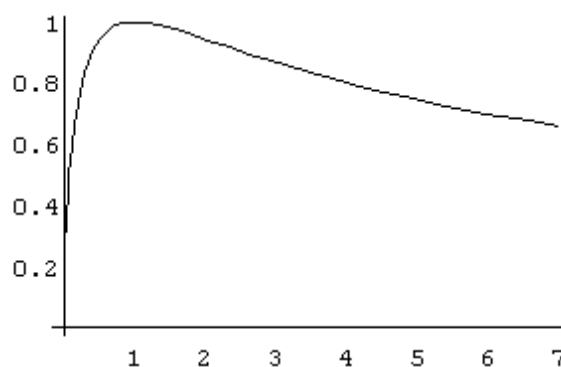


Figura 1.1: Gráfica de  $A(1, \theta)$  en un modelo exponencial.

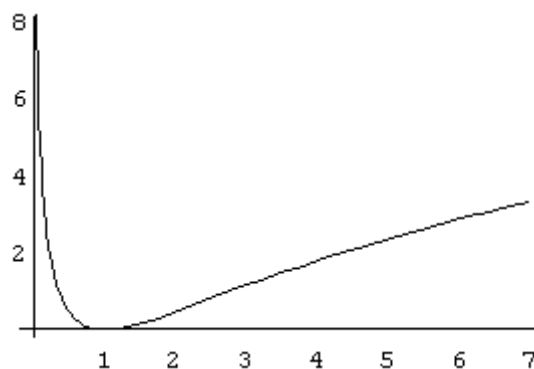


Figura 1.2: Gráfica de  $I(1, \theta)$  en un modelo exponencial.

En términos muy generales, estas gráficas significan que podemos discriminar mejor entre una  $\text{Exp}(1)$  y una  $\text{Exp}(\theta_0)$  cuando  $\theta_0 < 1$ .

**Ejemplo 1.2:** Consideremos  $X_1, \dots, X_n$  i.i.d.  $U(\theta - \frac{1}{2}, \theta + \frac{1}{2})$ .

En este caso, tenemos:  $f(x, \theta) = 1 \quad \theta - \frac{1}{2} \leq x \leq \theta + \frac{1}{2}$

Sean  $\theta_1$  y  $\theta_2$  tales que  $\text{rg}(X_{\theta_1}) \cap \text{rg}(X_{\theta_2}) \neq \emptyset$ , pues en otro caso la afinidad sería nula y por tanto la información valdría  $+\infty$ .

1.  $\theta_2 > \theta_1$ :

$$I(\theta_1, \theta_2) = -8 \log \int_{\theta_2 - 1/2}^{\theta_1 + 1/2} dx = -8 \log \left( \theta_1 + \frac{1}{2} - \theta_2 + \frac{1}{2} \right) = -8 \log(1 - \theta_2 + \theta_1)$$

2.  $\theta_1 > \theta_2$ :

$$I(\theta_1, \theta_2) = -8 \log \int_{\theta_1 - 1/2}^{\theta_2 + 1/2} dx = -8 \log \left( \theta_2 + \frac{1}{2} - \theta_1 + \frac{1}{2} \right) = -8 \log(1 - \theta_1 + \theta_2)$$

Resumiendo todos los casos, obtenemos:

$$I(\theta_1, \theta_2) = \begin{cases} -8 \log(1 - \theta_2 + \theta_1) & \text{si } \theta_2 \geq \theta_1 \text{ y } \text{rg}(X_{\theta_1}) \cap \text{rg}(X_{\theta_2}) \neq \emptyset \\ -8 \log(1 - \theta_1 + \theta_2) & \text{si } \theta_1 \geq \theta_2 \text{ y } \text{rg}(X_{\theta_1}) \cap \text{rg}(X_{\theta_2}) \neq \emptyset \\ +\infty & \text{si } \text{rg}(X_{\theta_1}) \cap \text{rg}(X_{\theta_2}) = \emptyset \end{cases}$$

Podemos entonces escribir:

$$I(\theta_1, \theta_2) = \begin{cases} -8 \log(1 - |\theta_1 - \theta_2|) & \text{si } |\theta_1 - \theta_2| \leq 1 \\ +\infty & \text{si } |\theta_1 - \theta_2| > 1 \end{cases}$$

y por tanto,

$$I_{\underline{X}}(\theta_1, \theta_2) = \begin{cases} -8n \log(1 - |\theta_1 - \theta_2|) & \text{si } |\theta_1 - \theta_2| \leq 1 \\ +\infty & \text{si } |\theta_1 - \theta_2| > 1 \end{cases}$$

Vamos a centrarnos en este capítulo en dos aspectos de esta medida de información. Veremos que reproduce todas las propiedades importantes de la Información de Fisher y estudiaremos la conexión existente entre ambas medidas. Dicha conexión nos permitirá obtener una definición más precisa y menos restrictiva de modelo regular.

Antes de entrar en ello, hemos de señalar que las medidas que hemos denominado I, J pueden considerarse equivalentes entre sí, en el sentido de que se da la siguiente relación entre ellas.

**Proposición:** Siendo  $I_\theta(h) = I(\theta, \theta + h)$  y  $J_\theta(h) = J(\theta, \theta + h)$ , se verifica

$$\lim_{h \rightarrow 0} \frac{J_\theta(h)}{I_\theta(h)} = \frac{1}{4}$$

**Dmt:**

$$\lim_{h \rightarrow 0} \frac{J_\theta(h)}{I_\theta(h)} = -\frac{1}{4} \lim_{h \rightarrow 0} \frac{1 - A_\theta(h)}{\log A_\theta(h)} = -\frac{1}{4} \lim_{h \rightarrow 0} \frac{-A'_\theta(h)}{\frac{A'_\theta(h)}{A_\theta(h)}} = \frac{1}{4} \lim_{h \rightarrow 0} A_\theta(h) = \frac{1}{4}$$

**Corolario:**

$$\lim_{h \rightarrow 0} \frac{I(\theta, \theta + h)}{|h|^k} \propto \lim_{h \rightarrow 0} \frac{J(\theta, \theta + h)}{|h|^k} \quad \forall k > 0$$

donde en la proporcionalidad anterior debe entenderse que o bien ambos son cero, o bien ambos son infinito y, si ambos son finitos, entonces son proporcionales. (En realidad bastaría definir la información  $J(\theta_1, \theta_2)$  como cuatro veces la distancia al cuadrado entre las distribuciones para que los límites fuesen iguales). Este corolario nos dice así que, para estudiar la sensibilidad de una familia ante pequeños cambios en los parámetros, es indiferente usar una u otra medida de información.

### **1.3.1. Propiedades de la medida de información general.**

La medida de información de Akahira verifica las siguientes propiedades.

**Propiedad 1:** Dados  $T_1$  y  $T_2$  estadísticos independientes, se verifica:

$$I_{(T_1, T_2)}(\theta_1, \theta_2) = I_{T_1}(\theta_1, \theta_2) + I_{T_2}(\theta_1, \theta_2)$$

**Propiedad 2:**  $T(X)$  es un estadístico auxiliar (es decir, su distribución es independiente del parámetro) si y sólo si  $I_T(\theta_1, \theta_2) = 0$ .

**Propiedad 3:**

a) Dados  $T_1$  y  $T_2$  se verifica:

$$I_{(T_1, T_2)}(\theta_1, \theta_2) = -8 \log \left\{ E_{T_2^*} \left[ A_{T_1|T_2}(\theta_1, \theta_2) \right] \right\} + I_{T_2}(\theta_1, \theta_2)$$

donde  $T_2^*$  es un estadístico cuya función de densidad es:

$$f^*(t_2) = \frac{f_{\theta_1}(t_2)^{1/2} f_{\theta_2}(t_2)^{1/2}}{\int (f_{\theta_1}(t_2)^{1/2} f_{\theta_2}(t_2)^{1/2}) dt_2}$$

**b)** Dados  $T_1$  y  $T_2$  siendo este último auxiliar, se tiene:

$$I_{(T_1, T_2)}(\theta_1, \theta_2) = -8 \log \left( E_{T_2} \left[ A_{T_1 T_2}(\theta_1, \theta_2) \right] \right)$$

donde  $E_{T_2}$  representa esperanza bajo la densidad de  $T_2$ .

**Propiedad 4:** Dado un estadístico  $T(X)$ , se verifica que:

$$I_T(\theta_1, \theta_2) \leq I_{\bar{X}}(\theta_1, \theta_2)$$

dándose la igualdad para todo par de valores  $\theta_1, \theta_2$  si y sólo si  $T$  es suficiente para el parámetro  $\theta$ .

**Dmt:**

**Propiedad 1:**

$$\begin{aligned} I_{(T_1, T_2)}(\theta_1, \theta_2) &= -8 \log \iint f_{\theta_1}(t_1, t_2)^{1/2} f_{\theta_2}(t_1, t_2)^{1/2} dt_1 dt_2 = \\ &= -8 \log \left[ \left( \int f_{\theta_1}(t_1)^{1/2} f_{\theta_2}(t_1)^{1/2} dt_1 \right) \left( \int f_{\theta_1}(t_2)^{1/2} f_{\theta_2}(t_2)^{1/2} dt_2 \right) \right] = \\ &= -8 \log \left[ A_{T_1}(\theta_1, \theta_2) A_{T_2}(\theta_1, \theta_2) \right] = I_{T_1}(\theta_1, \theta_2) + I_{T_2}(\theta_1, \theta_2) \end{aligned}$$

♣

Notemos que a partir de este resultado es inmediato establecer que la información proporcionada por una muestra aleatoria simple de tamaño  $n$  será  $n$  veces la proporcionada por una muestra de tamaño 1. Con la notación  $I(\theta_1, \theta_2)$  nos

referiremos en lo que sigue a la información proporcionada por una muestra de tamaño 1;  $I_{\bar{X}}(\theta_1, \theta_2)$  será la información de la muestra completa e  $I_T(\theta_1, \theta_2)$  la proporcionada por el estadístico  $T(X)$ .

**Propiedad 2:**

$$I_T(\theta_1, \theta_2) = -8 \log \int f_{\theta_1}(t)^{1/2} f_{\theta_2}(t)^{1/2} dt = 0 \Leftrightarrow$$

$$\Leftrightarrow A_T(\theta_1, \theta_2) = 1 \Leftrightarrow f_{\theta_1}(t) = f_{\theta_2}(t) \Leftrightarrow T \text{ auxiliar.}$$

♣

**Propiedad 3:**

a)

$$I_{(T_1, T_2)}(\theta_1, \theta_2) = -8 \log \iint f_{\theta_1}(t_1, t_2)^{1/2} f_{\theta_2}(t_1, t_2)^{1/2} dt_1 dt_2 =$$

$$= -8 \log \iint f_{\theta_1}(t_1 | t_2)^{1/2} f_{\theta_1}(t_2)^{1/2} f_{\theta_2}(t_1 | t_2)^{1/2} f_{\theta_2}(t_2)^{1/2} dt_1 dt_2 =$$

$$= -8 \log \int \left[ f_{\theta_1}(t_1 | t_2)^{1/2} \int f_{\theta_2}(t_1 | t_2)^{1/2} f_{\theta_2}(t_1 | t_2)^{1/2} dt_1 \right] f_{\theta_1}(t_2)^{1/2} f_{\theta_2}(t_2)^{1/2} dt_2 =$$

$$= -8 \log \int A_{(T_1|T_2)}(\theta_1, \theta_2) f_{\theta_1}(t_2)^{1/2} f_{\theta_2}(t_2)^{1/2} dt_2 =$$

$$-8 \log \left\{ \left[ \int A_{T_1|T_2}(\theta_1, \theta_2) \frac{f_{\theta_1}(t_2)^{1/2} f_{\theta_2}(t_2)^{1/2}}{\int f_{\theta_1}(t_2)^{1/2} f_{\theta_2}(t_2)^{1/2} dt_2} dt_2 \right] \left[ \int f_{\theta_1}(t_2)^{1/2} f_{\theta_2}(t_2)^{1/2} dt_2 \right] \right\} =$$

$$= -8 \log \left\{ E_{T_2^*} [A_{T_1|T_2}(\theta_1, \theta_2)] \right\} - 8 \log A_{T_2}(\theta_1, \theta_2) =$$

$$= -8 \log \left\{ E_{T_2^*} [A_{T_1|T_2}(\theta_1, \theta_2)] \right\} + I_{T_2}(\theta_1, \theta_2).$$

♣

b) Es una consecuencia directa de a), pues cuando  $T_2$  es auxiliar, la mencionada densidad de  $T_2^*$  sería:

$$f^*(t_2) = \frac{f(t_2)^{1/2} f(t_2)^{1/2}}{\int f(t_2)^{1/2} f(t_2)^{1/2} dt_2} = f(t_2)$$

con lo que  $T_2^*$  coincide con  $T_2$  y así queda probado 3.b).

♣

**Propiedad 4:**

$$\begin{aligned} A_{\bar{X}}(\theta_1, \theta_2) &= \int f_{\theta_1}(\bar{x})^{1/2} f_{\theta_2}(\bar{x})^{1/2} d\bar{x} = \\ &= \int f_{\theta_1}(\bar{x} | t)^{1/2} f_{\theta_1}(t)^{1/2} f_{\theta_2}(\bar{x} | t)^{1/2} f_{\theta_2}(t)^{1/2} d\bar{x} = \\ &= \int A_{\bar{X}|T}(\theta_1, \theta_2) f_{\theta_1}(t)^{1/2} f_{\theta_2}(t)^{1/2} dt \end{aligned}$$

Observando que la afinidad de la muestra condicionada al estadístico ha de ser menor o igual que 1, obtenemos que:

$$A_{\bar{X}}(\theta_1, \theta_2) \leq \int f_{\theta_1}(t)^{1/2} f_{\theta_2}(t)^{1/2} dt = A_T(\theta_1, \theta_2)$$

y, consecuentemente, al ser la información una función decreciente de la afinidad, se obtiene:

$$I_{\bar{X}}(\theta_1, \theta_2) \geq I_T(\theta_1, \theta_2)$$

Además, la igualdad se dará si y sólo si la afinidad reseñada anteriormente vale 1, es decir, si y sólo si  $f_{\theta_1}(\bar{x} | t) = f_{\theta_2}(\bar{x} | t) \quad \forall \theta_1 \neq \theta_2$ , lo que equivale a decir que la distribución de la muestra condicionada al estadístico es independiente de  $\theta$ , lo que a su vez es equivalente a decir que el estadístico es suficiente.

♣

**1.3.2 Relación entre las medidas de información generales y la Información de Fisher.**

La relación existente entre ambas medidas de información se establece a partir de la siguiente proposición que puede verse en Akahira, M. and Takeuchi, K. (1991).

**Proposición:** En los modelos regulares, para  $h$  suficientemente pequeño se verifica:

$$I(\theta, \theta+h) = I(\theta)h^2 + o(h^2)$$

**Dmt:**

$$\begin{aligned} I(\theta, \theta+h) &= -8 \log \int f(x, \theta)^{1/2} f(x, \theta+h)^{1/2} dx = \\ &= -8 \log \int \exp \left\{ \frac{1}{2} \log f(x, \theta) + \frac{1}{2} \log f(x, \theta+h) \right\} dx = \end{aligned}$$

Haciendo un desarrollo de Taylor de  $\log f(x, \theta+h)$  en torno a  $\theta$  obtendremos:

$$\begin{aligned} &= -8 \log \int \exp \left\{ \frac{1}{2} \log f(x, \theta) + \frac{1}{2} \left( \log f(x, \theta) + h \frac{\partial \log f(x, \theta)}{\partial \theta} + \frac{1}{2} h^2 \frac{\partial^2 \log f(x, \theta)}{\partial \theta^2} + o(h^2) \right) \right\} dx = \\ &= -8 \log \int \exp \left\{ \log f(x, \theta) + \frac{1}{2} h \frac{\partial \log f(x, \theta)}{\partial \theta} + \frac{1}{4} h^2 \frac{\partial^2 \log f(x, \theta)}{\partial \theta^2} + o(h^2) \right\} dx = \end{aligned}$$

Ahora, desarrollamos la función exponencial en torno a  $\log f(x, \theta)$  obteniendo:

$$= -8 \log \int \left\{ \begin{array}{l} \exp(\log f(x, \theta)) + \left( \frac{1}{2} h \frac{\partial \log f(x, \theta)}{\partial \theta} + \frac{1}{4} h^2 \frac{\partial^2 \log f(x, \theta)}{\partial \theta^2} + o(h^2) \right) \\ \exp(\log f(x, \theta)) + \frac{1}{2} (\dots + \dots + \dots)^2 \exp(\log f(x, \theta)) + o(h^2) \end{array} \right\} dx =$$



Al desarrollar el cuadrado obtenido anteriormente, sólo tendremos en cuenta el cuadrado del primer término, pues el resto de sumandos son  $o(h^2)$ , obteniendo de esta forma:

$$\begin{aligned}
 &= -8 \log \left\{ \int f(x, \theta) dx + \int \frac{1}{2} h \frac{\partial \log f(x, \theta)}{\partial \theta} f(x, \theta) dx + \int \frac{1}{4} h^2 \frac{\partial^2 \log f(x, \theta)}{\partial \theta^2} f(x, \theta) dx + \right. \\
 &\quad \left. \frac{1}{2} \int \frac{1}{4} h^2 \frac{\partial \log f(x, \theta)}{\partial \theta} f(x, \theta) dx + o(h^2) \right\} = \\
 &= -8 \log \left\{ 1 - \frac{h^2}{4} I(\theta) + \frac{h^2}{8} I(\theta) + o(h^2) \right\} = -8 \log \left\{ 1 - \frac{h^2}{8} I(\theta) + o(h^2) \right\} = \\
 &= -8 \left\{ \log 1 + \left( -\frac{h^2}{8} I(\theta) + o(h^2) \right) + o(h^2) \right\} = I(\theta) h^2 + o(h^2)
 \end{aligned}$$

♣

**Corolario:** En los modelos regulares, se verifica:

$$I(\theta) = \lim_{h \rightarrow 0} \frac{I(\theta, \theta + h)}{h^2}$$

De esta forma, a partir de una medida de información entre dos distribuciones (correspondientes a dos valores distintos del parámetro), podemos obtener la Información de Fisher como medida de información puntual. Por ello, según Pitman, E. J. (1979), el término “información” es inadecuado y propone que  $I(\theta) = E_{\theta} \left[ \left( (\partial \ell) / (\partial \theta) \right)^2 \right]$  debería ser la “sensibilidad” del modelo en  $\theta$ , ya que nos muestra la variación que experimentan las distribuciones ante pequeños cambios en los parámetros.

Vamos a comprobar directamente el corolario sobre algunos ejemplos, lo que permitirá clarificar algo más el significado de la información que hemos definido y su relación con la Información de Fisher.

**Ejemplo 2.3:** Volvamos a considerar el modelo  $\text{Exp}(\theta)$ .

Teniendo en cuenta la expresión obtenida anteriormente para  $I(\theta_1, \theta_2)$ , se tendrá:

$$I(\theta, \theta+h) = -8 \log \left( \frac{2\sqrt{\theta(\theta+h)}}{2\theta+h} \right)$$

Cálculos elementales de límites ponen de manifiesto que, efectivamente,

$$\lim_{h \rightarrow 0} \frac{I(\theta, \theta+h)}{h^2} = \frac{1}{\theta^2} = I(\theta)$$

Además, para poner de manifiesto más claramente el significado del corolario, vamos a considerar el valor de la información para un  $h$  fijo y “pequeño”. Por motivos de claridad en las gráficas que haremos a continuación, vamos a tomar  $h=0.25$  y estudiamos así la función  $I(\theta, \theta+0.25)$ . Esta función proporcionará una medida de “lo distintas” que son las densidades correspondientes a los valores  $\theta$  y  $\theta+h$  lo que ya hemos señalado que se relaciona con la capacidad de los datos de discriminar entre  $\theta$  y  $\theta+h$ . La representación gráfica de esta función es la que se ofrece a continuación.

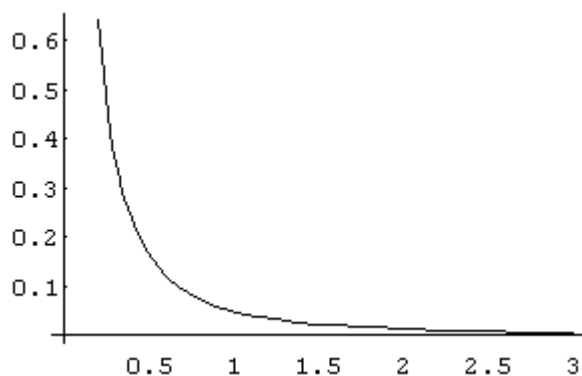


Figura 1.3: Gráfica de  $I(\theta, \theta+0.25)$  en un modelo exponencial.

Como es conocido, la información de Fisher para el modelo exponencial es  $I(\theta) = 1/\theta^2$ , cuya gráfica es la siguiente:

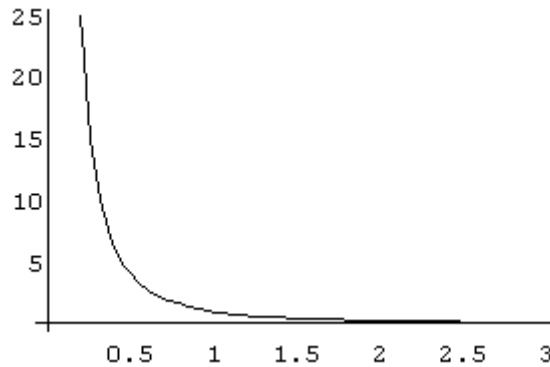


Figura 1.4: Gráfica de  $I(\theta)=1/\theta^2$  en un modelo exponencial.

Como puede apreciarse, a partir de ambas gráficas puede deducirse que la capacidad de discriminación es mucho mayor para valores próximos a cero, quedando así claro que las dos medidas de información tienen el mismo significado en el caso regular.

**Ejemplo 1.4:** Consideremos el modelo  $N(\mu, \sigma^2)$ . En este caso, la información correspondiente a dos valores distintos del espacio paramétrico será:

$$\begin{aligned}
 I((\mu_1, \mu_2), (\sigma_1, \sigma_2)) &= -8 \log \left( \int \frac{1}{\sqrt{2\pi\sigma_1\sigma_2}} \exp \left\{ -\frac{\sigma_2^2(x-\mu_1)^2 + \sigma_1^2(x-\mu_2)^2}{4\sigma_1^2\sigma_2^2} \right\} dx \right) = \\
 &= -8 \log \left( \sqrt{\frac{2\sigma_1\sigma_2}{\sigma_1^2 + \sigma_2^2}} \exp \left\{ -\frac{\mu_1 - \mu_2}{4(\sigma_1^2 + \sigma_2^2)} \right\} \right) = 4 \log \left( \frac{\sigma_1^2 + \sigma_2^2}{2\sigma_1\sigma_2} \right) + \frac{2(\mu_1 - \mu_2)^2}{\sigma_1^2 + \sigma_2^2}
 \end{aligned}$$

a) Consideremos ahora un modelo con varianza conocida, por ejemplo,  $N(\mu, 1)$ .

$$I(\mu_1, \mu_2) = (\mu_1 - \mu_2)^2 \Rightarrow I(\mu, \mu + h) = h^2$$

De esta forma, efectivamente, se verifica:

$$\lim_{h \rightarrow 0} \frac{I(\mu, \mu + h)}{h^2} = 1 = I(\mu)$$

b) Estudiamos ahora el caso de un modelo  $N(0, \sigma^2)$ .

$$I(\sigma, \sigma + h) = 4 \log \left( \frac{\sigma^2 + (\sigma + h)^2}{2\sigma(\sigma + h)} \right) = 4 \log \left( 1 + \frac{h^2}{2\sigma(\sigma + h)} \right)$$

$$\lim_{h \rightarrow 0} \frac{I(\sigma, \sigma + h)}{h^2} = 4 \log \left( \lim_{h \rightarrow 0} \left[ 1 + \frac{h^2}{2\sigma(\sigma + h)} \right]^{\frac{1}{h^2}} \right) = 4 \log \left( e^{\frac{1}{2\sigma^2}} \right) = \frac{2}{\sigma^2} = I(\sigma)$$

**Nota:** Obsérvese cómo la información puede calcularse sin problemas en el caso de más de un parámetro. De hecho, la definición de información puede establecerse de forma más general usando directamente medidas de probabilidad. En concreto, en Akahira, M. and Takeuchi, K. (1991), la definición que se ofrece es:

Dada una variable aleatoria  $X$  definida sobre un espacio muestral  $\chi$  y  $P$  y  $Q$  medidas absolutamente continuas respecto a una medida  $\sigma$ -finita  $\mu$ , definimos la cantidad de información entre  $P$  y  $Q$  como:

$$I_X(P, Q) = -8 \log \int \left( \frac{dP}{d\mu} \frac{dQ}{d\mu} \right)^{1/2} d\mu$$

La definición ofrecida inicialmente en este trabajo no es más que un caso particular de esta donde  $P$  y  $Q$  son las medidas de probabilidad inducidas por las

variables correspondientes a los parámetros  $\theta_1, \theta_2$ , la medida  $\sigma$ -finita considerada es la de Lebesgue y las derivadas respecto a esta medida son las funciones de densidad de las variables.

#### 1.4 Definición alternativa de modelo regular.

Según hemos visto, con las condiciones de regularidad exigidas en el apartado 1.1, se tiene que  $I(\theta, \theta+h)$  y  $J(\theta, \theta+h)$  son del orden de  $h^2$  y además

$$\lim_{h \rightarrow 0} \frac{I(\theta, \theta+h)}{h^2} \propto \lim_{h \rightarrow 0} \frac{J(\theta, \theta+h)}{h^2} \propto I(\theta).$$

Sin embargo, estas condiciones no son todas necesarias para obtener el resultado. Veamos esto con un ejemplo.

**Ejemplo 1.5:** Consideremos  $f(x, \theta) = \frac{1}{2} e^{(\theta-x)} (x-\theta)^2 \quad x \geq \theta$

Vamos a comprobar directamente que en este caso se verifican:

$$1. \lim_{h \rightarrow 0} \frac{I(\theta, \theta+h)}{|h|} = E \left[ \frac{\partial \log f(x, \theta)}{\partial \theta} \right] = 0.$$

$$(\text{y recordemos que } E \left[ \frac{\partial \log f(x, \theta)}{\partial \theta} \right] = 0 \Leftrightarrow \frac{\partial}{\partial \theta} \int f(x, \theta) dx = \int \frac{\partial f(x, \theta)}{\partial \theta} dx).$$

$$2. \lim_{h \rightarrow 0} \frac{I(\theta, \theta+h)}{h^2} = 1.$$

$$3. I(\theta) = E_{\theta} \left[ \left( \frac{\partial \ell}{\partial \theta} \right)^2 \right] = - E \left[ \frac{\partial^2 \ell}{\partial \theta^2} \right] = 1.$$

con lo cual tenemos un modelo en el que el recorrido depende del parámetro pero  $I(\theta, \theta+h)$  es de orden 2 y además al hacer el límite de  $I(\theta, \theta+h)$  dividido por  $h^2$  obtenemos  $I(\theta)$ .

Comencemos con el caso  $h < 0$ .

$$\begin{aligned}
 A(\theta, \theta+h) &= \int_{\theta}^{+\infty} \frac{1}{2} e^{\frac{\theta-x}{2}} (x-\theta) e^{\frac{\theta+h-x}{2}} (x-\theta-h) dx = \\
 &= \frac{1}{2} \int_{\theta}^{+\infty} e^{-x+\frac{h}{2}} (x-\theta)(x-\theta-h) dx = \\
 &= \frac{1}{2} \int_0^{+\infty} e^{-z+\frac{h}{2}} z(z-h) dz = \frac{1}{2} \left( -z(z-h) e^{-z+\frac{h}{2}} \right) \Big|_0^{+\infty} + \frac{1}{2} \int_0^{+\infty} e^{-z+\frac{h}{2}} (2z-h) dz = \\
 &= \frac{1}{2} \int_0^{+\infty} e^{-z+\frac{h}{2}} (2z-h) dz = \frac{1}{2} \left( -(2z-h) e^{-z+\frac{h}{2}} \right) \Big|_0^{+\infty} + \int_0^{+\infty} e^{-z+\frac{h}{2}} dz = \\
 &= -\frac{h}{2} e^{\frac{h}{2}} - e^{-z+\frac{h}{2}} \Big|_0^{+\infty} = -\frac{h}{2} e^{\frac{h}{2}} + e^{\frac{h}{2}} = e^{\frac{h}{2}} \left( 1 - \frac{h}{2} \right).
 \end{aligned}$$

Por lo tanto, la información quedará como:

$$I(\theta, \theta+h) = -8 \log A(\theta, \theta+h) = -4h - 8 \log \left( \frac{2-h}{h} \right)$$

De esta forma, tenemos

$$\lim_{h \rightarrow 0^-} \frac{I(\theta, \theta+h)}{|h|} = \lim_{h \rightarrow 0^-} \frac{-4h - 8 \log \left( \frac{2-h}{h} \right)}{-h} = \lim_{h \rightarrow 0^-} \left[ 4 + 8 \frac{\frac{1}{2}}{\frac{2-h}{h}} \right] =$$

$$= \lim_{h \rightarrow 0^+} \left[ 4 - \frac{8}{2-h} \right] = 4 - 4 = 0$$

$$\lim_{h \rightarrow 0^+} \frac{I(\theta, \theta+h)}{h^2} = \lim_{h \rightarrow 0^+} \frac{4 - \frac{8}{2-h}}{2h} = \lim_{h \rightarrow 0^+} \frac{8}{2(2-h)^2} = 1.$$

Para  $h > 0$ , se hace todo de forma análoga (las integrales se harán entre  $\theta+h$  e infinito) obteniéndose los mismos valores para ambos límites.

Por otra parte, también se verifica:

$$\begin{aligned} E \left[ \frac{\partial \log f(x, \theta)}{\partial \theta} \right] &= \int_0^{+\infty} \frac{\partial f(x, \theta)}{\partial \theta} dx = \frac{1}{2} \int_0^{+\infty} e^{(\theta-x)} (x-\theta)^2 dx - \int_0^{+\infty} e^{(\theta-x)} (x-\theta) dx = \\ &= 1 - \int_0^{+\infty} ze^{-z} dz = 1 - 1 = 0. \end{aligned}$$

De esta forma, la Información de Fisher será:

$$\begin{aligned} I(\theta) &= E \left[ \left( \frac{\partial \log f(x, \theta)}{\partial \theta} \right)^2 \right] = -E \left[ \frac{\partial^2 \log f(x, \theta)}{\partial \theta^2} \right] = \\ &= \int_0^{+\infty} \frac{2}{(x-\theta)^2} \frac{1}{2} e^{(\theta-x)} (x-\theta)^2 = \int_0^{+\infty} e^{-z} dz = 1. \end{aligned}$$

Este ejemplo nos hace ver que la condición que juega el papel fundamental sobre el orden de convergencia de  $I(\theta, \theta+h)$  es la posibilidad o no de intercambiar la derivada con la integral (es decir, que la esperanza de la derivada de  $\log f(x, \theta)$  sea cero o no).

Tras estas consideraciones, podemos adoptar la siguiente definición usada por Pitman, E.J. (1979).

**Definición:** Decimos que una familia de medidas de probabilidad (o de densidades) es regular en  $\theta_0$  si existe  $\lim_{h \rightarrow 0} \frac{I(\theta_0, \theta_0 + h)}{h^2}$ , es finito e igual a

$$I(\theta_0) = E_{\theta_0} \left[ \left( \frac{\partial \ell}{\partial \theta} \right)^2 \Bigg|_{\theta=\theta_0} \right].$$

Si la familia es regular en todo punto de un abierto  $N$ ,

decimos que la familia es regular en  $N$ .

Debemos aclarar que Pitman, E.J.(1979) trabaja con la medida de información que hemos llamado  $J(\theta_1, \theta_2)$  y entonces el límite anterior sale  $I(\theta_0)/4$ , aunque, como hemos dicho, lo importante realmente es la proporcionalidad con la Información de Fisher.

Puede verse también en Pitman, E.J. (1979) el siguiente teorema que da una serie de condiciones que son suficientes para asegurar la regularidad. Dichas condiciones son menos restrictivas que las impuestas en el apartado 1.

**Teorema:** Si en cada  $\theta$  perteneciente a un abierto  $N$  que contiene a  $\theta_0$  la función de densidad  $f$  del modelo tiene derivada respecto a  $\theta$  (salvo un conjunto de medida nula) y se verifican

$$\text{i) } \int \frac{\partial f(x, \theta)}{\partial \theta} = 0 = \frac{\partial}{\partial \theta} \int f(x, \theta) dx$$

$$\text{ii) } \frac{\partial}{\partial \theta} \int \sqrt{f(x, \theta) f(x, \theta_0)} dx = \int \frac{\partial \sqrt{f(x, \theta) f(x, \theta_0)}}{\partial \theta} dx = \int \frac{\frac{\partial f(x, \theta)}{\partial \theta} \sqrt{f(x, \theta_0)}}{2\sqrt{f(x, \theta)}} dx$$

$$\text{iii) } I(\theta) = \int \frac{\left( \frac{\partial f(x, \theta)}{\partial \theta} \right)^2}{f(x, \theta)} dx \text{ es una función continua de } \theta$$

entonces la familia es regular en  $\theta_0$ .



## 1.5 La inferencia condicional.

Los métodos frecuentistas o clásicos se justifican por sus propiedades en términos de repetición de muestras. Fisher pensó que los métodos de construcción de intervalos de confianza y de contrastación de hipótesis hacían excesivo énfasis en las propiedades muestrales en detrimento de la consideración de otras propiedades. En particular, pensó que los métodos con buenas propiedades muestrales no necesariamente usan toda la información disponible, en el sentido de que ignoran la existencia de los llamados subconjuntos relevantes, que fueron introducidos por Fisher y posteriormente formalizados por Buehler, R.J. (1959). Supongamos que  $C_\alpha(\bar{x})$  es un intervalo de confianza clásico a nivel  $1-\alpha$ , es decir,  $P[\theta \in C_\alpha(\bar{x}) | \theta] = 1-\alpha$ . Entonces, un subconjunto  $A(\bar{x}) \in \chi$  es un subconjunto relevante si para algún  $\varepsilon > 0$ , se tiene

$$P[\theta \in C_\alpha(\bar{x}) | A(\bar{x})] \geq 1-\alpha + \varepsilon$$

o bien

$$P[\theta \in C_\alpha(\bar{x}) | A(\bar{x})] \leq 1-\alpha - \varepsilon$$

para todos los valores del parámetro. En el primer caso se dirá que el subconjunto relevante es positivamente sesgado y en el segundo hablaremos de un subconjunto negativamente sesgado.

En presencia de subconjuntos relevantes, surge inmediatamente la siguiente cuestión; si el valor muestral obtenido resulta estar en  $A(\bar{x})$ , ¿cuál es el nivel de confianza del intervalo?. La respuesta a esta cuestión puede encontrarse en la siguiente cita de Goutis, C. and Casella, G. (1995):

*“El resultado final de un experimento es una inferencia, que habitualmente se hace después de obtener los datos (una inferencia “post-datos”). La teoría frecuentista clásica se desarrolla en torno a inferencias “pre-datos”, aquellas que pueden hacerse en la etapa del diseño del experimento, antes de haber obtenido los datos. Tales inferencias “pre-datos” a menudo no son razonables como inferencias “post-datos”, llevando a frecuencias que no concuerdan con las de la inferencia condicionada a los datos observados”<sup>4</sup>.*

Existen multitud de ejemplos en los que aparecen subconjuntos relevantes, con el inconveniente que ello supone para la construcción de intervalos de confianza clásicos, ya que el intervalo obtenido para una muestra concreta puede resultar absurdo.

**Ejemplo 1.6:**  $X_1, \dots, X_n$  i.i.d.  $\sim U(\theta - \frac{1}{2}, \theta + \frac{1}{2})$ .

En este modelo,

$$f(\bar{x} | \theta) = 1 \quad \text{si} \quad -\frac{1}{2} \leq x_{(1)} - \theta \leq x_{(n)} - \theta \leq \frac{1}{2}$$

Consideremos el estadístico  $M = \frac{X_{(1)} + X_{(n)}}{2}$ . Mediante un simple cambio de variable, puede obtenerse fácilmente su función de distribución, que viene dada por

---

<sup>4</sup> Goutis, C. & Casella, G. (1995, pág. 325). El texto original es: “The end result of an experiment is an inference, which is typically made after the data have been seen (a post-data inference). Classical frequency theory has evolved around pre-data inferences, those that can be made in the planning stages

$$F_M(m) = \begin{cases} 2^{n-1} \left( m + \frac{1}{2} - \theta \right)^n & \theta - \frac{1}{2} \leq m \leq \theta \\ 1 - 2^{n-1} \left( \theta + \frac{1}{2} - m \right)^n & \theta \leq m \leq \theta + \frac{1}{2} \end{cases}$$

Podemos obtener un intervalo de confianza central buscando a y b tales que

$$\frac{\alpha}{2} = F_M(a + \theta) = 2^{n-1} \left( a + \frac{1}{2} \right)^n \rightarrow a = \frac{\alpha^{1/n} - 1}{2}$$

y

$$1 - \frac{\alpha}{2} = F_M(b + \theta) = 1 - 2^{n-1} \left( \frac{1}{2} - b \right)^n \rightarrow b = \frac{1 - \alpha^{1/n}}{2}$$

Así,  $(M - b, M - a) \equiv \left( M - \frac{1 - \alpha^{1/n}}{2}, M + \frac{1 - \alpha^{1/n}}{2} \right)$  es un intervalo de confianza

de nivel  $1 - \alpha$ , ya que

$$\begin{aligned} P[\theta \in (M - b, M - a) | \theta] &= P[M - b \leq \theta \leq M - a | \theta] = P[-b \leq \theta - M \leq -a | \theta] = \\ &= P[-b - \theta \leq -M \leq -a - \theta | \theta] = P[a + \theta \leq M \leq b + \theta | \theta] = F_M(b + \theta) - F_M(a + \theta) = \end{aligned}$$

$$1 - \frac{\alpha}{2} - \frac{\alpha}{2} = 1 - \alpha$$

Supongamos que  $1 - \alpha = 0.9$ ,  $n = 10$ ,  $x_{(1)} = 0.01$  y  $x_{(10)} = 0.99$ , con lo que el intervalo para estos datos sería  $(0.397, 0.603)$ . Pero, para estos datos, estamos seguros de que  $0.49 \leq \theta \leq 0.51$ , con lo que el intervalo anterior contiene al parámetro con total seguridad. El problema entonces es que la afirmación “ $I(\bar{X})$  es un intervalo de

---

*of an experiment, before data are collected. Such pre-data inference are often not reasonable as post-*

---

confianza de nivel  $1-\alpha$ ” es correcta como afirmación “pre-datos”, pues es cierta cuando integramos en todo el espacio muestral. Para un determinado subconjunto del espacio muestral (al que pueden pertenecer los datos obtenidos), la afirmación no tiene por qué ser correcta.

En este ejemplo, es evidente que el recorrido muestral contiene información que no se ha usado al construir el intervalo de confianza. Si el recorrido es 1 (el valor máximo), entonces se tienen los extremos de la distribución y por tanto obtenemos que el punto medio muestral coincide con  $\theta$  con total seguridad. Así, si el recorrido  $R(\bar{x})$  es próximo a uno, el intervalo debería ser más corto que si dicho recorrido es próximo a 0. Es decir, en vez de integrar sobre todo el espacio muestral (y por tanto sobre todos los valores posibles de  $R$ ), deberíamos condicionar al valor observado  $r$  para obtener el intervalo e integrar sólo sobre el subconjunto del espacio muestral dado por  $\{\bar{x} \in \chi \text{ t.q. } R(\bar{x}) = r\}$ .

De hecho, en Goutis, C. and Casella, G., (1995), puede verse que en este ejemplo existen subconjuntos relevantes determinados por el recorrido. Concretamente,

1.  $P\left[\theta \in I(\bar{X}) \mid \theta, X_{(n)} - X_{(1)} \geq \varepsilon_1\right] > 1 - \alpha \quad \forall \varepsilon_1 > \frac{\alpha^{1/n} - \alpha}{1 - \alpha}$
2.  $P\left[\theta \in I(\bar{X}) \mid \theta, X_{(n)} - X_{(1)} \leq \varepsilon_2\right] < 1 - \alpha \quad \forall \varepsilon_2 < \frac{\alpha^{1/n} - \alpha}{1 - \alpha}$

En nuestro ejemplo, estamos en la situación 1, pues

$$\varepsilon_1 = x_{(10)} - x_{(1)} = 0'98 > \frac{(0'9)^{1/10} - 0'9}{1 - 0'9} = 0'8952$$

Como se señaló antes, existen multitud de ejemplos de modelos en los que existen subconjuntos relevantes (ver, por ejemplo, Robinson, G.K.(1975), Buehler, R.J. and Feddersen, A.P. (1963), Brown, L.P. (1967), etc.). Cabe destacar, por su

---

*data inferences, leaving a frequentist with no inference conditional on the observed data”.*

---

importancia, la existencia de ellos en el modelo Normal. En efecto, en el modelo  $N(\mu, \sigma^2)$  con ambos parámetros desconocidos, de todos es conocido el intervalo  $I_\alpha(\bar{X})$  para  $\mu$  de nivel  $1-\alpha$  basado en la distribución t-Student. Pues bien, en Brown L.P. (1967), puede verse que existen constantes estrictamente positivas  $K$  y  $\varepsilon$  tales que

$$P \left[ \mu \in I(\bar{X}) \mid \mu, \sigma, \frac{|\bar{X}|}{S} \leq K \right] \geq 1 - \alpha + \varepsilon$$

con lo cual el subconjunto  $A(\bar{x}) = \left\{ \bar{x} \in \chi \text{ t.q. } \frac{|\bar{x}|}{s} \leq K \right\}$  es un subconjunto relevante positivamente sesgado.

Tras estas consideraciones, surge la cuestión de cómo hacer inferencias con buenas propiedades condicionales, es decir, hemos de elegir el subconjunto del espacio muestral sobre el que condicionar a la hora de hacer la inferencia.

De los trabajos de Fisher se desprende que una función de los datos con distribución que no dependa de los parámetros desconocidos y que aporte información sobre la “dispersión” de la verosimilitud ha de ser vista como “*ancillary*” (auxiliar) a la inferencia, en el sentido de que afecta a la precisión de las inferencias pero no a su posición. (En el ejemplo  $U(\theta-1/2, \theta+1/2)$ , el recorrido, que es auxiliar, no afecta a la localización del intervalo sino a su amplitud). Fisher llamó a estas funciones *estadísticos auxiliares* y argumentó que la inferencia debe hacerse condicionando a estadísticos auxiliares. Estas ideas son las que dieron base al llamado *Principio de condicionar*, que puede enunciarse como sigue:

**Principio de condicionar:** *Si existe un estadístico auxiliar para un modelo, entonces toda inferencia debe hacerse condicionando al estadístico auxiliar.*

Es evidente que un claro inconveniente del principio de condicionar viene determinado por los problemas derivados de existencia (y búsqueda) y unicidad de estadísticos auxiliares a los que condicionar. Estos problemas están resueltos en familias de modelos importantes, aunque como problema genérico no está resuelto.

A continuación, vamos a aplicar el principio de condicionar al ejemplo anterior.

**Ejemplo 1.6 (continuación):**  $X_1, \dots, X_n$  i.i.d.  $\sim U(\theta - 1/2, \theta + 1/2)$ .

Siendo  $R = X_{(n)} - X_{(1)}$ , tenemos que  $R$  es un estadístico auxiliar, ya que

$$f_R(r) = n(n-1)r^{n-2}(1-r) \quad 0 \leq r \leq 1$$

La distribución de  $M = \frac{X_{(1)} + X_{(n)}}{2}$  condicionada a  $R=r$  es

$$f_{M|R}(m|r) = \frac{1}{1-r} \quad \theta - \frac{1}{2} + \frac{r}{2} \leq m \leq \theta + \frac{1}{2} - \frac{r}{2}$$

y su función de distribución es

$$F_{M|R}(m|r) = \frac{m - \theta + \frac{1}{2} - \frac{r}{2}}{1-r} \quad \theta - \frac{1}{2} + \frac{r}{2} \leq m \leq \theta + \frac{1}{2} - \frac{r}{2}$$

Por tanto,  $Z=M-\theta$  es una cantidad pivotal. La función de distribución de  $Z|R$  es

$$F_{Z|R}(z|r) = \frac{z + \frac{1}{2} - \frac{r}{2}}{1-r} \quad -\frac{1}{2} + \frac{r}{2} \leq z \leq \frac{1}{2} - \frac{r}{2}$$

Así, el percentil de orden  $0 \leq \gamma \leq 1$  viene dado por

$$F_{Z|R}(z_\gamma | r) = \frac{z_\gamma + \frac{1}{2} - \frac{r}{2}}{1-r} = \gamma \quad \Rightarrow \quad z_\gamma = \gamma(1-r) - \frac{1}{2} + \frac{r}{2}$$

El intervalo de confianza centrado de nivel  $1-\alpha$  queda por tanto:

$$\Rightarrow I(\bar{x} | r) = \left( m + (1-r)\frac{\alpha}{2} + \frac{r}{2} - \frac{1}{2}, m - (1-r)\frac{\alpha}{2} - \frac{r}{2} + \frac{1}{2} \right)$$

ya que

$$\begin{aligned} 1-\alpha &= P \left[ Z_{\alpha/2} \leq Z \leq Z_{1-\alpha/2} | r \right] = P \left[ \frac{\alpha}{2}(1-r) - \frac{1}{2} + \frac{r}{2} \leq M - \theta \leq \left(1 - \frac{\alpha}{2}\right)(1-r) - \frac{1}{2} + \frac{r}{2} \right] = \\ &= P \left[ M - \left(1 - \frac{\alpha}{2}\right)(1-r) + \frac{1}{2} - \frac{r}{2} \leq \theta \leq M - \frac{\alpha}{2}(1-r) + \frac{1}{2} - \frac{r}{2} \right] = \\ &= P \left[ M - 1 + r + \frac{\alpha}{2}(1-r) + \frac{1}{2} - \frac{r}{2} \leq \theta \leq M - \frac{\alpha}{2}(1-r) + \frac{1}{2} - \frac{r}{2} \right] = \\ &= P \left[ M + \frac{\alpha}{2}(1-r) - \frac{1}{2} + \frac{r}{2} \leq \theta \leq M - \frac{\alpha}{2}(1-r) + \frac{1}{2} - \frac{r}{2} \right] \end{aligned}$$

Un intervalo unilateral de probabilidad  $1-\alpha$  sería  $\left( m - \frac{1}{2} + \frac{r}{2}, m - \alpha(1-r) + \frac{1}{2} - \frac{r}{2} \right)$  ya que

$$\begin{aligned} 1-\alpha &= P[Z \geq z_\alpha | r] = P \left[ M - \theta \geq \alpha(1-r) - \frac{1}{2} + \frac{r}{2} \right] = \\ &= P \left[ -\theta \geq -M + \alpha(1-r) - \frac{1}{2} + \frac{r}{2} \right] = P \left[ \theta \leq M - \alpha(1-r) + \frac{1}{2} - \frac{r}{2} \right] \end{aligned}$$

Observemos que la longitud del intervalo es  $(1-\alpha)(1-r)$ , que toma el valor 0 cuando  $r=1$  y su valor máximo  $1-\alpha$  cuando  $r=0$ . Así, el valor del estadístico auxiliar  $R$  determina la precisión de la inferencia efectuada.

Una interpretación del principio de condicionar puede hacerse atendiendo a la misma idea que se aplica al utilizar un estadístico suficiente para hacer inferencia (ver Reid, N. (1995)). En efecto, siendo  $T$  un estadístico suficiente para un modelo generador de datos dado y una muestra aleatoria simple de tamaño  $n$ , podemos escribir:

$$f(\bar{x}, \theta) = f(\bar{x} | t, \theta) f_t(t, \theta)$$

y, puesto que  $T$  es suficiente, la distribución de  $\bar{X} | T = t$  no depende de  $\theta$ , con lo que podemos escribir

$$f(\bar{x}, \theta) = f(\bar{x} | t) f_T(t, \theta)$$

y de esta forma, concluimos que toda la información sobre el parámetro que hay en la muestra está contenida en la distribución marginal del estadístico suficiente  $T$ , con lo que podemos basar nuestra inferencia en dicho estadístico.

Ahora, si  $V$  es un estadístico auxiliar y no existe un estadístico suficiente de la misma dimensión que el parámetro, podemos escribir:

$$f(\bar{x}, \theta) = f(\bar{x} | v, \theta) f_v(v, \theta)$$

y, puesto que  $V$  es auxiliar, la distribución de  $V$  no depende de  $\theta$ , con lo que podemos escribir



$$f(\bar{x}, \theta) = f(\bar{x} | v, \theta) f_v(v)$$

y de esta forma, concluimos que toda la información sobre el parámetro que hay en la muestra está contenida en la distribución de la muestra condicionada al estadístico auxiliar, con lo que podemos basar nuestra inferencia en dicha distribución. De forma análoga, siendo  $(\hat{\theta}, V)$  suficiente, concluiremos que toda la información sobre el parámetro se encuentra en la distribución de  $\hat{\theta} | V$ , que es quizás la forma más habitual de aplicar el principio de condicionar. Indiquemos que, si el estadístico máximo-verosímil es suficiente (y completo), entonces será independiente de cualquier estadístico auxiliar, con lo que la inferencia condicionada y sin condicionar llevan a los mismos resultados.

Otro aspecto que refuerza la importancia del principio de condicionar es el de la **recuperación de la información**. A partir de las propiedades de la Información de Fisher y de la medida de información generalizada, vimos que dado un estadístico  $(T, S)$  suficiente donde  $S$  es auxiliar, se tenía:

1.  $E_S [I_{T|S}(\theta)] = I_{\bar{X}}(\theta)$
2.  $I_{\bar{X}}(\theta_1, \theta_2) = -8 \log(E_S [A_{T|S}(\theta_1, \theta_2)])$

Así, si en la situación anterior  $T$  no es suficiente, en media, recuperamos la información contenida en la muestra al condicionar al estadístico auxiliar. Resaltamos que estas igualdades **no** se verifican para cada valor dado del estadístico auxiliar, lo que en principio puede resultar sorprendente dado que la distribución del mismo no depende del parámetro. No obstante, podemos encontrar una explicación a este hecho si tenemos en cuenta que los valores del estadístico auxiliar, aunque no afectan a la localización, sí que determinan la precisión de las estimaciones. Además, ocurre que los valores del estadístico auxiliar que determinan estimaciones de mayor precisión son también los que proporcionan una mayor información. Nos encontramos así con un rango de valores del estadístico auxiliar para los que las distribuciones

---

condicionadas contienen menor información que la muestra y otro rango en el que se da la situación contraria, de forma que al tomar esperanza bajo todos los posibles valores obtenemos la información muestral.

Podemos pensar que estamos en una situación parecida a la planteada cuando nos preguntábamos por el significado de la expresión “el intervalo  $I(\bar{x})$  tiene un nivel de confianza  $1-\alpha$ ”. En efecto, cuando hablamos de la información que proporciona una muestra en un modelo, en cierto sentido estamos haciendo una afirmación “pre-datos” pues para el cálculo de la misma hacemos un promedio teniendo en cuenta todas las posibles muestras, cada una de las cuales puede aportar una información distinta. Cuando tenemos en cuenta el valor muestral a través de un estadístico auxiliar, es decir promediamos sólo para aquellas muestras en las que el valor del estadístico auxiliar es igual al observado, encontramos las diferencias.

La información de que disponemos al utilizar la inferencia condicionada depende pues del valor concreto del estadístico auxiliar y, en promedio, coincide con la información de la muestra. No obstante, hemos de tener en cuenta que si trabajamos con un estadístico no suficiente, perderemos información **seguro**. Además, también ocurre que a medida que aumentamos el tamaño muestral, en general, los estadísticos auxiliares tienden a concentrarse en torno a los valores que proporcionan una mayor precisión en las estimaciones de los parámetros ( y por tanto una mayor información).

A continuación, vamos a ilustrar la propiedad anterior con dos ejemplos.

**Ejemplo 1.7:** Consideremos  $X \sim M(n, \theta - \frac{1}{2}, 1 - \theta, \frac{1}{2})$  donde  $\frac{1}{2} < \theta < 1$ .

La función de verosimilitud para este modelo trinomial vendría dada por:

$$L(\theta; r_1, r_2, r_3) = \frac{n!}{r_1! r_2! r_3!} (\theta - \frac{1}{2})^{r_1} (1 - \theta)^{r_2} (\frac{1}{2})^{r_3}$$

Consideramos ahora el estadístico  $R_3$  definido por el número de veces que aparece el tercer valor de  $X$ . Sabemos que entonces  $R_3 \sim \text{Bi}(n, \frac{1}{2})$  con lo que será un estadístico auxiliar, ya que su distribución no depende del parámetro  $\theta$ .

Además, tenemos que  $(R_1, R_2) | R_3 \sim M\left(n - r_3, \frac{\theta - \frac{1}{2}}{\frac{1}{2}}, \frac{1 - \theta}{\frac{1}{2}}\right)$ , ya que:

$$f(r_1, r_2 | r_3) = \frac{\frac{n!}{r_1! r_2! r_3!} (\theta - \frac{1}{2})^{r_1} (1 - \theta)^{r_2} (\frac{1}{2})^{r_3}}{\frac{n!}{r_3! (n - r_3)!} (\frac{1}{2})^n} = \frac{(n - r_3)!}{r_1! r_2!} \left(\frac{\theta - \frac{1}{2}}{\frac{1}{2}}\right)^{r_1} \left(\frac{1 - \theta}{\frac{1}{2}}\right)^{r_2}$$

A continuación, vamos a calcular la cantidad de Información de Fisher correspondiente a la muestra.

$$\ell(\theta; r_1, r_2, r_3) = k + r_1 \log(\theta - \frac{1}{2}) + r_2 \log(1 - \theta)$$

$$\frac{\partial \ell}{\partial \theta} = \frac{r_1}{\theta - \frac{1}{2}} - \frac{r_2}{1 - \theta} \quad ; \quad \frac{\partial^2 \ell}{\partial \theta^2} = -\frac{r_1}{(\theta - \frac{1}{2})^2} - \frac{r_2}{(1 - \theta)^2}$$

$$I_{\bar{X}}(\theta) = \frac{n(\theta - \frac{1}{2})}{(\theta - \frac{1}{2})^2} + \frac{n(1 - \theta)}{(1 - \theta)^2} = \frac{n}{\theta - \frac{1}{2}} + \frac{n}{1 - \theta}$$

Veamos cuánto vale la información para el estadístico  $(R_1, R_2) | R_3$  y que, efectivamente, al tomar esperanza bajo la distribución de  $R_3$  obtenemos la información de la muestra.

$$\ell(\theta, r_1, r_2 | r_3) = k + r_1 \log(\theta - \frac{1}{2}) + r_2 \log(1 - \theta)$$

$$\frac{\partial^2 \ell}{\partial \theta^2} = -\frac{r_1}{(\theta - \frac{1}{2})^2} - \frac{r_2}{(1 - \theta)^2} \Rightarrow I_{(R_1, R_2) | R_3}(\theta) = \frac{(n - r_3) \frac{\theta - \frac{1}{2}}{\frac{1}{2}}}{(\theta - \frac{1}{2})^2} + \frac{(n - r_3) \left(\frac{1 - \theta}{\frac{1}{2}}\right)}{(1 - \theta)^2} =$$

$$= 2 \frac{n-r_3}{\theta - \frac{1}{2}} + 2 \frac{n-r_3}{1-\theta}$$

Como vemos, esta información depende de  $r_3$ . Al calcular la esperanza de esta cantidad bajo la distribución de  $R_3$  obtenemos:

$$E_{R_3} (I_{(R_1, R_2) | R_3}(\theta)) = 2 \frac{n - \frac{n}{2}}{\theta - \frac{1}{2}} + 2 \frac{n - \frac{n}{2}}{1 - \theta} = \frac{n}{\theta - \frac{1}{2}} + \frac{n}{1 - \theta} = I_{\bar{X}}(\theta).$$

**Ejemplo 1.2 (continuación):**

Consideremos  $X_1, \dots, X_n$  i.i.d.  $U(\theta - \frac{1}{2}, \theta + \frac{1}{2})$ .

En este caso, obtuvimos:

$$I(\theta_1, \theta_2) = \begin{cases} -8 \log(1 - |\theta_1 - \theta_2|) & \text{si } |\theta_1 - \theta_2| \leq 1 \\ +\infty & \text{si } |\theta_1 - \theta_2| > 1 \end{cases}$$

y por tanto,

$$I_{\bar{X}}(\theta_1, \theta_2) = \begin{cases} -8n \log(1 - |\theta_1 - \theta_2|) & \text{si } |\theta_1 - \theta_2| \leq 1 \\ +\infty & \text{si } |\theta_1 - \theta_2| > 1 \end{cases}$$

Vamos a considerar ahora los estadísticos  $T_1 = X_{(n)} - X_{(1)}$  y  $T_2 = (X_{(n)} + X_{(1)})/2$ , donde con  $X_{(i)}$  representamos la muestra ordenada. El estadístico  $T_1$  es auxiliar como se verá a continuación, mientras que  $T_2$  es un estimador máximo verosímil (que en este caso no es único, ya que la función de verosimilitud es plana entre  $X_{(1)}$  y  $X_{(n)}$ ).

A partir de la distribución conjunta de  $(X_{(1)}, X_{(n)})$  se obtiene que la distribución conjunta de  $(T_1, T_2)$  viene dada por:

$$f(t_1, t_2) = n(n-1)t_1^{n-2} \quad \text{si} \quad \begin{cases} \theta - \frac{1}{2} \leq \frac{2t_2 - t_1}{2} \leq \frac{2t_2 + t_1}{2} \leq \theta + \frac{1}{2} \\ 0 \leq t_1 \leq 1 \end{cases}$$

Así, podemos calcular la distribución marginal de  $T_1$  que será:

$$f(t_1) = \int_{\frac{t_1 + 2\theta - 1}{2}}^{\frac{t_1 + 2\theta - 1}{2}} n(n-1)t_1^{n-2} dt_2 = n(n-1)t_1^{n-2} \left( -\frac{t_1}{2} + \frac{2\theta - 1}{2} - \frac{t_1}{2} - \frac{2\theta - 1}{2} \right)$$

y por tanto,

$$f(t_1) = n(n-1)t_1^{n-2}(1-t_1) \quad \text{si} \quad 0 < t_1 < 1$$

de donde podemos deducir que  $T_1$  es auxiliar ya que su distribución no depende del parámetro desconocido.

La distribución de  $T_2|T_1$  vendrá dada por:

$$f(t_2 | t_1) = \frac{n(n-1)t_1^{n-2}}{n(n-1)t_1^{n-2}(1-t_1)} = \frac{1}{1-t_1} \quad \text{si} \quad \begin{cases} \theta - \frac{1-t_1}{2} \leq t_2 \leq \theta + \frac{1-t_1}{2} \\ 0 \leq t_1 \leq 1 \end{cases}$$

Vamos a calcular la información asociada a estas tres distribuciones y a comprobar directamente que se verifica la propiedad 3.

1. Sabemos que  $(X_{(1)}, X_{(n)})$  es suficiente por el criterio de factorización y así  $(T_1, T_2)$  que es una función biyectiva del mínimo y el máximo es suficiente, con lo que

$$I_{(T_1, T_2)}(\theta_1, \theta_2) = I_{\bar{X}}(\theta_1, \theta_2).$$

2.  $T_1$  es auxiliar y por tanto  $I_{T_1}(\theta_1, \theta_2) = 0$ .

3. A la vista de la función de densidad de  $T_2|T_1$  tenemos que

$$T_2 | T_1 \sim U\left(\theta - \frac{1-t_1}{2}, \theta + \frac{1-t_1}{2}\right)$$

con lo que, aprovechando el resultado anterior, se obtiene

$$I_{T_2|T_1}(\theta_1, \theta_2) = \begin{cases} -8 \log\left(1 - \frac{|\theta_1 - \theta_2|}{1-t_1}\right) & \text{si } |\theta_1 - \theta_2| \leq 1-t_1 \\ +\infty & \text{si } |\theta_1 - \theta_2| < 1-t_1 \end{cases}$$

Vemos así cómo esta información depende de  $t_1$ . Vamos a comprobar directamente que al tomar esperanza bajo la distribución de  $T_1$  como se indica en la propiedad 3 se obtiene la información proporcionada por la muestra.

$$A_{T_2|T_1}(\theta_1, \theta_2) = \begin{cases} 1 - \frac{|\theta_1 - \theta_2|}{1-t_1} & \text{si } |\theta_1 - \theta_2| < 1-t_1 \\ +\infty & \text{si } |\theta_1 - \theta_2| < 1-t_1 \end{cases}$$

$$\begin{aligned} E_{T_1} [A_{T_2|T_1}(\theta_1, \theta_2)] &= \int_0^{1-|\theta_1-\theta_2|} \left(1 - \frac{|\theta_1 - \theta_2|}{1-t_1}\right) n(n-1)t_1^{n-2}(1-t_1) dt_1 = \\ &= n(n-1) \int_0^{1-|\theta_1-\theta_2|} \left(\frac{1-t_1 - |\theta_1 - \theta_2|}{1-t_1}\right) t_1^{n-2}(1-t_1) dt_1 = \\ &= n(n-1) \int_0^{1-|\theta_1-\theta_2|} (1-t_1 - |\theta_1 - \theta_2|) t_1^{n-2} dt_1 = (1-|\theta_1 - \theta_2|)^n. \end{aligned}$$

De esta forma, tenemos

$$-8 \log E_{T_1} [A_{T_2|T_1}(\theta_1, \theta_2)] = -8n \log(1 - |\theta_1 - \theta_2|) = I_{(T_1, T_2)}(\theta_1, \theta_2) = I_{\bar{X}}(\theta_1, \theta_2).$$

Con objeto de ilustrar los resultados obtenidos, vamos a suponer que  $\theta = 0,5$  y a analizar la información correspondiente a los valores  $\theta_1 = 0'8$  y  $\theta_2 = 0'81$  que aportan la muestra, el estadístico no suficiente  $T_2$  y las distribuciones  $T_2 | T_1 = t_1$ .

Para ello, necesitamos la distribución marginal de  $T_2$ . Tras integrar la expresión de la función de densidad conjunta con respecto a  $t_1$  obtenemos que

$$f(t_2) = n(1 - 2|\theta - t_2|)^{n-1}, \quad \theta - \frac{1}{2} \leq t_2 \leq \theta + \frac{1}{2}$$

La figura 1.5 muestra  $I_{\bar{X}}(0.8, 0.81)$  e  $I_{T_2}(0.8, 0.81)$  así como  $I_{T_2|T_1=t_1}(0.8, 0.81)$  para todos los posibles valores  $t_1$  y para los tamaños muestrales  $n = 10$  y  $n = 50$ .

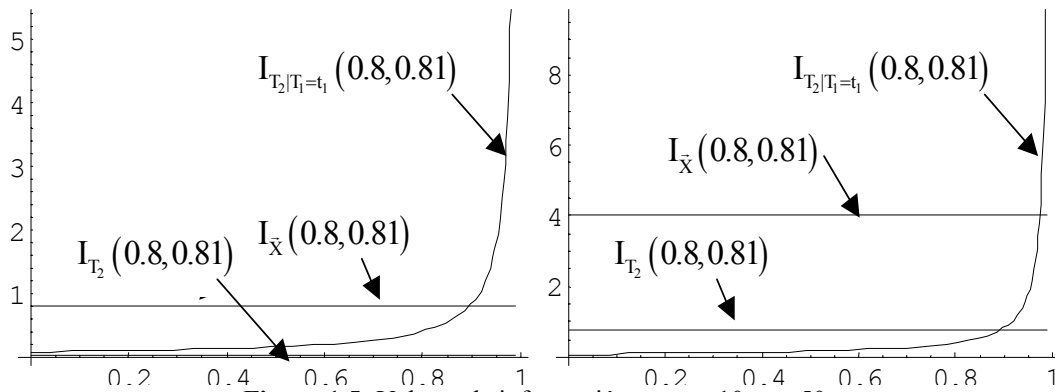


Figura 1.5: Valores de información para  $n=10$  y  $n=50$ .

Vemos cómo valores de  $t_1$  próximos a 1 significan disponer de mayor información y cómo el uso del estadístico  $T_2 | T_1 = t_1$  para nuestra inferencia es preferible al uso del estadístico no suficiente  $T_2$ .

A medida que aumenta el tamaño muestral  $n$ , aumentan  $I_{\bar{x}}(0.8, 0.81)$  e  $I_{T_2}(0.8, 0.81)$  pero también ocurre que la distribución de  $T_1$  se va concentrando en torno a 1, lo que hace muy probable que  $I_{T_2|T_1=t_1}(0.8, 0.81)$  sea elevada.

En la tabla 1.I hemos recogido para  $n=10, 20, 50$  y  $100$  los valores de  $I_{\bar{x}}(0.8, 0.81)$  e  $I_{T_2}(0.8, 0.81)$  así como  $I_{T_2|T_1=t_1}(0.8, 0.81)$  para una única muestra simulada del modelo. Observemos principalmente cómo al aumentar  $n$ , el valor de  $t_1$  tiende a aproximarse a 1.

	$I_{\bar{x}}(0.8, 0.81)$	$I_{T_2}(0.8, 0.81)$	$t_1$	$I_{T_2 T_1=t_1}(0.8, 0.81)$
$n=10$	0,8040	0,0385	0,7654	0,3484
$n=20$	1,6080	0,1436	0,9245	1,1367
$n=50$	4,0201	0,7656	0,9870	9,9840
$n=100$	8,0403	2,4820	0,9883	15,6576

**Tabla 1.I:**  $I_{\bar{x}}(0.8, 0.81)$ ,  $I_{T_2}(0.8, 0.81)$  e  $I_{T_2|T_1=t_1}(0.8, 0.81)$  en el modelo  $U(\theta - 1/2, \theta + 1/2)$ ,  $\theta = 1/2$



## **CAPÍTULO 2: DISTRIBUCIONES A PRIORI EN MODELOS REGULARES: LA REGLA DE JEFFREYS.**

En este capítulo hacemos una recapitulación de los resultados que hemos considerado más importantes concernientes a la obtención de una distribución a priori no informativa en el caso de modelos regulares, que es la situación más habitual en la literatura. Este fue el tema de un trabajo de investigación anterior titulado “La Regla de Jeffreys: un Estudio de su Significado y Propiedades” y en este sentido el presente capítulo constituye un resumen de los resultados en él presentados con un interés especial en mostrar los métodos de obtención de distribuciones a priori que han tenido una mayor aceptación.

En el caso de modelos regulares con un único parámetro desconocido existe acuerdo total en elegir la distribución a priori no informativa que se obtiene a través de la Regla de Jeffreys, dadas las buenas propiedades que presenta esta propuesta con respecto a distintos criterios, sobre todo por sus propiedades de invarianza y por las propiedades frecuentistas de los intervalos bayesianos que se obtienen. Cuando el modelo tiene más de un parámetro desconocido, la generalización directa de la regla mantiene la invarianza ante reparametrizaciones pero muestra serias deficiencias, lo que ha llevado a que sea más utilizada la llamada Regla de Jeffreys modificada, aplicable a modelos de localización-escala (y a modelos ortogonales).

En el capítulo siguiente, intentaremos generalizar los procedimientos más usados en el caso regular de forma que puedan ser aplicados también a modelos no regulares.

## 2.1 Invarianza ante reparametrizaciones.

En la inferencia en general, y en particular en la Bayesiana, una propiedad deseable es la invarianza de los métodos respecto a reparametrizaciones arbitrarias en la especificación del modelo generador de datos. Obsérvese que, por ejemplo, un simple cambio en las unidades de medida en el problema puede significar una reparametrización del modelo y entonces es deseable que esto no afecte a los resultados obtenidos mediante un procedimiento en concreto.

En el caso que nos ocupa, supongamos que una regla **R** nos permite obtener una distribución a priori  $p(\theta)$  para el parámetro de un modelo. Entonces para cualquier reparametrización biyectiva  $\phi = \phi(\theta)$  la regla proporcionará una distribución a priori  $q(\phi)$  que, de forma genérica, no guardará ninguna relación específica con  $p(\theta)$ . Pues bien, la regla **R** tendrá la propiedad de invarianza si  $q(\phi)$  puede ser obtenida aplicando a  $p(\theta)$  el cambio de variable  $\phi = \phi(\theta)$ . De esta forma, la propiedad de invarianza se tendrá cuando se verifique

$$q(\phi) \propto p(\theta) \left| \frac{\partial \theta}{\partial \phi} \right|$$

El esquema que sigue puede ayudar a comprender el significado de la propiedad de invarianza. Para una regla **R**, representamos:

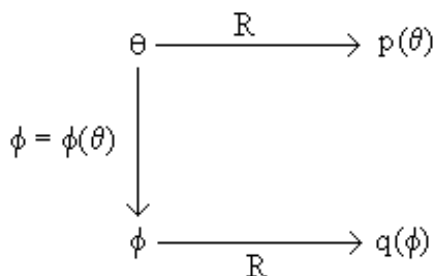


Figura 2.1.a: Propiedad de invarianza

Entonces, la regla  $\mathbf{R}$  tendrá la propiedad de invarianza si podemos "cerrar el esquema" y al aplicar el cambio de variable  $\phi = \phi(\theta)$  a la distribución  $p(\theta)$  llegamos a  $q(\phi)$ , obteniendo:

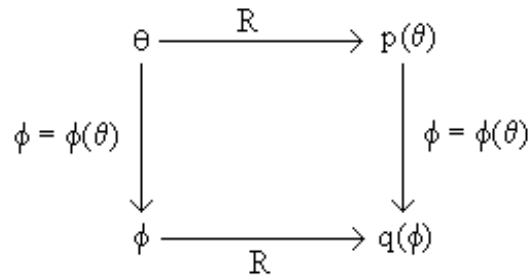


Figura 2.1.b: Propiedad de invarianza

Una vez explicado lo que significa la propiedad de invarianza ante reparametrizaciones, con respecto a la elección de una distribución a priori, podemos ver en qué consiste la regla propuesta por Sir Harold Jeffreys para obtener distribuciones a priori no informativas, centrándonos fundamentalmente en el caso uniparamétrico.

## 2.2 La regla de Jeffreys. Caso uniparamétrico.

Recordemos que si llamamos  $\ell(\theta|x) = \log L(\theta|x) = \log f(x|\theta)$ , entonces la cantidad de información de Fisher proporcionada por una observación sobre el parámetro  $\theta$  se definía como

$$I(\theta) = E \left[ \left( \frac{\partial \ell(\theta|x)}{\partial \theta} \right)^2 \right]$$

Dada una reparametrización biyectiva  $\phi = \phi(\theta)$ , se tiene

$$\frac{\partial \ell(\phi | x)}{\partial \phi} = \frac{\partial \ell(\theta | x)}{\partial \theta} \left( \frac{\partial \theta}{\partial \phi} \right)$$

y puesto que  $\frac{\partial \theta}{\partial \phi}$  no depende de  $x$ , se verifica

$$E \left[ \left( \frac{\partial \ell(\phi | x)}{\partial \phi} \right)^2 \right] = E \left[ \left( \frac{\partial \ell(\theta | x)}{\partial \theta} \right)^2 \right] \left( \frac{\partial \theta}{\partial \phi} \right)^2$$

es decir, que tenemos la relación:

$$I(\phi) = I(\theta) \left( \frac{\partial \theta}{\partial \phi} \right)^2$$

Por tanto, al tomar como distribución a priori

$$p(\theta) \propto \sqrt{I(\theta)}$$

tenemos que, efectivamente, se cumple la propiedad de invarianza, ya que entonces la regla de Jeffreys dice que la distribución para  $\phi$  es

$$q(\phi) \propto \sqrt{I(\phi)}$$

Y por tanto tenemos

$$q(\phi) \propto \sqrt{I(\phi)} = \sqrt{I(\theta) \left( \frac{\partial \theta}{\partial \phi} \right)^2} = \sqrt{I(\theta)} \left| \frac{\partial \theta}{\partial \phi} \right|$$

de donde vemos que, efectivamente, se cumple

$$q(\phi) \propto p(\theta) \left| \frac{\partial \theta}{\partial \phi} \right|$$

**Ejemplo 2.1:** Vamos a estudiar la media de la población Normal con  $\sigma$  conocida.

$$\ell(\mu | x) = -\frac{1}{2} \frac{(x - \mu)^2}{\sigma^2} + k \Rightarrow \frac{\partial^2 \ell}{\partial \mu^2} = -\frac{1}{\sigma^2}$$

y por tanto la distribución a priori no informativa será  $p(\mu) \propto k$ , es decir, una distribución uniforme.

**Ejemplo 2.2:** Estudiamos ahora la varianza en una distribución normal donde la media es conocida.

$$\ell(\sigma^2 | x) = \ell(\varphi | x) = -\frac{1}{2} \log \varphi - \frac{1}{2} \frac{(x - \theta)^2}{\varphi} + k$$
$$\frac{\partial^2 \ell}{\partial \varphi^2} = \frac{1}{2\varphi^2} - \frac{(x - \theta)^2}{\varphi^3}$$

Por tanto, tendremos que

$$I(\varphi) = -E \left[ \frac{\partial^2 \ell}{\partial \varphi^2} \right] = -\frac{1}{2\varphi^2} + \frac{E[(x - \theta)^2]}{\varphi^3} = -\frac{1}{2\varphi^2} + \frac{\varphi}{\varphi^3} = \frac{1}{2\varphi^2}$$

y de esta forma se obtiene como distribución a priori no informativa

$$p(\varphi) \propto \frac{1}{\varphi}$$

Notemos que hemos aplicado un cambio de variable, concretamente  $\varphi = \varphi(\sigma) = \sigma^2$ . Dada la invarianza de la regla de Jeffreys, podemos obtener la distribución a priori no informativa para  $\sigma$  sin más que aplicar el correspondiente cambio de variable a la distribución obtenida para  $\varphi$ , obteniéndose

$$p(\sigma) \propto q(\varphi) \left| \frac{\partial \varphi}{\partial \sigma} \right| = \frac{1}{\varphi} \left| \frac{\partial \varphi}{\partial \sigma} \right| = \frac{1}{\sigma^2} 2\sigma = \frac{2}{\sigma} \propto \frac{1}{\sigma}$$

**Ejemplo 2.3:** Vamos a estudiar ahora la población  $\gamma(p, \theta)$  donde  $p$  es conocido.

$$\ell(\theta | x) = p \log \theta - \theta x + k \Rightarrow \frac{\partial^2 \ell}{\partial \theta^2} = -\frac{p}{\theta^2} \Rightarrow I(\theta) = \frac{p}{\theta^2}$$

De esta forma, la distribución a priori no informativa es

$$p(\theta) \propto \frac{1}{\theta}$$

Hemos visto directamente a partir de la definición cómo la distribución a priori de Jeffreys es invariante ante reparametrizaciones, dado que si  $\phi = \phi(\theta)$  es una reparametrización biyectiva entonces  $I(\phi) = I(\theta) \left( \frac{\partial \theta}{\partial \phi} \right)^2$ . Sin embargo, es muy importante resaltar que el argumento seguido por Jeffreys, H. (1946) para la obtención de su propuesta parte de la consideración de las medidas de información que en la sección 1.3 hemos denotado como  $J$  y  $\tilde{J}$  y de la relación de las mismas con la Información de Fisher. Jeffreys hace constar que estas medidas son invariantes pues pueden definirse en términos de las medidas de probabilidad involucradas, sin tener en cuenta la posible parametrización elegida para representar las funciones de densidad. Además, señala que estas medidas son de segundo orden en la diferencia de los parámetros de las leyes cuando esta diferencia es pequeña (como ya habíamos

señalado y demostrado en el capítulo 1). Esta propiedad es cierta tanto para el caso uni como multiparamétrico y lleva a que  $J$  y  $\tilde{J}$  sean el cuadrado de un elemento de distancia. En concreto, aunque con otras notaciones, para  $\theta \in \mathbb{R}^m$ , el resultado que obtiene es

$$\tilde{J}(\theta, \theta + h) \approx \sum_{i,k} (I(\theta))_{ik} h_i h_k$$

lo que le permite concluir, dada la invarianza de  $\tilde{J}$ , que  $I(\phi) = I(\theta) \left( (\partial\theta) / (\partial\phi) \right)^2$  y esto le lleva a definir su regla. Observemos que este resultado se tendría para las medidas  $J$ ,  $\tilde{J}$  e  $I$  ya que todas ellas son equivalentes cuando  $h$  es pequeño. Observemos además, que en el caso uniparamétrico, la relación sería  $\tilde{J}(\theta) \approx I(\theta)h^2$ , que es el resultado visto en la sección 1.3.2

En este mismo artículo, Jeffreys destaca que  $\tilde{J}$  no siempre es de orden 2, pues, por ejemplo, en el modelo  $U(0, \theta)$  se verifica que  $\tilde{J}$  es de orden 1 y por ello, en este caso, no sería aplicable su regla general. Esta observación hecha por Jeffreys va a resultar crucial en el capítulo siguiente, pues en ella está la esencia de la propuesta que haremos para obtener distribuciones a priori no informativas.

Una de las críticas más extendida a la regla de Jeffreys se basa en que la información de Fisher es una característica inherente al modelo generador de datos y la distribución a priori se define en base a ella. No obstante, si, como en Box, G.E. and Tiao, G.C. (1973), admitimos que ha de usarse una distribución no informativa cuando conocemos poco a priori sobre el parámetro en relación a lo que se espera que aporte el experimento, la dependencia anterior podría justificarse ya que distintos experimentos se reflejarían básicamente en distintos modelos generadores de datos.

La dependencia de la distribución a priori de Jeffreys del experimento en cuestión se observa bastante bien en el siguiente ejemplo.

**Ejemplo 2.4:** Supongamos que estamos interesados en estudiar la proporción de éxitos  $\theta$  en un modelo de Bernoulli. Planteamos para ello dos experimentos diferentes:

a) Repetir el experimento  $n$  veces y contar el número de éxitos.

b) Contar el número de experimentos necesarios para obtener  $r$  éxitos.

**Caso a)** En este caso tendremos una observación  $x$  de un modelo binomial, y por tanto:

$$L(\theta | x) = \binom{n}{x} \theta^x (1-\theta)^{n-x} \Rightarrow \ell(\theta | x) = x \log \theta + (n-x) \log(1-\theta) + k$$

$$\frac{\partial^2 \ell}{\partial \theta^2} = -\frac{x}{\theta^2} - \frac{n-x}{(1-\theta)^2} \Rightarrow I(\theta) = \frac{E[X]}{\theta^2} + \frac{n-E[X]}{(1-\theta)^2} = \frac{n\theta}{\theta^2} + \frac{n-n\theta}{(1-\theta)^2} = \frac{n}{\theta(1-\theta)}$$

Vemos de esta forma que la distribución a priori es

$$p(\theta) \propto \sqrt{\frac{1}{\theta(1-\theta)}} = \theta^{-1/2} (1-\theta)^{-1/2}$$

**Caso b)** Ahora tendremos una observación  $x$  de un modelo binomial negativo, y por tanto:

$$L(\theta | x) = \binom{x-1}{r-1} \theta^r (1-\theta)^{x-r} \quad x = r, r+1, \dots$$

$$\ell(\theta | x) = r \log \theta + (x-r) \log(1-\theta) + k$$

$$\frac{\partial^2 \ell}{\partial \theta^2} = -\frac{r}{\theta^2} - \frac{(x-r)}{(1-\theta)^2} \Rightarrow$$



$$\Rightarrow I(\theta) = \frac{r}{\theta^2} + \frac{E[X] - r}{(1-\theta)^2} = \frac{r}{\theta^2} + \frac{r/\theta - r}{(1-\theta)^2} = r \left[ \frac{1}{\theta^2} + \frac{1}{\theta(1-\theta)} \right] \Rightarrow$$

$$I(\theta) = \frac{r}{\theta^2(1-\theta)}$$

Vemos por tanto que ahora la distribución a priori para el parámetro sería

$$p(\theta) \propto \frac{1}{\theta\sqrt{1-\theta}}$$

Supongamos, por fijar ideas, que en el caso **a)**  $n=20$  y que se obtienen  $x=15$  éxitos y que en el caso **b)**  $r=15$  y resulta que necesitamos  $x'=20$  experimentos. Al ser diferentes las distribuciones a priori y ser el núcleo de las verosimilitudes el mismo (pues estamos suponiendo que resulta  $r=x$ ,  $n=x'$ ) se obtienen distintas distribuciones a posteriori, y esto llevará a estimaciones distintas, intervalos distintos,...

Concretamente se tiene:

**Caso a)**

$$p(\theta | x) \propto L(\theta | x)p(\theta) \propto \theta^x (1-\theta)^{n-x} \theta^{-1/2} (1-\theta)^{-1/2} = \theta^{x-1/2} (1-\theta)^{n-x-1/2} \Rightarrow$$

$$\Rightarrow \theta | x \sim \text{Be}(x+1/2, n-x+1/2)$$

Siendo  $n=20$  y  $x=15$ , tenemos que  $\theta | x \sim \text{Be}(15.5, 5.5)$  y así la media de la distribución (que sería la estimación para una función de pérdida cuadrática) es 0.7380952.

**Caso b)**

$$p(\theta | x') \propto L(\theta | x')p(\theta) \propto \theta^r (1-\theta)^{x'-r} \theta^{-1/2} (1-\theta)^{-1/2} = \theta^{r-1/2} (1-\theta)^{x'-r-1/2} \Rightarrow$$

$$\Rightarrow \theta | x' \sim \text{Be}(r, x'-r+1/2)$$

Para  $r=15$  y  $x'=20$ , la distribución es  $\theta|x' \sim \text{Be}(15,5.5)$  y así la media de la distribución es 0.7317073.

Con este ejemplo se ve claramente la dependencia señalada y que es considerada como inaceptable por muchos autores. Incluso aun siendo las diferencias pequeñas, éstas existen; el caso **b)** añade una información que no se tiene en **a)** y es que el último experimento es un éxito y esto hace cambiar las distribuciones a priori y consecuentemente las distribuciones a posteriori.

Una explicación clara de que en el modelo Binomial Negativo se introduce una información adicional relevante puede obtenerse a partir del trabajo de Ye, K. (1993). Este autor propone que, cuando en un experimento la regla de parada depende del parámetro de interés, la distribución a priori correspondiente debe incorporar la información que proporciona la regla de parada. Concretando más, el resultado es el que sigue.

Sean  $X_1, X_2, X_3, \dots$  independientes e idénticamente distribuidas con función de densidad  $f(x|\theta)$  regular, donde  $x \in \mathbb{R}^k, \theta \in \mathbb{R}^p$ . Supongamos que tenemos una regla de parada  $\tau$  consistente en  $\tau_0, \tau_1(X^1), \tau_2(X^2), \dots$  donde  $X^n = (X_1, X_2, \dots, X_n)$  y  $\tau_i(x^i)$  es la probabilidad de parar tras observar  $x^i$ . Sea  $N$  el número de observaciones hechas antes de parar.

Entonces, siendo  $\pi(\theta)$  la distribución a priori de Jeffreys de  $\theta$  **para**  $x_1$ , la distribución de Jeffreys con regla de parada conocida es:

$$\pi^*(\theta) \propto [E_\theta(N)]^{p/2} \pi(\theta)$$

Aplicando esto al ejemplo anterior, en ambos casos tenemos que  $X_1 \sim \text{Be}(\theta)$ , con lo que  $\pi(\theta) \propto \theta^{-1/2} (1-\theta)^{-1/2}$ . Sin embargo, en el experimento Binomial  $N$  es una

variable aleatoria degenerada pues conocemos de antemano el tamaño muestral que es el que determina el criterio de parada mientras que en el Binomial Negativo N es una variable aleatoria propiamente dicha.

En concreto, con las notaciones seguidas hasta ahora se tiene  $E[N] = r/\theta$  y así  $E[N] \propto \theta^{-1}$  con lo que se obtiene

$$\pi^*(\theta) \propto \theta^{-1/2} \theta^{-1/2} (1-\theta)^{-1/2} \propto \theta^{-1} (1-\theta)^{-1/2}$$

que, como puede observarse, es el resultado obtenido anteriormente al calcular la información de Fisher para una observación de una variable Binomial Negativa.

Así, ambos modelos surgen de la realización sucesiva de variables Bernoulli pero la forma de detener el experimento es distinta y esto es lo que hace que las distribuciones a priori sean diferentes. Al considerar, como hicimos antes, una observación de un modelo Binomial y una observación de un modelo Binomial Negativo, esa información adicional sobre el criterio de parada ya se encontraba incorporada en el propio modelo, con lo que quedan explicadas las diferencias obtenidas.

### **2.2.1 Interpretación geométrica de la regla de Jeffreys.**

En Box, G.E. and Tiao, G.C. (1973), se estudia el tema de la búsqueda de distribuciones a priori no informativas y se propone una solución basada fundamentalmente en una idea geométrica. Al coincidir esta distribución con la regla de Jeffreys, obtenemos una interpretación geométrica para la misma.

- **Verosimilitud trasladada por los datos.**

Se plantea que una distribución es no informativa para un parámetro  $\theta$  si es uniforme para  $\phi = \phi(\theta)$ , donde  $\phi$  es una reparametrización biyectiva para la que la verosimilitud es "trasladada por los datos", es decir, la forma de  $L(\theta | x)$  es siempre la misma para cualesquiera que sean los datos y el único cambio que producen los distintos conjuntos de datos es una traslación de la verosimilitud a lo largo del eje de  $\phi$ . Obviamente, si  $p(\phi) \propto k$ , obtendremos  $p(\theta)$  aplicando el correspondiente cambio de variable y entonces

$$p(\theta) \propto \left| \frac{\partial \phi}{\partial \theta} \right|$$

Para entender el por qué de este método, hay que señalar que los autores dicen que se tomará una distribución a priori no informativa cuando "se conoce poco a priori en relación con lo que se espera que nos aporten los datos". Así, cuando la verosimilitud  $L(\theta | x)$  es trasladada por los datos, queda completamente determinada a priori excepto por su localización, que dependerá a su vez de los datos (generalmente a través de un estadístico). De esta forma, a priori se está en un estado de indiferencia que hará que se tenga la misma "predisposición" a aceptar un valor de  $\phi$  u otro y este estado, según los autores, puede expresarse asumiendo que la distribución de  $\phi$  es uniforme.

Antes de ver la definición de verosimilitud trasladada por los datos vamos a ver algún ejemplo.

**Ejemplo 2.5:** Consideremos  $X_1, \dots, X_n$  muestra aleatoria simple de una distribución  $N(\theta, \sigma)$ , donde  $\sigma$  es conocida. En este caso la verosimilitud es:

$$L(\theta | x) \propto \exp\left\{-\frac{n}{2\sigma^2}(\theta - \bar{x})^2\right\}$$

Así, la verosimilitud estandarizada como función de  $\theta$  es una normal de media la media muestral y desviación típica  $\sigma n^{-1/2}$ , que es constante. Así, las distintas muestras posibles sólo determinan la localización de la curva en el eje  $\theta$ . La situación explicada se representa en la figura que sigue para  $n=10$  y para diversos valores de la media muestral (concretamente para  $\bar{x} = 3.5$ ,  $\bar{x} = 5$ ,  $\bar{x} = 6.5$  y  $\bar{x} = 8$ ).

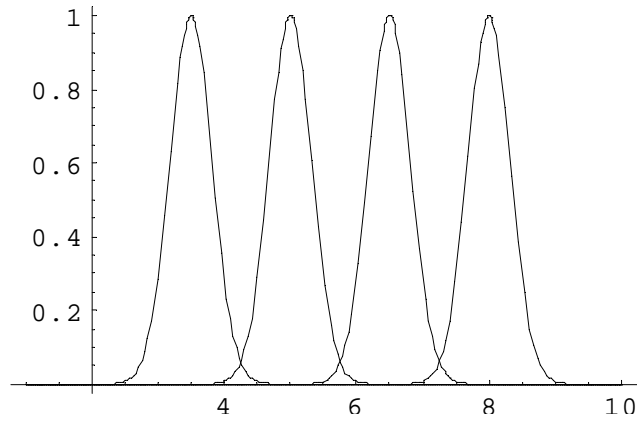


Figura 2.2: Verosimilitudes de  $N(\theta, \sigma^2)$ ,  $\sigma^2$  conocido

**Ejemplo 2.6:** Consideremos ahora  $X_1, \dots, X_n$  m.a.s. de una distribución  $N(\theta, \sigma)$  donde  $\theta$  se supone conocido. La verosimilitud es:

$$L(\sigma | x) \propto \sigma^{-n} \exp\left\{-\frac{nS^2}{2\sigma^2}\right\} \quad \text{con} \quad S^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \theta)^2$$

En este caso, como se aprecia en la figura adjunta efectuada para  $n=2$  y para  $S^2=1, S^2=2, S^2=3$  y  $S^2=5$ , la verosimilitud expresada en función de  $\sigma$  no es trasladada por los datos, y por tanto la distribución no informativa no puede tomarse uniforme en  $\sigma$ .

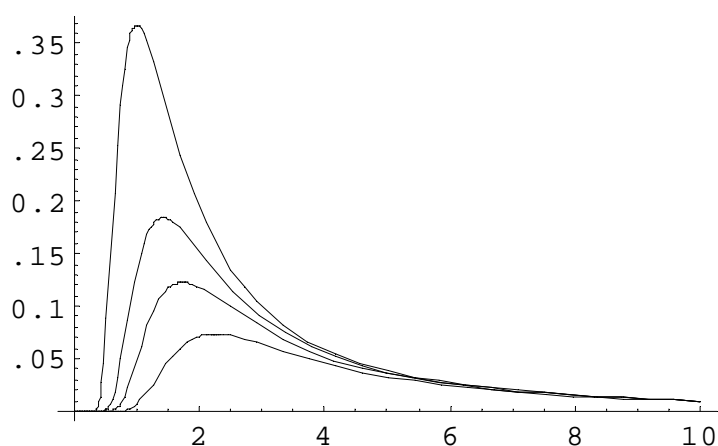


Figura 2.3: Verosimilitudes de  $N(\theta, \sigma^2)$ ,  $\sigma^2$  conocido.

Ahora bien, si tomamos  $\phi(\sigma) = \log \sigma$ , la verosimilitud puede reescribirse como:

$$L((\log \sigma) | x) \propto \exp \left\{ -n(\log \sigma - \log S) - \frac{n}{2} \exp[-2(\log \sigma - \log S)] \right\}$$

y por tanto, la verosimilitud en término de  $\phi$  queda como:

$$L(\phi | x) \propto \exp \left\{ -n(\phi - \log S) - \frac{n}{2} \exp[-2(\phi - \log S)] \right\}$$

La representación de  $L(\phi | x)$  para  $n=2$  y para distintos valores de  $\log S$  (concretamente  $\log(1)$ ,  $\log(2)$ ,  $\log(3)$  y  $\log(5)$ ) es:

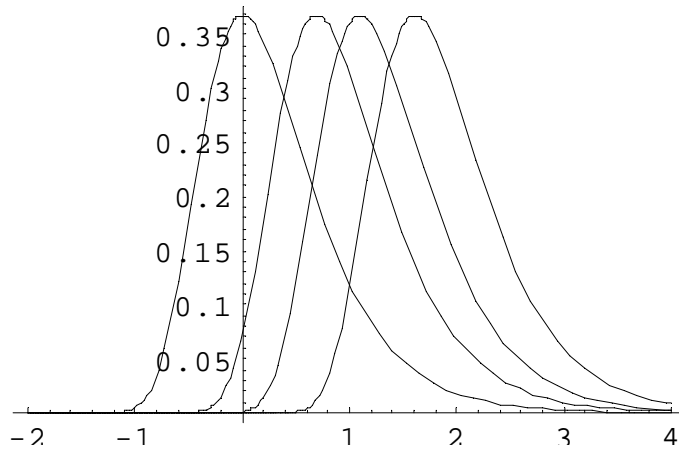


Figura 2.4: Verosimilitudes de  $N(\theta, \phi)$ ,  $\theta$  conocido,  $\phi(\sigma) = \log(\sigma)$

Como puede apreciarse, la verosimilitud es trasladada por los datos en este nuevo parámetro. Así, la distribución no informativa será uniforme en  $\phi = \log \sigma$ , es decir  $p(\phi) \propto k$  y por tanto

$$p(\sigma) \propto \left| \frac{\partial \log \sigma}{\partial \sigma} \right| = \frac{1}{\sigma}$$

Tras los ejemplos, pasamos a ver cuándo una verosimilitud es trasladada por los datos.

**Definición:** La verosimilitud  $L(\phi | x)$  es trasladada por los datos si puede expresarse de la forma

$$L(\phi | x) = g[\phi - f(x)]$$

donde  $g(y)$  es una función conocida independiente de los datos y  $f(x)$  es una función sólo de los datos (es decir,  $f(x)$  es un estadístico que además será suficiente sin más que aplicar el criterio de factorización).

Vamos a ver que en los ejemplos comentados se daba esta situación.

**Ejemplo 2.5: (continuación)**

$$L(\theta | x) = \left( \frac{1}{\sqrt{2\pi\sigma}} \right)^n \exp \left\{ -\frac{n}{2\sigma^2} (\theta - \bar{x})^2 \right\}$$

En este caso el criterio se cumple tomando:

$$\begin{cases} \phi(\theta) = \theta \\ f(x) = \bar{x} \\ g(y) = \left( \frac{1}{\sqrt{2\pi\sigma}} \right)^n \exp \left\{ -\frac{n}{2\sigma^2} y^2 \right\} \end{cases}$$

**Ejemplo 2.6: (continuación)**

$$L(\sigma | x) = \left( \frac{1}{2\pi} \right)^{n/2} \exp \left\{ -n(\log \sigma - \log S) - \frac{n}{2} \exp[-2(\log \sigma - \log S)] \right\}$$

El criterio se verifica para:

$$\begin{cases} \phi(\sigma) = \log \sigma \\ f(x) = \log S \\ g(y) = (2\pi)^{-n/2} \exp \left\{ -ny - \frac{n}{2} \exp(-2y) \right\} \end{cases}$$

• **Verosimilitud casi trasladada por los datos.**

Como parece obvio, no resultará fácil en general obtener una reparametrización para la que la verosimilitud sea trasladada por los datos. De hecho,

---



la condición es tan restrictiva que en el caso de la recta real y variables continuas sólo las distribuciones Normal y Gamma la cumplen (ver por ejemplo Kass, R.E. and Wasserman, L. (1995)). Este resultado puede deducirse también del artículo de Lindley, D.V. (1958). De esta forma se plantea un siguiente paso consistente en estudiar cuándo una verosimilitud es "aproximadamente" trasladada por los datos.

Es fundamental señalar que, en este punto, se parte de que la distribución cumple ciertas condiciones de regularidad y de que  $n$  es suficientemente grande. Bajo estas condiciones, la verosimilitud es aproximadamente normal y entonces:

$$\log L(\theta | x) = \ell(\theta | x) \approx \ell(\hat{\theta} | x) - \frac{n}{2}(\theta - \hat{\theta})^2 \left( -\frac{1}{n} \frac{\partial^2 \ell}{\partial \theta^2} \right)_{\hat{\theta}}$$

Vamos a adoptar la notación:

$$J(\hat{\theta}) = \left( -\frac{1}{n} \frac{\partial^2 \ell}{\partial \theta^2} \right)_{\hat{\theta}}$$

Si  $J(\hat{\theta})$  no depende de los datos  $x$ , entonces la distribución a priori no informativa se elegirá uniforme en la reparametrización  $\phi(\theta)$  que cumpla

$$\frac{\partial \phi}{\partial \theta} \propto \sqrt{J(\theta)}$$

lo que provocará que  $J(\hat{\theta})$  sea una constante independiente de  $\hat{\theta}$ . Así,

$$p(\theta) \propto \left| \frac{\partial \phi}{\partial \theta} \right| \propto \sqrt{J(\theta)}$$

Si  $J(\hat{\theta})$  depende de los datos, entonces básicamente lo que se hace es repetir el razonamiento pero tomando antes esperanza bajo la distribución de  $x|\theta$ , con lo que

$$I(\theta) = -E_{x|\theta} \left[ \frac{\partial^2 \ell}{\partial \theta^2} \right]$$

no depende de  $x$  y repetimos entonces el razonamiento anterior, con lo que tenemos que la distribución a priori no informativa resultará ser

$$p(\theta) \propto \sqrt{I(\theta)}$$

que, como puede observarse, es la distribución propuesta por Jeffreys.

Es importante señalar que, trivialmente, si  $J(\hat{\theta})$  no depende de los datos, entonces  $J(\hat{\theta}) = I(\hat{\theta})$ . Este resultado lo que significa es que, si existe una reparametrización  $\phi = \phi(\theta)$  para la que la verosimilitud es trasladada por los datos y tomamos una distribución a priori uniforme en  $\phi$ , al hacer el correspondiente cambio de variable para  $\theta$  obtendremos la distribución de Jeffreys.

Obsérvese que la idea que se está usando es que al ser  $J(\hat{\theta})$  una constante, tendremos:

$$\ell(\phi | x) \approx \ell(\hat{\phi} | x) - \frac{n}{2} (\phi - \hat{\phi})^2 k$$

con lo que los datos sólo actúan a través de  $\hat{\phi}$  y lo que hacen es determinar la localización de la verosimilitud.

**Ejemplo 2.7:** Sea  $X_1, \dots, X_n$  una m.a.s. de una población  $Be(\pi)$ . En este caso, llamando  $y = \sum x_i$  la verosimilitud es:

$$L(\pi | y) = \prod \pi^{x_i} (1-\pi)^{n-x_i} = \pi^y (1-\pi)^{n-y}$$

Por tanto, el logaritmo de la verosimilitud es:

$$\ell(\pi | y) = y \log \pi + (n-y) \log(1-\pi)$$

De esta expresión, deducimos las siguientes igualdades:

$$\begin{aligned} \frac{\partial^2 \ell}{\partial \pi^2} &= -\frac{y}{\pi^2} - \frac{n-y}{(1-\pi)^2} \\ E\left[\frac{\partial^2 \ell}{\partial \pi^2}\right] &= -n\left(\frac{1}{\pi} + \frac{1}{1-\pi}\right) \\ E\left[-\frac{1}{n} \frac{\partial^2 \ell}{\partial \pi^2}\right]_{\hat{\pi}} &= \frac{1}{\hat{\pi}} + \frac{1}{1-\hat{\pi}} = \frac{1}{\hat{\pi}(1-\hat{\pi})} \end{aligned}$$

De todo ello, podemos obtener la distribución a priori no informativa, que será

$$p(\pi) \propto \frac{1}{\sqrt{\pi(1-\pi)}}$$

- **Generalización del concepto de Box-Tiao.**

En Kass, R.E. (1990) puede verse una generalización del concepto de traslación dado por Box-Tiao. Recordemos que la exigencia de que la verosimilitud sea exactamente trasladada por los datos significa que la verosimilitud pueda escribirse en la forma:

$$L(\phi | y) = L_y(\phi) = f(\phi - t(y))$$

donde  $f(\cdot)$  no depende de  $y$  (es decir,  $t(y)$  es un estadístico suficiente).

Esta ecuación es equivalente a decir que:

$$L_y(\phi) = L_{y^*}(\phi^*) \quad \text{con} \quad \phi^* = \phi + \{t(y^*) - t(y)\}$$

es decir, que podemos definir el grupo de traslación en  $\phi$  y en la imagen de  $t(\cdot)$  de forma que la función de verosimilitud es invariante bajo su acción. La idea es entonces relajar las condiciones, exigiendo sólo que  $f(\cdot)$  dependa de los datos a través de un estadístico auxiliar  $\mathbf{a}$ . Así, lo que se pide es que:

$$L_y(\phi) = f_{\mathbf{a}}(\phi - t(y))$$

donde  $f_{\mathbf{a}}(\cdot)$  depende de los datos sólo a través de  $\mathbf{a}$ , siendo  $\mathbf{a}$  un estadístico auxiliar tal que  $(\hat{\phi}, \mathbf{a})$  son conjuntamente suficientes.

Con esta exigencia sobre  $\mathbf{a}$ , puede verse fácilmente que se tendrá:

$$\frac{L_{(\hat{\phi}, \mathbf{a})}(\phi)}{L_{(\hat{\phi}, \mathbf{a})}(\hat{\phi})} = \frac{L_{(\hat{\phi}^*, \mathbf{a})}(\phi^*)}{L_{(\hat{\phi}^*, \mathbf{a})}(\hat{\phi}^*)} \quad \text{para} \quad \phi^* = \phi + (\hat{\phi}^* - \hat{\phi})$$

y así, para dos conjuntos de datos  $y$  e  $y^*$  tales que  $\mathbf{a}(y) = \mathbf{a}(y^*)$  las verosimilitudes  $L_y(\phi)$  y  $L_{y^*}(\phi)$  serán idénticas salvo la traslación  $\hat{\phi} \rightarrow \hat{\phi}^*$ .

### 2.2.2 Coincidencia de resultados con la inferencia fiducial y clásica.

Una de las propiedades más interesantes de la regla de Jeffreys es la similitud de los resultados bayesianos de nuestra inferencia obtenidos con ella y los resultados proporcionados por la inferencia fiducial y clásica.

Un camino bastante lógico para buscar una distribución a priori no informativa puede ser "dejar que los datos hablen por sí solos", es decir, intentar que los resultados obtenidos con la distribución a priori sean los mismos que obtenemos cuando no utilizamos información a priori. Bajo esta idea, vamos a ver que la distribución obtenida a partir de la regla de Jeffreys muestra un excelente comportamiento en el caso uniparamétrico.

El primer ejemplo que llamó la atención en este sentido fue el del caso Normal con varianza conocida (que podemos suponer sin ninguna restricción que es unitaria). Llamando  $\theta$  a la media desconocida de esta población, para una muestra de tamaño  $n$ , el conjunto  $A_x = \left[ \bar{X} - \frac{Z_{\alpha/2}}{\sqrt{n}}, \bar{X} + \frac{Z_{\alpha/2}}{\sqrt{n}} \right]$  es un intervalo de confianza clásico de nivel  $1-\alpha$  y, si utilizamos como distribución a priori la distribución de Jeffreys (es decir,  $p(\theta) \propto \text{cte.}$ ), es también un intervalo que tiene probabilidad a posteriori  $1-\alpha$ . Puede verse también que la igualdad de resultados se obtiene usando la regla de Jeffreys si suponemos que la media es conocida y que el parámetro desconocido es la desviación típica.

Jeffreys, H. (1961) señaló este acuerdo entre sus métodos y los de Fisher en muchos ejemplos, aunque sin entrar a estudiar las características de los problemas en que se daba la coincidencia.

Antes de ver las propiedades de la distribución de Jeffreys al respecto, es interesante hacer constar que esta idea por sí sola da origen a un método de elección de distribuciones a priori, consistente en tomar aquella distribución que hace que los

---

intervalos bayesianos de probabilidad  $1-\alpha$  sean también intervalos clásicos a nivel  $1-\alpha$ , a ser posible de forma exacta y si no asintóticamente. Se obtienen así las llamadas “*coverage matching priors*”. No obstante, parece lógico pensar que será preferible la elección de una distribución a priori a partir de algún tipo de planteamiento “lógico” y que además, posteriormente, comprobemos que los intervalos bayesianos obtenidos tienen buen comportamiento en el muestreo.

En Lindley, D.V. (1958) se establece un resultado del que se deduce que no siempre va a existir una distribución a priori de la que se obtengan intervalos bayesianos que sean también intervalos clásicos de forma exacta. Entre otras cosas, se demuestra que si  $X$  es una variable aleatoria unidimensional con un único parámetro  $\theta$ , entonces se tiene que la condición necesaria y suficiente para que la distribución fiducial de  $\theta$ , dado  $x$ , sea una distribución de Bayes (es decir, pueda obtenerse como una distribución a posteriori para una cierta distribución a priori  $\psi(\theta)$ ) es que existan transformaciones de  $x$  en  $u$  y de  $\theta$  en  $\gamma$  tales que  $\gamma$  sea un parámetro de localización para  $u$ . En este caso, la distribución que lleva a la coincidencia mencionada es constante en  $\gamma$ .

Observemos que así se está asegurando que, cuando existe una reparametrización  $\gamma = \gamma(\theta)$  y una función  $u=u(x)$  tal que  $p(u | \gamma) = f(u - \gamma)$ , entonces la distribución a priori de Jeffreys produce el mismo resultado que la inferencia fiducial (y también que la inferencia clásica), ya que en el caso de parámetros de localización la distribución a priori de Jeffreys es uniforme. Notemos además que la situación anterior implica que la verosimilitud para  $x$  en términos de  $\gamma$  es trasladada por los datos en el sentido de Box-Tiao, ya que se tendría  $L(\gamma | x) = f(u(x) - \gamma) = \tilde{f}(\gamma - u(x))$ .

Hay varios comentarios interesantes que se pueden hacer a este resultado.

1) En primer lugar, nótese que si  $\theta$  es un parámetro de escala, es decir,  $p(x|\theta) = \theta^{-1}h(x\theta^{-1})$ , entonces con los cambios  $u = \log(x)$  y  $\gamma = \log(\theta)$  se tiene:

$$\begin{aligned} p(x|\theta) &\propto \frac{x}{\theta} h\left(\frac{x}{\theta}\right) = \exp\{\log x - \log \theta\} h(\exp\{\log x - \log \theta\}) = \\ &= \exp\{u - \gamma\} h(\exp(u - \gamma)) = \tilde{h}(u - \gamma) \end{aligned}$$

y así  $\gamma$  es un parámetro de localización para  $u$ , con lo que  $p(\gamma) \propto k$ , es decir,  $p(\theta) \propto \theta^{-1}$ . De esta forma, se ve que la regla de Jeffreys aplicada a la desviación típica de la distribución Normal producirá los mismos resultados que la inferencia clásica.

2) En todo el procedimiento de la demostración se supone que  $X$  es una variable continua y que el rango de  $\theta$  no es discreto (pues se deriva tanto con respecto a  $x$  como con respecto a  $\theta$ ).

3) Cuando se extrae una muestra de tamaño  $n$ , para generalizar el resultado, Lindley supone que existe un estadístico suficiente para el parámetro.

4) Lindley demostró que, dentro de la familia exponencial, las funciones señaladas existen sólo si el modelo es Gamma o Normal, con lo que sólo en estos casos podemos conseguir la igualdad de resultados. Recordemos que esto significa también que sólo en estos modelos podemos conseguir reparametrizaciones para las que la verosimilitud sea trasladada por los datos.

En el artículo de Lindley, se pone de manifiesto que en general no es posible encontrar una distribución a priori tal que la correspondiente distribución a posteriori coincida con la distribución fiducial. Por tanto, se hace necesario un estudio de condiciones para ver cuándo ambos resultados son “aproximadamente” iguales, lo cual se entenderá en términos asintóticos. La mayoría de los estudios hechos en este

sentido se han centrado en ver cuándo los intervalos de confianza unilaterales obtenidos por ambos métodos son iguales salvo infinitésimos de alguna potencia de  $n$ .

Welch, B.L. and Peers, H.W. (1963) obtienen un resultado para el caso de un único parámetro en este sentido y que nos proporciona una justificación importante para la regla de Jeffreys. En concreto, si denotamos por  $S$  a una muestra de la variable  $X$  y  $h(S, \alpha)$  es una función tal que  $P[\theta < h(S, \alpha) | \theta] = \alpha$  (es decir, un intervalo unilateral clásico para  $\theta$ ) y  $g(S, \alpha)$  es una función tal que  $P[\theta < g(S, \alpha) | S] = \alpha$  (es decir, un intervalo Bayesiano para  $\theta$ ), donde la distribución a posteriori para  $\theta$  se ha obtenido partiendo de la distribución a priori de Jeffreys, entonces se tiene:

$$P[\theta < g(S, \alpha) | \theta] = \alpha + O(1/n)$$

y así estamos diciendo que, salvo un infinitésimo de orden  $1/n$ , los intervalos bayesianos unilaterales obtenidos a partir de la regla de Jeffreys también son intervalos unilaterales clásicos.

Aclaremos que al notar  $O(1/n^k)$  el significado del término es:

$$X_n = O(1) \Leftrightarrow \forall \varepsilon > 0, \exists M_\varepsilon \text{ y } N_\varepsilon \text{ t.q. } F_n(M_\varepsilon) - F_n(-M_\varepsilon) > 1 - \varepsilon \quad \forall n > N_\varepsilon$$

y en el caso continuo nos quedaría:

$$X_n = O(1) \Leftrightarrow \forall \varepsilon > 0, \exists M_\varepsilon \text{ y } N_\varepsilon \text{ t.q. } P[|X_n| \leq M_\varepsilon] > 1 - \varepsilon \quad \forall n > N_\varepsilon$$

Señalemos también que la notación más habitual es  $O_p(1/n^k)$  aunque mantendremos por comodidad  $O(1/n^k)$ .



Sin embargo, se hace necesaria una reflexión tras el artículo de Hartigan, J.A. (1966), pues en él se establece que, tomando intervalos centrados, cualquier distribución a priori continua verificará la propiedad deseada. Este artículo puede hacer pensar que este método de trabajo carece de interés, pues cuando  $n$  tiende a infinito, cualquier distribución es buena según este criterio.

No obstante, desde otra perspectiva, puede entenderse que este artículo viene a reforzar la importancia del trabajo de Welch and Peers y de la distribución a priori de Jeffreys en el caso uniparamétrico. Efectivamente, a primera vista resulta extraño que si trabajamos con intervalos centrados sirva cualquier distribución y si lo hacemos con intervalos unilaterales sólo haya una distribución que verifique lo buscado.

Lo que puede ocurrir entonces es que se trabaja con desarrollos en torno al estimador máximo-verosímil  $\hat{\theta}$ ; como es bien sabido, este tipo de desarrollos tiene un comportamiento bueno en la zona central, pero no podemos afirmar lo mismo a medida que nos alejamos del punto en torno al que se hace el desarrollo, es decir, no podemos asegurar que el “ajuste” sea bueno en las colas de la distribución.

El resultado de Hartigan, J.A. (1966) viene a significar que la distribución a posteriori se ajusta bien a la zona central de la verosimilitud cuando  $n$  se hace grande para cualquier distribución a priori. Sin embargo, si pretendemos que este ajuste sea también bueno en las colas, entonces hemos de utilizar la distribución a priori de Jeffreys.

## **2.3 La regla de Jeffreys. Caso multiparamétrico.**

### **2.3.1 La regla general de Jeffreys y la regla modificada.**

En el caso en que el parámetro desconocido  $\theta \in \mathbb{R}^m$ , se definió en el capítulo 1 la matriz de Información de Fisher como

---

$$I(\theta) = E_{\theta} \left[ \left( \frac{\partial \ell}{\partial \theta} \right) \left( \frac{\partial \ell}{\partial \theta} \right)^t \right]$$

que en el caso regular (es decir, que el modelo sea regular con respecto a cada una de las componentes fijadas las otras), también puede escribirse como

$$I(\theta) = -E_{\theta} \left[ \frac{\partial^2 \ell}{\partial \theta \partial \theta^t} \right] \Leftrightarrow (I(\theta))_{ij} = -E_{\theta} \left[ \frac{\partial^2 \ell}{\partial \theta_i \partial \theta_j} \right]$$

Sabemos que la distribución a priori  $p(\theta) \propto \sqrt{|I(\theta)|}$  es invariante ante reparametrizaciones biyectivas cualesquiera, según vimos en la sección anterior. Sin embargo, en esta sección vamos a ver que esta elección conlleva una serie de dificultades que hacen que sea preferible usar otras distribuciones a priori (al menos en algunas situaciones concretas) aun cuando con estas elecciones alternativas se pierda la propiedad de invarianza general.

**Ejemplo 2.8:** Población normal en la que la media y la varianza son desconocidas.

Llamando  $\varphi = \sigma^2$ , tenemos:

$$\ell(\theta, \varphi | x) = -\frac{1}{2} \log \varphi - \frac{1}{2} \frac{(x - \theta)^2}{\varphi} + k$$

$$\frac{\partial^2 \ell}{\partial \theta^2} = -\frac{1}{\varphi}$$

$$\frac{\partial^2 \ell}{\partial \theta \partial \varphi} = -\frac{(x - \theta)}{\varphi^2}$$

$$\frac{\partial^2 \ell}{\partial \varphi^2} = -\frac{1}{2\varphi^2} - \frac{(x-\theta)^2}{\varphi^3}$$

Teniendo en cuenta que  $E[X]=\theta$  y que  $E[(X-\theta)^2]=\varphi$ , tenemos

$$I(\theta, \varphi) = \begin{pmatrix} 1/\varphi & 0 \\ 0 & 1/(2\varphi^2) \end{pmatrix} \Rightarrow |I(\theta, \varphi)| = \frac{1}{2\varphi^3}$$

y por tanto, la distribución a priori no informativa obtenida a partir de la regla es

$$p(\theta, \varphi) \propto \varphi^{-3/2}$$

En base a la invarianza de la regla podemos obtener la distribución para los parámetros  $(\theta, \sigma)$ . Teniendo en cuenta que el jacobiano de la reparametrización es  $2\sigma$ , se verifica

$$p(\theta, \sigma) \propto p(\theta, \varphi) 2\sigma = (\sigma^2)^{-3/2} 2\sigma \propto \sigma^{-2}$$

En este ejemplo, el propio Jeffreys desestimó la distribución  $p(\mu, \sigma) \propto \sigma^{-2}$ , recomendando el uso de la densidad  $p(\mu, \sigma) \propto \sigma^{-1}$ . Llamando  $\lambda$  a la media y  $\sigma$  a la desviación típica, Jeffreys, H. (1961) dice (p.182): “...Pero si  $\lambda$  y  $\sigma$  son ambos considerados,  $\|g_{ik}\|^{1/2} d\lambda d\sigma \propto d\lambda d\sigma / \sigma^2$  (es decir,  $p(\lambda, \sigma) \propto 1/\sigma^2$ ) en vez de  $d\lambda d\sigma / \sigma$ . Si se aplica el mismo método a una distribución conjunta para varias variables con parámetros independientes se añadiría un factor extra  $1/\sigma$  por cada una. El índice en la distribución t correspondiente sería siempre  $(n+1)/2$  aunque se estimen muchos valores. **Esto es inaceptable.**”<sup>5</sup>

---

<sup>5</sup> Jeffreys, H. (1961), p.182. El texto original es: “*But if  $\lambda$  and  $\sigma$  are both varied,  $\|g_{ik}\|^{1/2} d\lambda d\sigma \propto d\lambda d\sigma / \sigma^2$  instead of  $d\lambda d\sigma / \sigma$ . If the same method was applied to a joint distribution for*”

**Definición:** Decimos que un modelo generador de datos pertenece a la familia de localización-escala si la función de densidad es de la forma

$$f(x | \mu, \sigma) = \prod_{i=1}^n \frac{1}{\sigma} h\left(\frac{x_i - \mu}{\sigma}\right) \quad -\infty < x_i < +\infty, \quad \mu \in \mathbb{R}, \quad \sigma > 0$$

donde  $h$  es una función de densidad conocida.

Observemos que el efecto de cambiar  $\mu$  es simplemente una traslación de la distribución y por ello decimos que  $\mu$  es un parámetro de localización. En cambio, el efecto de cambiar  $\sigma$  es hacer que la densidad esté más o menos concentrada y por ello decimos que  $\sigma$  es un parámetro de escala. Si el parámetro  $\sigma$  es conocido, hablaremos de un modelo de localización y, análogamente, si  $\mu$  es conocido hablaremos de un modelo de escala. Notemos también que podemos generalizar sin dificultad la definición para un parámetro de localización multidimensional.

Como ejemplos, podemos señalar que el modelo  $N(\mu, \sigma)$  es de localización-escala con  $h(x) = (2\pi)^{-1/2} \exp(-x^2/2)$  siendo la media el parámetro de localización y la desviación típica el de escala y que el modelo  $\text{Exp}(\theta)$  es de escala para el parámetro  $\sigma = \theta^{-1}$  y  $h(x) = \exp(-x)$ .

Jeffreys, H. (1961), concluye que si  $\lambda$  y  $\sigma$  son parámetros de localización y escala respectivamente, hemos de tomar como distribución a priori  $p(\lambda, \sigma) \propto 1/\sigma$  y señala que esta distribución es invariante ante reparametrizaciones del tipo  $\lambda' = \lambda + k\sigma$ . Esta distribución se obtiene también al calcular  $\det(I(\sigma))^{1/2}$  donde  $\lambda$  se toma constante lo que conduce al mismo resultado que considerar que **ambos parámetros son independientes a priori** y así obtener las distribuciones marginales

---

*several variables about independent true values, an extra factor  $1/\sigma$  would appear for each. The index in the corresponding  $t$  distribution would always be  $1/2(n+1)$  however many true values were estimated. This is unacceptable."*

---

unidimensionales para posteriormente obtener la distribución a priori conjunta multiplicando dichas marginales ya que la distribución unidimensional correspondiente al parámetro de escala es constante. En general, si hay  $k$  parámetros de localización y uno de escala, podemos generalizar tomando  $p(\lambda_1, \dots, \lambda_k, \sigma) \propto \det(I(\sigma))^{1/2}$  donde  $\lambda_1, \dots, \lambda_k$  se toman constantes. Este procedimiento es el que llamaremos **regla de Jeffreys para el caso de localización-escala** o **regla de Jeffreys modificada**.

Vemos por tanto que se abandona así la propiedad de invarianza general pero a cambio se evitan otra serie de dificultades importantes. Vamos a ver a continuación con dos ejemplos cuáles son los inconvenientes que puede presentar el uso de la regla general de Jeffreys en el caso de más de un parámetro.

- **El modelo de regresión lineal.**

En el modelo de regresión lineal con una variable explicativa, perteneciente a la familia de localización-escala, vamos a ver que el uso de la regla de Jeffreys modificada conduce a intervalos bayesianos coincidentes numéricamente con los clásicos, mientras que el uso de la regla general no tiene esta propiedad deseable.

Para ello consideremos el modelo de regresión:

$$y_t = \beta_0 + \beta_1 x_t + u_t \quad \text{donde } u_t \sim N(0, \sigma^2) \text{ indep.}$$

Notemos que en este caso tenemos dos parámetros de localización que son  $\beta_0$  y  $\beta_1$  y uno de escala que es  $\sigma$ , pues la función de verosimilitud viene dada por

$$L(\beta_0, \beta_1, \sigma | x_t, y_t) = \left( \frac{1}{\sqrt{2\pi\sigma}} \right)^n \exp \left\{ -\frac{1}{2\sigma^2} \sum_t (y_t - \beta_0 - \beta_1 x_t)^2 \right\} =$$

$$= \left( \frac{1}{\sqrt{2\pi}} \right)^n \prod_t \frac{1}{\sigma} \exp \left\{ -\frac{1}{2} \left( \frac{y_t - \beta_0 - \beta_1 x_t}{\sigma} \right)^2 \right\}$$

$$\ell(\beta_0, \beta_1, \sigma | x_t, y_t) = -\frac{n}{2} \log(2\pi) - n \log \sigma - \frac{1}{2\sigma^2} \sum_t (y_t - \beta_0 - \beta_1 x_t)^2$$

Las segundas derivadas del logaritmo de la verosimilitud serían:

$$\frac{\partial \ell}{\partial \beta_0} = \frac{1}{2\sigma^2} 2 \sum_t (y_t - \beta_0 - \beta_1 x_t)$$

$$\frac{\partial \ell}{\partial \beta_1} = \frac{1}{2\sigma^2} 2 \sum_t (y_t - \beta_0 - \beta_1 x_t) x_t$$

$$\frac{\partial \ell}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^3} \sum_t (y_t - \beta_0 - \beta_1 x_t)^2$$

$$\frac{\partial^2 \ell}{\partial \beta_0^2} = -\frac{n}{\sigma^2}$$

$$\frac{\partial^2 \ell}{\partial \beta_0 \partial \beta_1} = -\frac{\sum_t x_t}{\sigma^2}$$

$$\frac{\partial^2 \ell}{\partial \beta_0 \partial \sigma} = -\frac{2}{\sigma^3} \sum_t (y_t - \beta_0 - \beta_1 x_t)$$

$$\frac{\partial^2 \ell}{\partial \beta_1^2} = -\frac{\sum_t x_t^2}{\sigma^2}$$

$$\frac{\partial^2 \ell}{\partial \beta_1 \partial \sigma} = -\frac{2}{\sigma^3} \sum_t (y_t - \beta_0 - \beta_1 x_t) x_t$$

$$\frac{\partial^2 \ell}{\partial \sigma^2} = \frac{n}{\sigma^2} - \frac{3}{\sigma^4} \sum_t (y_t - \beta_0 - \beta_1 x_t)^2$$

Las esperanzas de las derivadas segundas que no son constantes son:

$$E\left[\frac{\partial^2 \ell}{\partial \beta_0 \partial \sigma}\right] = -\frac{2}{\sigma^3} \sum_t E[(y_t - \beta_0 - \beta_1 x_t)] = 0$$

$$E\left[\frac{\partial^2 \ell}{\partial \beta_1 \partial \sigma}\right] = -\frac{2}{\sigma^3} \sum_t x_t E[(y_t - \beta_0 - \beta_1 x_t)] = 0$$

$$E\left[\frac{\partial^2 \ell}{\partial \sigma^2}\right] = \frac{n}{\sigma^2} - \frac{3n\sigma^2}{\sigma^4} = -\frac{2n}{\sigma^2}$$

Así, la matriz de Información de Fisher viene dada por:

$$I_n(\beta_0, \beta_1, \sigma) = \begin{pmatrix} \sigma^{-2}n & \sigma^{-2} \sum_t x_t & 0 \\ \sigma^{-2} \sum_t x_t & \sigma^{-2} \sum_t x_t^2 & 0 \\ 0 & 0 & \sigma^{-2}2n \end{pmatrix}$$

cuyo determinante es:

$$|I_n| = \frac{2n^2 \sum_t x_t^2}{\sigma^6} - \frac{2n \left(\sum_t x_t\right)^2}{\sigma^6} = \frac{2n \left(n \sum_t x_t^2 - \left(\sum_t x_t\right)^2\right)}{\sigma^6}$$

de lo que deducimos inmediatamente que la distribución a priori de Jeffreys conjunta será:

$$p(\beta_0, \beta_1, \sigma) \propto \sqrt{|I_n|} \propto \sqrt{\frac{1}{\sigma^6}} \propto \frac{1}{\sigma^3}$$

La distribución a posteriori conjunta es:

$$p(\beta_0, \beta_1, \sigma | x_t, y_t) \propto \ell(\beta_0, \beta_1, \sigma | x_t, y_t) p(\beta_0, \beta_1, \sigma) \propto \left(\frac{1}{\sigma}\right)^{n-3} \exp\left\{-\frac{1}{2\sigma^2} \sum_t (y_t - \beta_0 - \beta_1 x_t)^2\right\}$$

Vamos a considerar ahora los estimadores mínimo-cuadráticos de  $\beta_0$  y  $\beta_1$  y el estimador insesgado de  $\sigma^2$ , es decir:

$$\hat{\beta}_1 = \frac{\text{Cov}(X, Y)}{\text{Var}(X)}$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$$

$$\hat{\sigma}^2 = \frac{\sum_t (y_t - \hat{\beta}_0 - \hat{\beta}_1 x_t)^2}{n-2} = \frac{\sum_t \hat{u}_t^2}{n-2}$$

Recordemos que con estos estimadores se verificaban las igualdades:

$$\sum_t \hat{u}_t = \sum_t \hat{u}_t x_t = 0$$

Manipulando la distribución a posteriori y usando las igualdades señaladas, puede obtenerse una expresión más adecuada para obtener la distribución marginal de  $\sigma$  como sigue.

$$\begin{aligned} \sum_t (y_t - \beta_0 - \beta_1 x_t)^2 &= \sum_t \left[ y_t - \hat{\beta}_0 - \hat{\beta}_1 x_t - (\beta_0 - \hat{\beta}_0) - (\beta_1 x_t - \hat{\beta}_1 x_t) \right]^2 = \\ &= \sum_t (y_t - \hat{\beta}_0 - \hat{\beta}_1 x_t)^2 + \sum_t (\beta_0 - \hat{\beta}_0)^2 + \sum_t (\beta_1 - \hat{\beta}_1)^2 x_t^2 + 2 \sum_t (\beta_0 - \hat{\beta}_0) (\beta_1 - \hat{\beta}_1) x_t - \\ &\quad - 2 \sum_t (y_t - \hat{\beta}_0 - \hat{\beta}_1 x_t) (\beta_0 - \hat{\beta}_0) - 2 \sum_t (y_t - \hat{\beta}_0 - \hat{\beta}_1 x_t) (\beta_1 - \hat{\beta}_1) x_t \end{aligned}$$

Al ser los dos últimos sumandos de esta expresión, respectivamente,  $2(\beta_0 - \hat{\beta}_0) \sum_t \hat{u}_t$  y  $2(\beta_1 - \hat{\beta}_1) \sum_t x_t \hat{u}_t$ , ambos serán nulos, resultando definitivamente que la distribución a posteriori podemos expresarla como:



$$p(\beta_0, \beta_1, \sigma | x_t, y_t) = \left(\frac{1}{\sigma}\right)^{n+3} \exp \left\{ -\frac{1}{2\sigma^2} \left[ (n-2)\hat{\sigma}^2 + n(\beta_0 - \hat{\beta}_0)^2 + \left(\sum_t x_t^2\right)(\beta_1 - \hat{\beta}_1)^2 + 2\sum_t x_t(\beta_0 - \hat{\beta}_0)(\beta_1 - \hat{\beta}_1) \right] \right\}$$

La distribución marginal de  $\sigma$  será:

$$p(\sigma | x_t, y_t) \propto \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} p(\beta_0, \beta_1, \sigma | x_t, y_t) d\beta_0 d\beta_1 \propto \left(\frac{1}{\sigma}\right)^{n+3} \exp \left\{ -\frac{1}{2\sigma^2} (n-2)\hat{\sigma}^2 \right\} \iint \exp \left\{ -\frac{1}{2\sigma^2} \left[ n(\beta_0 - \hat{\beta}_0)^2 + \left(\sum_t x_t^2\right)(\beta_1 - \hat{\beta}_1)^2 + 2\sum_t x_t(\beta_0 - \hat{\beta}_0)(\beta_1 - \hat{\beta}_1) \right] \right\} d\beta_0 d\beta_1$$

donde la doble integral puede resolverse fácilmente por ser el núcleo de una distribución Normal Bivariante, resultando entonces que:

$$p(\sigma | x_t, y_t) \propto \left(\frac{1}{\sigma}\right)^{n+3} \exp \left\{ -\frac{1}{2\sigma^2} (n-2)\hat{\sigma}^2 \right\} \sigma^2 \propto \left(\frac{1}{\sigma}\right)^{\frac{n+1}{2}} \exp \left\{ -\frac{(n-2)\hat{\sigma}^2}{2\sigma^2} \right\}$$

Definimos ahora  $u(\sigma) = \frac{(n-2)\hat{\sigma}^2}{\sigma^2}$ , con lo que  $\frac{\partial u}{\partial \sigma} \propto -\frac{1}{\sigma^3}$ , y de esta forma, aplicando el correspondiente cambio de variable, tenemos que la distribución de  $u$  será:

$$p(u | x_t, y_t) \propto \left(\frac{u}{(n-2)\hat{\sigma}^2}\right)^{\frac{n+1}{2}} e^{-\frac{u}{2}} \left(\frac{u}{(n-2)\hat{\sigma}^2}\right)^{-\frac{3}{2}} \propto u^{\frac{n-2}{2}} e^{-\frac{u}{2}} \propto u^{\frac{n}{2}-1} e^{-\frac{u}{2}}$$

Vemos así que  $u(\sigma) = \frac{(n-2)\hat{\sigma}^2}{\sigma^2} \sim \chi_n^2$  si tomamos como distribución a priori

la obtenida al aplicar la regla general de Jeffreys.

Siguiendo la recomendación del propio Jeffreys para el caso de localización-escala, los resultados serían los siguientes.

$$p(\beta_0, \beta_1, \sigma | x_t, y_t) \propto \sqrt{-E \left[ \frac{\partial^2 L}{\partial \sigma^2} \right]} \propto \sqrt{\frac{2n}{\sigma^2}} \propto \frac{1}{\sigma}$$

El procedimiento entonces es prácticamente igual que el desarrollado anteriormente, obteniéndose en este caso:

$$p(\sigma | x_t, y_t) \propto \left( \frac{1}{\sigma} \right)^{n+1} \exp \left\{ -\frac{1}{2\sigma^2} (n-2) \hat{\sigma}^2 \right\} \sigma^2 \propto \left( \frac{1}{\sigma} \right)^{n-1} \exp \left\{ -\frac{(n-2) \hat{\sigma}^2}{2\sigma^2} \right\}$$

Haciendo el mismo cambio de variable que se hizo en el caso anterior, obtenemos la distribución de  $u$ , que es:

$$p(u | x_t, y_t) \propto u^{\frac{n-4}{2}} e^{-\frac{u}{2}} \propto u^{\frac{n-2}{2}-1} e^{-\frac{u}{2}}$$

Así,  $u(\sigma) = \frac{(n-2) \hat{\sigma}^2}{\sigma^2} \sim \chi_{n-2}^2$  y por tanto los intervalos Bayesianos obtenidos

coincidirán numéricamente con los obtenidos a través de la Inferencia Clásica.

Observemos que en este ejemplo se aprecia lo señalado por Jeffreys, es decir, que si aplicamos la regla general para obtener la distribución a priori entonces la distribución  $\chi^2$  tiene  $n$  grados de libertad independientemente del número de parámetros en estudio, cuando lo que parece razonable es que los grados de libertad disminuyan a medida que aumenta el número de parámetros desconocidos.

- **El ejemplo de Stein.**

El ejemplo quizás más sorprendente de discrepancia entre la inferencia clásica y Bayesiana usando la regla general de Jeffreys es el proporcionado por Stein, C. (1959).

Sean  $X_1, \dots, X_n$  variables aleatorias normales independientes de medias  $\theta_1, \dots, \theta_n$  respectivamente y varianza 1, es decir,  $X_i \sim N(\theta_i, 1)$ . Planteamos construir un intervalo unilateral a nivel  $\alpha$  de la forma  $[c, +\infty)$  para la función  $\sum \theta_i^2$ . Sea  $I_{n,\alpha}$  el intervalo obtenido utilizando la distribución a priori de Jeffreys. El resultado dice, no sólo que la probabilidad de cubrimiento en sentido frecuentista no es  $\alpha$ , sino que, sorprendentemente, esta probabilidad tiende a cero cuando  $n$  tiende a  $\infty$ .

Antes de demostrar el resultado, es conveniente aclarar lo siguiente. Stein construye  $I_{n,\alpha}$  calculando la distribución Fiducial de  $(\theta_1, \dots, \theta_n)$  y a partir de ella la de  $\sum \theta_i^2$ . Ahora bien, puesto que dicha distribución fiducial coincide con la distribución a posteriori de  $\sum \theta_i^2$  si tomamos como distribución a priori para  $(\theta_1, \dots, \theta_n)$  la de Jeffreys, el intervalo es el mismo.

En la demostración se hace uso de una hipótesis adicional que es bastante razonable y que cubre la mayoría de las situaciones reales. Dicha hipótesis es que  $\sum \theta_i^2$  es  $o(n^2)$ , es decir,  $\lim_{n \rightarrow \infty} \frac{\sum \theta_i^2}{n^2} = 0$ . Una vez hechas estas aclaraciones, vamos a ver que, efectivamente, se tiene el resultado.

Como se dijo, partimos de una muestra de tamaño  $n$ , de forma que  $X_i \sim N(\theta_i, 1)$  La función de verosimilitud y su logaritmo vienen dados por:

$$L(\theta | x) = \left( \frac{1}{\sqrt{2\pi}} \right)^n \exp \left\{ -\frac{1}{2} \sum_i (x_i - \theta_i)^2 \right\}$$

$$\ell(\theta | x) = -\frac{n}{2} \log(2\pi) - \frac{1}{2} \sum_i (x_i - \theta_i)^2$$

La matriz de Información de Fisher resultante para este modelo es la Identidad de orden  $n$ , ya que

$$\frac{\partial \ell}{\partial \theta_i} = x_i - \theta_i$$

$$\frac{\partial^2 \ell}{\partial \theta_i^2} = -1$$

$$\frac{\partial^2 \ell}{\partial \theta_i \partial \theta_j} = 0 \quad \forall i \neq j$$

y por tanto la distribución a priori de Jeffreys en este caso es constante. La distribución a posteriori de  $\theta$  dados los valores de  $x$ , es entonces una Normal Multivariante de vector de medias  $(x_1, \dots, x_n)$ , siendo la matriz de varianzas-covarianzas la identidad de orden  $n$ , pues:

$$p(\theta | x) \propto L(\theta | x) p(\theta) \propto \exp \left\{ -\frac{1}{2} \sum_i (x_i - \theta_i)^2 \right\}$$

Deducimos así que la distribución a posteriori para  $\sum \theta_i^2$  dados los datos, es una  $\chi_n^2(\sum x_i^2)$ . Un intervalo unilateral de probabilidad  $\alpha$  sería entonces  $I_{n,\alpha} = (\phi_{\alpha,n}(\sum x_i^2), +\infty)$  donde se verifica  $P[\chi_n^2(\sum x_i^2) > \phi_{\alpha,n}(\sum x_i^2)] = \alpha$ . Vamos a notar a partir de ahora  $\phi$  en vez de  $\phi_{\alpha,n}(\sum x_i^2)$  por simplicidad.

Tenemos ahora en cuenta que para  $n$  suficientemente grande, una distribución  $\chi_n^2(\sum x_i^2)$  puede aproximarse por una distribución  $N(n + \sum x_i^2, 2n + 4\sum x_i^2)$  y así,  $U = \sum \theta_i^2 | x \approx N(n + \sum x_i^2, 2n + 4\sum x_i^2)$ .

Siendo  $t_\alpha$  el percentil de orden  $\alpha$  de una Normal tipificada, vamos a probar que  $\forall t'_\alpha > t_\alpha$  se verifica que  $\phi > n + \sum x_i^2 - t'_\alpha \sqrt{2n + 4\sum x_i^2}$ .

En efecto, teníamos  $P[\chi_n^2(\sum x_i^2) > \phi_{\alpha,n}(\sum x_i^2)] = \alpha$ , i.e.,  $P[U > \phi] = \alpha$ .

Teniendo en cuenta la aproximación hecha para la distribución de  $U$ , se obtiene:

$$\begin{aligned} \alpha = P[U > \phi] &= P\left[\frac{U - (n + \sum x_i^2)}{\sqrt{2n + 4\sum x_i^2}} > \frac{\phi - (n + \sum x_i^2)}{\sqrt{2n + 4\sum x_i^2}}\right] = P\left[Z > \frac{\phi - (n + \sum x_i^2)}{\sqrt{2n + 4\sum x_i^2}}\right] \Rightarrow \\ &\Rightarrow P\left[Z < \frac{\phi - (n + \sum x_i^2)}{\sqrt{2n + 4\sum x_i^2}}\right] = 1 - \alpha \Rightarrow \frac{\phi - (n + \sum x_i^2)}{\sqrt{2n + 4\sum x_i^2}} = t_{1-\alpha} \\ &\Rightarrow t_\alpha = \frac{-\phi + n + \sum x_i^2}{\sqrt{2n + 4\sum x_i^2}} \Rightarrow \phi = n + \sum x_i^2 - t_\alpha \sqrt{2n + 4\sum x_i^2} \end{aligned}$$

Al ser  $t'_\alpha > t_\alpha$  concluimos que

$$\phi > n + \sum x_i^2 - t'_\alpha \sqrt{2n + 4\sum x_i^2}$$

como se quería demostrar.

Así, para  $\theta_1, \dots, \theta_n$  fijos, podemos establecer la siguiente cota para la probabilidad de cubrimiento del intervalo bayesiano.

$$\begin{aligned} P\left[\sum \theta_i^2 \geq \phi\right] &\leq P\left[\sum \theta_i^2 \geq n + \sum x_i^2 - t'_\alpha \sqrt{2n + \sum x_i^2}\right] = \\ &= P\left[\sum x_i^2 \leq \sum \theta_i^2 - n + t'_\alpha \sqrt{2n + \sum x_i^2}\right] \end{aligned}$$

Ahora, aplicando la desigualdad de Markov podemos obtener que  $P\left[2n + 4\sum x_i^2 > \varepsilon n^2\right] < \varepsilon \quad \forall n \geq n_0$ . (Recordemos que la desigualdad de Markov dice que dada una variable aleatoria  $T$  y una función  $g(T) \geq 0$  se verifica que  $P[T \geq k] < E[g(T)]/k$ ).

En efecto, sea  $T = \sum x_i^2$  y  $g(T) = 2n + 4T \geq 0$ . Sabemos que  $T \sim \chi_n^2(\sum \theta_i^2) \approx N(n + \sum \theta_i^2, 2n + 4\sum \theta_i^2)$  y por tanto:

$$\begin{aligned} P\left[2n + 4\sum x_i^2 > \varepsilon n^2\right] &= P\left[2n + 4T > \varepsilon n^2\right] < \frac{E[2n + 4T]}{\varepsilon n^2} = \frac{2n + 4E[T]}{\varepsilon n^2} = \\ &= \frac{2n + 4n + 4\sum \theta_i^2}{\varepsilon n^2} = \frac{6n + 4\sum \theta_i^2}{\varepsilon n^2} = \frac{1}{\varepsilon} \left( \frac{6}{n} + 4 \frac{\sum \theta_i^2}{n^2} \right) \end{aligned}$$

Usamos aquí la hipótesis  $\lim_{n \rightarrow \infty} \frac{\sum \theta_i^2}{n^2} = 0$  y entonces

$$\exists n_0, \forall n \geq n_0 \Rightarrow \frac{6}{n} + \frac{4\sum \theta_i^2}{n^2} < \varepsilon^2 \text{ con lo que } P\left[2n + 4\sum x_i^2 > \varepsilon n^2\right] < \frac{\varepsilon^2}{\varepsilon} = \varepsilon.$$

De este resultado podemos deducir que

$$P\left[\sum x_i^2 \leq \sum \theta_i^2 - n + t'_\alpha \sqrt{2n + 4\sum x_i^2}\right] \leq P\left[\sum x_i^2 \leq \sum \theta_i^2\right] + \left(\frac{1}{t'_\alpha}\right)^2 \quad (*)$$

Dado  $\varepsilon > 0$ , como  $t'_\alpha$  es arbitrario (con la condición de que sea mayor que  $t_\alpha$ ),

podemos elegir  $t'_\alpha$  suficientemente grande como para que  $\left(\frac{1}{t'_\alpha}\right)^2 < \varepsilon$  y así:

$$P\left[\sum x_i^2 \leq \sum \theta_i^2 - n + t'_\alpha \sqrt{2n + 4\sum x_i^2}\right] \leq P\left[\sum x_i^2 \leq \sum \theta_i^2\right] + \varepsilon$$

Ahora, vamos a calcular  $P\left[\sum x_i^2 \leq \sum \theta_i^2\right]$ . Puesto que  $\sum x_i^2$  se puede aproximar por una  $N\left(n + \sum \theta_i^2, 2n + 4\sum \theta_i^2\right)$ , tendremos:

$$\begin{aligned} P\left[\sum x_i^2 \leq \sum \theta_i^2\right] &= P\left[Z \leq \frac{\sum \theta_i^2 - n - \sum \theta_i^2}{\sqrt{2n + 4\sum \theta_i^2}}\right] = \\ &= P\left[Z \leq \frac{-n}{\sqrt{2n + 4\sum \theta_i^2}}\right] = P\left[Z > \frac{n}{\sqrt{2n + 4\sum \theta_i^2}}\right] \\ \Rightarrow \lim_{n \rightarrow \infty} P\left[\sum x_i^2 \leq \sum \theta_i^2\right] &= \lim_{n \rightarrow \infty} P\left[Z > \frac{n}{\sqrt{2n + 4\sum \theta_i^2}}\right] = \\ &= P\left[Z > \lim_{n \rightarrow \infty} \frac{1}{\sqrt{\frac{2}{n} + \frac{4\sum \theta_i^2}{n^2}}}\right] = P[Z > +\infty] = 0 \end{aligned}$$

(obsérvese que hemos vuelto a usar la hipótesis de que  $\lim_{n \rightarrow \infty} \frac{\sum \theta_i^2}{n^2} = 0$ ).

En conclusión, hemos obtenido:

$$P\left[\sum \theta_i^2 \geq \phi\right] \leq P\left[\sum x_i^2 \leq \sum \theta_i^2 - n + t'_\alpha \sqrt{2n + 4\sum x_i^2}\right] \leq P\left[\sum x_i^2 \leq \sum \theta_i^2\right] + \varepsilon \quad \forall \varepsilon > 0$$

y también

$$\lim_{n \rightarrow \infty} P\left[\sum x_i^2 \leq \sum \theta_i^2\right] = 0$$

De ambos resultados, se obtiene que

$$\lim_{n \rightarrow \infty} P\left[\sum \theta_i^2 \geq \phi\right] \leq \varepsilon \quad \forall \varepsilon > 0 \quad \Rightarrow \quad \lim_{n \rightarrow \infty} P\left[\sum \theta_i^2 \geq \phi\right] = 0$$

que es el resultado que se pretendía demostrar.

Antes de continuar vamos a ver que se verifica (\*).

En efecto,

$$P\left[\sum x_i^2 \leq \sum \theta_i^2 - n + t'_\alpha \sqrt{2n + 4 \sum x_i^2}\right] = P[w \leq t - r]$$

donde hemos notado  $w = \sum x_i^2$ ,  $r = n - t'_\alpha \sqrt{2n + 4 \sum x_i^2}$  y  $t = \sum \theta_i^2$ .

Ahora bien, tendremos:

$$\begin{aligned} P[w \leq t - r] &= P[w + r \leq t] = P[\{w + r \leq t\} \cap \{r < 0\}] + P[\{w + r \leq t\} \cap \{r \geq 0\}] \leq \\ &\leq P[r < 0] + P[\{w + r \leq t\} \cap \{r \geq 0\}] \leq P[r < 0] + P[w \leq t] = \\ &= P\left[n - t'_\alpha \sqrt{2n + 4 \sum x_i^2} < 0\right] + P\left[\sum x_i^2 \leq \sum \theta_i^2\right] = \\ &= P\left[\sqrt{2n + 4 \sum x_i^2} > \frac{n}{t'_\alpha}\right] + P\left[\sum x_i^2 \leq \sum \theta_i^2\right] \leq \\ &= P\left[2n + 4 \sum x_i^2 > n^2 \left(\frac{1}{t'_\alpha}\right)^2\right] + P\left[\sum x_i^2 \leq \sum \theta_i^2\right] < \left(\frac{1}{t'_\alpha}\right)^2 + P\left[\sum x_i^2 \leq \sum \theta_i^2\right] \end{aligned}$$

como se quería demostrar.

♣



Este ejemplo ha sido largamente discutido y se han encontrado más dificultades relacionadas con problemas de estimación que indicamos a continuación.

Concretamente, si en la misma situación, el problema consiste en estimar  $\theta$ , puesto que  $\theta | x \sim N(x, I)$ , tendríamos  $\hat{\theta} = E[\theta | x] = x = (x_1, \dots, x_n)^t$ . Stein, C. (1956) probó que este estimador es inadmisibles en el sentido de que siendo  $L(\theta, \delta) = \sum (\theta_i - \delta_i)^2$ , entonces  $\exists \gamma = (\gamma_1, \dots, \gamma_n)^t$  t.q.  $E_\theta [L(\theta, \gamma)] \leq E_\theta [L(\theta, x)] \quad \forall \theta$  dándose la desigualdad estricta en algún  $\theta_0$ .

El estimador de  $\delta = \sum \theta_i^2$  también presenta problemas, pues  $\hat{\delta} = E[\chi_n^2(\sum x_i^2)] = n + \sum x_i^2$ , mientras que el UMVUE para  $\delta$  es  $\tilde{\delta} = \sum x_i^2 - n$  (y el estimador de máxima verosimilitud es  $\bar{\delta} = \sum x_i^2$ ). Por tanto, parece que  $\hat{\delta}$  es excesivamente grande para estimar  $\delta$ , sobre todo a medida que aumentamos  $n$ . Notemos que hay una estrecha relación entre esta dificultad y la observada para el intervalo unilateral. En ambos casos, los resultados obtenidos “están excesivamente a la derecha” correspondiéndose entonces esto con el hecho de que la distribución a posteriori de  $\delta$  parece concentrar excesiva masa de probabilidad en valores grandes del rango, es decir, la distribución “se escapa a la derecha del verdadero valor de  $\delta$ ”.

### **2.3.2 Coincidencia de intervalos bayesianos y clásicos para más de un parámetro.**

Hemos visto con los dos ejemplos anteriores que en el caso de más de un parámetro la regla general de Jeffreys no tiene buenas propiedades con respecto a las probabilidades de cubrimiento frecuentistas de los intervalos Bayesianos obtenidos a partir de ella. Vamos a señalar algunos trabajos en esta dirección en los que se pone

de manifiesto esta circunstancia y de los que deduciremos también que en modelos de localización-escala el uso de la regla modificada lleva a intervalos con buenas propiedades frecuentistas.

- **La ecuación de Peers.**

En el trabajo de Peers, H.W. (1965) la situación general es que tenemos el parámetro desconocido multidimensional  $\theta=(\theta_1, \dots, \theta_k)$  de un modelo regular donde  $\theta_1$  es el parámetro objeto de estudio. Entonces, los intervalos bayesianos de probabilidad  $\alpha$  obtenidos a partir de una distribución a priori  $p(\theta)$  tendrán probabilidad de cubrimiento frecuentista  $\alpha + o(1/n)$  si se verifica la siguiente ecuación en derivadas parciales

$$\sum_{j=1}^k \left[ k^{1j} (k^{11})^{-1/2} \frac{\partial \psi}{\partial \theta_j} + \frac{\partial}{\partial \theta_j} \left\{ k^{1j} (k^{11})^{-1/2} \right\} \right] = 0$$

donde  $\psi(\theta) = \log p(\theta)$ ,  $k_{ij} = (I(\theta))_{ij}$  y  $k^{ij} = (I^{-1}(\theta))_{ij}$ .

**Ejemplo 2.9:** Consideremos la distribución  $N(\mu, \sigma)$ .

Para este caso, como sabemos, se tiene:

$$I(\mu, \sigma) = \begin{pmatrix} 1/\sigma^2 & 0 \\ 0 & 2/\sigma^2 \end{pmatrix}; \quad I^{-1}(\mu, \sigma) = \begin{pmatrix} \sigma^2 & 0 \\ 0 & \sigma^2/2 \end{pmatrix}$$

A) Supongamos que el parámetro de interés es  $\mu$ .

Para  $j=1$  el término correspondiente en el sumatorio es:

$$k^{11} (k^{11})^{-1/2} \frac{\partial \psi}{\partial \mu} + \frac{\partial}{\partial \mu} \left\{ k^{11} (k^{11})^{-1/2} \right\}$$

Para  $j=2$ , al ser  $k^{12}=0$ , el sumando correspondiente es cero, quedando entonces la ecuación en la forma:

$$\sqrt{k^{11}} \frac{\partial \psi}{\partial \mu} + \frac{\partial \sqrt{k^{11}}}{\partial \mu} = 0$$

En este caso, la ecuación se resuelve fácilmente, ya que:

$$\sqrt{k^{11}} \frac{\partial \psi}{\partial \mu} + \frac{\partial \sqrt{k^{11}}}{\partial \mu} = 0 \Leftrightarrow \sigma \frac{\partial \psi}{\partial \mu} + \frac{\partial \sigma}{\partial \mu} = 0 \Leftrightarrow \frac{\partial \psi}{\partial \mu} = 0 \Leftrightarrow p(\mu, \sigma) \propto c(\sigma)$$

Veremos más adelante con mayor detenimiento que en Nicolau, A. (1993) se hace otra interpretación de este resultado. En efecto, puesto que el parámetro de interés es  $\mu$  y la distribución queda determinada en realidad en lo que se refiere a dicha componente, pues aparece en la solución una función arbitraria del parámetro  $\sigma$ , en esencia lo que estamos determinando con este procedimiento es la distribución del parámetro de interés conocido el perturbador, es decir, que entonces sería  $p(\mu | \sigma) \propto c(\sigma)$ .

**B)** Supongamos que el parámetro de interés es  $\sigma$ .

Ahora, la ecuación correspondiente sería:

$$\sqrt{k^{11}} \frac{\partial \psi}{\partial \sigma} + \frac{\partial \sqrt{k^{11}}}{\partial \sigma} = 0$$

que también resulta fácilmente resoluble, quedando como solución general

$$\psi(\sigma, \mu) = \log p(\sigma, \mu) \propto -\log(\sigma) + c(\mu) \Leftrightarrow p(\sigma, \mu) \propto \frac{1}{\sigma} c(\mu)$$

Observemos así que la distribución obtenida con la regla de Jeffreys modificada verifica la propiedad buscada, pero no ocurre lo mismo con la distribución obtenida a partir de la regla general.

**Ejemplo 2.10:** En el ejemplo 1 se aprecia la gran simplificación que supone el hecho  $k^{12} = 0$ , simplificación que sigue teniéndose siempre que  $k^{lj} = 0 \quad \forall j \neq 1$ . Vamos a suponer así que los parámetros perturbadores  $\theta_2, \dots, \theta_k$  son **ortogonales** a  $\theta_1$  en el sentido de Cox, D.R. and Reid, N. (1987), es decir, que  $(I(\theta))_{1j} = k_{1j} = 0 \quad \forall j$ . Por simplicidad, trabajaremos con sólo dos parámetros, pues el tener más no aporta nada esencial al problema. De esta forma, supongamos que  $\theta = (\theta_1, \theta_2)^t$  y que  $k_{12} = 0$ , es decir, que la matriz de Información de Fisher sería de la forma:

$$I = \begin{pmatrix} I_{11}(\theta) & 0 \\ 0 & I_{22}(\theta) \end{pmatrix} = \begin{pmatrix} I_{11}(\theta_1, \theta_2) & 0 \\ 0 & I_{22}(\theta_1, \theta_2) \end{pmatrix}$$

La ecuación entonces sería:

$$\sqrt{k^{11}} \frac{\partial \psi}{\partial \theta_1} + \frac{\partial \sqrt{k^{11}}}{\partial \theta_1} = 0$$

cuya solución puede obtenerse fácilmente, pues:

$$\begin{aligned} \sqrt{k^{11}} \frac{\partial \psi}{\partial \theta_1} + \frac{\partial \sqrt{k^{11}}}{\partial \theta_1} = 0 &\Leftrightarrow \sqrt{k^{11}} \frac{\partial \psi}{\partial \theta_1} + \frac{\frac{\partial k^{11}}{\partial \theta_1}}{2\sqrt{k^{11}}} = 0 \Leftrightarrow 2k^{11} \frac{\partial \psi}{\partial \theta_1} + \frac{\partial k^{11}}{\partial \theta_1} = 0 \Leftrightarrow \\ &\Leftrightarrow \frac{\partial \psi}{\partial \theta_1} = -\frac{1}{2} \frac{\frac{\partial k^{11}}{\partial \theta_1}}{k^{11}} = -\frac{1}{2} \frac{\partial \log k^{11}}{\partial \theta_1} = \frac{\partial \log(k^{11})^{-1/2}}{\partial \theta_1} \Rightarrow \end{aligned}$$

$$\begin{aligned}\Rightarrow \log p(\theta_1, \theta_2) &= \psi(\theta_1, \theta_2) = \log\left(\frac{1}{\sqrt{k^{11}}}\right) + c(\theta_2) = \log\sqrt{k_{11}} + c(\theta_2) \Rightarrow \\ &\Rightarrow p(\theta_1, \theta_2) \propto \sqrt{I_{11}(\theta_1, \theta_2)} \quad c(\theta_2)\end{aligned}$$

Observemos que la regla general de Jeffreys proporciona la densidad

$$p(\theta_1, \theta_2) \propto \sqrt{I_{11}(\theta_1, \theta_2)} \quad \sqrt{I_{22}(\theta_1, \theta_2)}$$

que verifica la ecuación de Peers si  $I_{22}(\theta_1, \theta_2)$  sólo depende de  $\theta_2$ .

El ejemplo 2 proporciona una idea general para la búsqueda de distribuciones a priori con más de un parámetro, consistente en la obtención de una reparametrización ortogonal en el sentido de Cox, D.R. and Reid, N. (1987) y entonces aplicar la regla de Jeffreys modificada con los nuevos parámetros. Este tipo de reparametrizaciones han sido utilizadas en diversas facetas de la Inferencia Estadística (ver, por ejemplo, Huzurbazar, V.S. (1950,1956), Jeffreys, H. (1961), Cox, D.R. and Reid, N. (1987), Nicolau, A. (1993), etc.). Cuando los parámetros son ortogonales, entonces la función de verosimilitud factoriza aproximadamente en sus componentes marginales, la primera dependiente sólo de  $\theta_1$  y la segunda de  $\theta_2$ , por lo que la información en cada componente puede considerarse por separado (ver, por ejemplo, Nicolau, A. (1993)).

Una ortogonalización total (matriz de información diagonal) no es en general posible, pues requiere del cumplimiento de un sistema en derivadas parciales con más ecuaciones que incógnitas. Sin embargo, relajando los requerimientos es posible en muchas situaciones obtener una reparametrización ortogonal u ortogonal por bloques adecuada al problema en cuestión.

En Jeffreys, H. (1961), podemos encontrar la ecuación obtenida por Huzurbazar, V.S. (1950) para el caso de dimensión 2 y bajo las condiciones de regularidad habituales. Siendo el parámetro del modelo  $\theta=(\theta_1,\theta_2)$ , tomamos  $\varphi=(\varphi_1, \varphi_2)$  en la forma  $\varphi_1= \theta_1$  y  $\varphi_2=\varphi_2(\theta_1,\theta_2)$ , con lo que la transformación inversa sería  $\theta_1=\varphi_1$  y  $\theta_2=\theta_2(\varphi_1, \varphi_2)$ . Aplicando la regla de la cadena, obtenemos:

$$\begin{aligned} E\left[\frac{\partial^2 \ell}{\partial \varphi_1 \partial \varphi_2}\right] &= \frac{\partial \theta_1}{\partial \varphi_1} \frac{\partial \theta_1}{\partial \varphi_2} E\left[\frac{\partial^2 \ell}{\partial \theta_1^2}\right] + \frac{\partial \theta_1}{\partial \varphi_1} \frac{\partial \theta_2}{\partial \varphi_2} E\left[\frac{\partial^2 \ell}{\partial \theta_1 \partial \theta_2}\right] + \\ &+ \frac{\partial \theta_2}{\partial \varphi_1} \frac{\partial \theta_1}{\partial \varphi_2} E\left[\frac{\partial^2 \ell}{\partial \theta_1 \partial \theta_2}\right] + \frac{\partial \theta_2}{\partial \varphi_1} \frac{\partial \theta_2}{\partial \varphi_2} E\left[\frac{\partial^2 \ell}{\partial \theta_2^2}\right] \end{aligned}$$

Se tiene  $\frac{\partial \theta_1}{\partial \varphi_1} = 1$  y  $\frac{\partial \theta_1}{\partial \varphi_2} = 0$  y además suponemos  $\frac{\partial \theta_2}{\partial \varphi_2} \neq 0$  (pues en otro

caso  $\varphi_2$  no intervendría en la reparametrización) con lo que la condición de ortogonalidad queda:

$$E\left[\frac{\partial^2 \ell}{\partial \theta_1 \partial \theta_2}\right] + \frac{\partial \theta_2}{\partial \varphi_1} E\left[\frac{\partial^2 \ell}{\partial \theta_2^2}\right] = 0$$

que es una ecuación diferencial en  $\theta_2(\varphi_1, \varphi_2)$ . Es importante la apreciación de que en la solución aparece una constante arbitraria con respecto a  $\varphi_1$  (es decir una función arbitraria  $g(\varphi_2)$ ). Según Jeffreys, esta puede tomarse igual a  $\varphi_2$ , aunque no argumenta nada a favor de dicha elección.

En Nicolau, A. (1993) se estudia este procedimiento de construcción de distribuciones a priori en el caso regular. Como se señaló anteriormente, al considerar que  $\theta_1$  es el parámetro de interés y  $\theta_2$  el perturbador, la distribución a priori que se obtiene es  $\sqrt{I_{11}(\theta)} c(\theta_2)$ . Nicolau dice ante esta situación: *“esencialmente, esta distribución es sólo la distribución a priori del parámetro de interés dado que el*

*parámetro perturbador está determinado: las propiedades de confianza son independientes de la distribución a priori del parámetro perturbador. Aunque en general la marginal a priori del parámetro perturbador no afecta al término de error en la probabilidad de cubrimiento de los intervalos Bayesianos resultantes, una elección sensata de la misma mejora numéricamente su comportamiento frecuentista”<sup>6</sup>.*

La interpretación es clara, pues

$$p(\theta_1, \theta_2) \propto \sqrt{I_{11}(\theta_1, \theta_2)} c(\theta_2) \Leftrightarrow p(\theta_1 | \theta_2) \propto \frac{\sqrt{I_{11}(\theta_1, \theta_2)} c(\theta_2)}{p(\theta_2)} \propto \sqrt{I_{11}(\theta_1, \theta_2)} k(\theta_2)$$

Hagamos notar también que, como indica Nicolás, la arbitrariedad en términos del parámetro perturbador refleja el hecho de que la reparametrización ortogonal no es única.

Según vemos, este procedimiento da buenos resultados cuando sólo uno de los parámetros es objeto de estudio. No obstante, en ciertos casos podemos obtener una distribución a priori conjunta con buenas propiedades frecuentistas para ambos parámetros. En efecto, por el procedimiento anterior podemos obtener  $p(\theta_1 | \theta_2)$  y  $p(\theta_2 | \theta_1)$  intercambiando los papeles del parámetro de interés y del perturbador. A continuación, buscamos una distribución conjunta  $p(\theta_1, \theta_2)$  compatible con ambas marginales.

En Barry, C.A. et al. (1992) podemos ver que las dos distribuciones condicionadas definen una distribución conjunta compatible con ambas si

---

<sup>6</sup> Nicolau (1993), pág. 378. El texto original es: “Essentially, it is only the conditional prior density of the parameter of interest given the nuisance parameter that is determined; the confidence property is independent of the prior density of the nuisance parameter. Although in general no marginal prior for the nuisance parameter can affect the error term in the coverage probability of the resulting Bayes intervals, a sensible choice of it improves numerically their frequency behaviour.”

---

$$\frac{p(\theta_1 | \theta_2)}{p(\theta_2 | \theta_1)} = \frac{h(\theta_1)}{h(\theta_2)}$$

y además la distribución conjunta compatible es

$$p(\theta_1, \theta_2) = p(\theta_1 | \theta_2)h(\theta_2)$$

Destaquemos un caso especialmente interesante. Siendo los parámetros ortogonales, si además se verifica

$$\begin{aligned} I_{11}(\theta_1, \theta_2) &= a(\theta_1)b(\theta_2) \\ I_{22}(\theta_1, \theta_2) &= a^*(\theta_1)b^*(\theta_2) \end{aligned}$$

entonces, una distribución a priori conjunta compatible con las dos condicionadas viene dada por:

$$p(\theta_1, \theta_2) \propto \sqrt{a(\theta_1)b^*(\theta_2)}$$

**Ejemplo 2.9 (continuación)** Consideremos la distribución  $N(\mu, \sigma)$ .

Para este caso, como sabemos, se tiene:

$$I(\mu, \sigma) = \begin{pmatrix} 1/\sigma^2 & 0 \\ 0 & 2/\sigma^2 \end{pmatrix}; \quad I^{-1}(\mu, \sigma) = \begin{pmatrix} \sigma^2 & 0 \\ 0 & \sigma^2/2 \end{pmatrix}$$

y así los parámetros son ortogonales y además

$$\begin{aligned} I_{11}(\mu, \sigma) = \frac{1}{\sigma^2} &\rightarrow a(\mu) = 1, \quad b(\sigma) = \frac{1}{\sigma^2} \\ I_{22}(\mu, \sigma) = \frac{2}{\sigma^2} &\rightarrow a^*(\mu) = 2, \quad b^*(\sigma) = \frac{1}{\sigma^2} \end{aligned}$$



con lo que la distribución conjunta es

$$p(\mu, \sigma) \propto \sqrt{a(\mu)b^*(\sigma)} \propto \frac{1}{\sigma}$$

• **La ecuación de Datta&Gosh.**

En Datta, G.S. and Gosh, J.K. (1995) se obtiene una ecuación similar a la de Peers, H.W. (1965) para una situación ligeramente distinta. Concretamente, se parte de  $X_1, \dots, X_n$  independientes e idénticamente distribuidas con función de densidad  $f(x, \theta)$  donde  $\theta = (\theta_1, \dots, \theta_p)^T$  y el parámetro de interés es  $\delta = t(\theta)$  donde  $t$  es una función real de clase  $C^2$  y se busca una densidad a priori  $p(\theta)$  con la siguiente propiedad:

$$P_\theta \left[ \frac{\sqrt{n}(t(\theta) - t(\hat{\theta}))}{\sqrt{b}} \leq z \right] = P_p \left[ \frac{\sqrt{n}(t(\theta) - t(\hat{\theta}))}{\sqrt{b}} \leq z \mid x \right] + O(1/n)$$

siendo  $\hat{\theta}$  la moda a posteriori o el estimador máximo verosímil de  $\theta$  y  $b$  la varianza a posteriori asintótica del numerador de la fracción.

Entonces, una distribución a priori  $p(\theta)$  cumple lo requerido si y sólo si verifica la ecuación diferencial:

$$\sum_{j=1}^p \frac{\partial}{\partial \theta_j} \{ \eta_j(\theta) p(\theta) \} = 0$$

donde  $\eta(\theta) = (\eta_1(\theta), \dots, \eta_p(\theta))^T$  viene dado por  $\eta(\theta) = \frac{I^{-1}(\theta)\nabla_t(\theta)}{\sqrt{\nabla_t(\theta)^T I^{-1}(\theta)\nabla_t(\theta)}}$

satisfaciendo la ecuación  $\eta(\theta^T)I(\theta)\eta(\theta) = 1 \quad \forall \theta$ . Hemos representado con  $I(\theta)$  a la matriz de información de Fisher para  $\theta$  con una única observación y

$$\nabla_t(\theta) = \left( \frac{\partial t(\theta)}{\partial \theta_1}, \dots, \frac{\partial t(\theta)}{\partial \theta_p} \right)^T.$$

Observemos que esta ecuación generaliza a la de Peers, H.W. (1965). En efecto, suponiendo (por simplicidad)  $p = 2$ , si consideramos  $t(\theta) = t(\theta_1, \theta_2) = \theta_1$  entonces puede comprobarse sin mayor dificultad que la ecuación de Datta and Gosh coincide con la de Peers.

**CAPÍTULO 3:    MODELOS NO REGULARES.**  
**GENERALIZACIÓN DE LA REGLA DE JEFFREYS.**

En este capítulo, vamos a intentar generalizar las ideas vistas con respecto a la elección de distribuciones a priori en modelos regulares, proporcionando un criterio de obtención de tales distribuciones que es aplicable también a modelos no regulares y que conduce en el caso regular a la misma distribución que se obtiene al aplicar la Regla de Jeffreys y la Regla de Jeffreys modificada en los casos uni y multiparamétrico respectivamente.

La base fundamental de la propuesta que haremos reside en los comentarios hechos por el propio Jeffreys, H. (1946). Partiendo de una medida de información basada en la discrepancia o distancia entre las funciones de densidad correspondientes a dos valores distintos del parámetro, Jeffreys, H. (1946) observa que, bajo condiciones de regularidad, para cada valor  $\theta$  del espacio paramétrico, la información o discrepancia entre las densidades correspondientes a  $\theta$  y a  $\theta + \Delta\theta$  es del orden de  $(\Delta\theta)^2$ , con constante de proporcionalidad igual a la información de Fisher del modelo en  $\theta$ , que es la función base para definir la distribución a priori no informativa. También señala que esto no siempre es correcto, poniendo como ejemplo el modelo uniforme con recorrido en  $[0, \theta]$ . Pondremos de manifiesto que tanto en este modelo uniforme como en todos los modelos no regulares usuales ocurre que la información correspondiente a  $\theta$  y a  $\theta + \Delta\theta$  es del orden de  $|\Delta\theta|$  y que la función del parámetro que actúa como constante de proporcionalidad es la que debe darnos la clave para definir la distribución a priori en el caso no regular. En el caso multiparamétrico, buscaremos reparametrizaciones ortogonales para buscar las distribuciones a priori por separado.

### 3.1 Elección de la distribución a priori. Caso uniparamétrico.

Como se indicó en el capítulo anterior, Jeffreys, H. (1946) propone su regla para la elección de distribuciones tras observar que la Información de Kullbar-Leiber (que nosotros hemos denotado por  $\tilde{J}$ ) es de orden 2 ante pequeñas variaciones en los parámetros; además, también indicamos que en este artículo se hace la observación de que en el modelo  $U(0,\theta)$ , dicha información es de orden 1 y que en este comentario reside la clave de la propuesta que vamos a efectuar. Antes de continuar, vamos a comprobar que se da esta propiedad. En realidad, vamos a trabajar con la Información usada por Akahira, M. and Takeuchi, K. (1991) (que nosotros hemos denotado por  $I$ ), pero ya hemos visto que ambas medidas son equivalentes en el sentido de que son del mismo orden ante pequeñas variaciones en los parámetros.

**Ejemplo 3.1:**  $X \sim U(0,\theta), \quad \theta \in (0,+\infty)$

En este caso, tenemos:  $f(x, \theta) = \frac{1}{\theta} \quad 0 \leq x \leq \theta$

1.  $h > 0$ :

$$\begin{aligned} I(\theta, \theta+h) &= -8 \log \int_0^\theta \left( \frac{1}{\theta} \frac{1}{\theta+h} \right)^{1/2} dx = -8 \log \left( \frac{\theta}{\sqrt{\theta(\theta+h)}} \right) \\ &= -8 \log \left( \sqrt{\frac{\theta}{\theta+h}} \right) = -4 \log \left( \frac{\theta}{\theta+h} \right) \end{aligned}$$

2.  $h < 0$ :

$$\begin{aligned} I(\theta, \theta+h) &= -8 \log \int_0^{\theta+h} \left( \frac{1}{\theta} \frac{1}{\theta+h} \right)^{1/2} dx = -8 \log \left( \frac{\theta+h}{\sqrt{(\theta+h)\theta}} \right) = \\ &= -8 \log \left( \sqrt{\frac{\theta+h}{\theta}} \right) = -4 \log \left( \frac{\theta+h}{\theta} \right) \end{aligned}$$

De esta forma, se tiene

$$\lim_{h \rightarrow 0} \frac{I(\theta, \theta + h)}{h^2} = +\infty.$$

Sin embargo,

$$\lim_{h \rightarrow 0^+} \frac{I(\theta, \theta + h)}{|h|} = \lim_{h \rightarrow 0^+} \frac{I(\theta, \theta + h)}{h} = \frac{4}{\theta}$$

$$\lim_{h \rightarrow 0^-} \frac{I(\theta, \theta + h)}{|h|} = \lim_{h \rightarrow 0^-} \frac{I(\theta, \theta + h)}{-h} = \frac{4}{\theta}$$

con lo que vemos que, efectivamente, la información es de orden 1.

Notemos que en el modelo uniforme, no se verifica la igualdad

$$\frac{\partial}{\partial \theta} \int f(x, \theta) dx = \int \frac{\partial}{\partial \theta} f(x, \theta) dx$$

ya que

$$\int \frac{\partial}{\partial \theta} f(x, \theta) dx = \int_0^\theta -\frac{1}{\theta^2} dx = -\frac{1}{\theta} \neq 0 = \frac{\partial}{\partial \theta} \int f(x, \theta) dx$$

y, consecuentemente,

$$E \left[ \frac{\partial \log f(x, \theta)}{\partial \theta} \right] = \int \frac{\partial}{\partial \theta} f(x, \theta) dx \neq 0.$$

Obsérvese que para que ambos límites puedan coincidir es necesario dividir por  $|h|$  en vez de por  $h$ , ya que la información es siempre positiva. Notemos que, en definitiva, estamos diciendo que

$$\lim_{h \rightarrow 0} \left| \frac{I(\theta, \theta + h)}{h} \right| = \frac{4}{\theta}$$

Hemos visto así que en todo caso se tiene  $\lim_{h \rightarrow 0} I(\theta, \theta + h) = 0$  (recordemos que  $I(\theta, \theta) = 0$ ), si bien la velocidad de esta convergencia es más rápida en los modelos regulares que en el modelo uniforme.

Ahora, teniendo en cuenta que la distribución a priori de Jeffreys es

$$\pi(\theta) \propto (I(\theta))^{1/2}$$

y que habíamos probado que en los modelos regulares se cumple

$$\lim_{h \rightarrow 0} \frac{I(\theta, \theta + h)}{h^2} = I(\theta)$$

podemos escribir la distribución de Jeffreys como:

$$\pi(\theta) \propto \left( \lim_{h \rightarrow 0} \frac{I(\theta, \theta + h)}{|h|^2} \right)^{1/2} \propto \lim_{h \rightarrow 0} \frac{\sqrt{I(\theta, \theta + h)}}{|h|}$$

Según el ejemplo visto anteriormente de la distribución uniforme, y puesto que la convergencia es del orden de  $|h|$ , vamos a proponer la siguiente distribución a priori para el parámetro  $\theta$ :

$$\pi(\theta) \propto \lim_{h \rightarrow 0} \frac{I(\theta, \theta + h)}{|h|}$$

supuesto que dicho límite exista.

De forma más genérica, la distribución a priori vendría dada por

$$\pi(\theta) \propto \lim_{h \rightarrow 0} \left( \frac{I(\theta, \theta + h)}{|h|^k} \right)^{1/k}$$

donde  $k > 0$  es tal que

$$\lim_{h \rightarrow 0} \frac{I(\theta, \theta + h)}{|h|^k} = C(\theta)$$

donde  $C(\theta)$  es una función que puede ser constante (pero no idénticamente nula ni infinito).

Evidentemente, dicho valor de  $k$ , caso de existir, es único. Sin embargo, es claro que la definición conlleva problemas de existencia en su planteamiento genérico. Podemos señalar que, si el modelo es regular entonces ya hemos visto que  $k=2$  y que con todos los modelos no regulares con los que hemos trabajado resulta ser  $k=1$ . No obstante, quedan abiertas las cuestiones de si existe un tal valor de  $k$  para cualquier modelo y si dicho valor puede ser no entero.

Un poco más adelante se verá que bajo ciertas condiciones ese límite existe y se justificará en términos más rigurosos el hecho de elevar a  $1/k$ . De momento, este hecho lo justificaremos simplemente por la analogía de que en el caso regular se eleva el límite a  $1/2$  por ser la convergencia del orden de  $|h|^2$ . Siendo la convergencia del orden de  $|h|^k$ , elevamos a  $1/k$ . Podríamos plantear también que en todos los casos hemos de buscar una medida que sea del orden de  $|h|$  y que la distribución a priori sería proporcional a la sensibilidad obtenida a partir de dicha medida (dividiendo por  $|h|$  y tomando límite cuando  $h$  tiende a cero). Así, en los modelos en los que

$I(\theta, \theta+h)$  es del orden de  $h^2$ , tomaríamos como medida  $\sqrt{I(\theta, \theta+h)}$  que sería entonces del orden de  $|h|$  y la distribución a priori resultante entonces quedaría  $\pi(\theta) \propto \sqrt{I(\theta)}$ . En realidad, en ambos casos la justificación es conseguir invarianza ante reparametrizaciones.

**Ejemplo 3.1: (continuación)**

En el ejemplo anterior de la distribución uniforme en  $(0, \theta)$ , como obtuvimos que el límite era  $4/\theta$ , tendremos que

$$\pi(\theta) \propto \frac{1}{\theta}$$

que es la distribución que se acepta comúnmente como no informativa para este modelo. Esta distribución a priori es la propuesta en Basulto, J.(1997).

**Ejemplo 3.2:**  $X \sim U(\theta - 1/2, \theta + 1/2)$   $\theta \in \mathbb{R}$

En este caso, tenemos:  $f(x, \theta) = 1$   $\theta - \frac{1}{2} \leq x \leq \theta + \frac{1}{2}$

$h > 0$ :

$$I(\theta, \theta+h) = -8 \log \int_{\theta+h-1/2}^{\theta+1/2} dx = -8 \log \left( \theta + \frac{1}{2} - \theta - h + \frac{1}{2} \right) = -8 \log(1-h)$$

2.  $h < 0$ :

$$I(\theta, \theta+h) = -8 \log \int_{\theta-1/2}^{\theta+h+1/2} dx = -8 \log \left( \theta + h + \frac{1}{2} - \theta + \frac{1}{2} \right) = -8 \log(1+h)$$



Puede comprobarse sin dificultad que también en este caso la velocidad de convergencia a 0 es del orden de  $|h|$  y que se verifica

$$\lim_{h \rightarrow 0} \frac{I(\theta, \theta + h)}{|h|} = 8$$

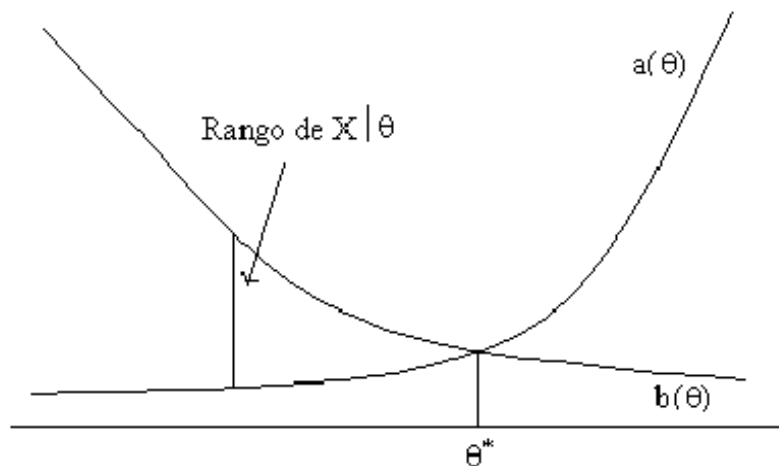
y por tanto, la distribución a priori sería para este caso:

$$\pi(\theta) \propto 1$$

**Ejemplo 3.3:** Consideremos la familia de modelos siguiente:

$$f(x, \theta) = \frac{1}{g(\theta)} \quad a(\theta) \leq x \leq b(\theta)$$

donde  $a(\theta)$  y  $b(\theta)$  pueden ser constantes, pero suponemos que si no lo son, entonces  $a(\theta)$  es estrictamente creciente y  $b(\theta)$  es estrictamente decreciente, siendo ambas funciones derivables. En este caso, se tendrá que  $g(\theta) = b(\theta) - a(\theta) \geq 0$  y además  $g(\theta)$  es estrictamente decreciente y diferenciable. La situación en estos modelos puede apreciarse mejor en la figura adjunta.



**Figura 3.1:** Rango del modelo  $f(x, \theta) = g^{-1}(\theta)$ ,  $a(\theta) \leq x \leq b(\theta)$

Señalemos que si suponemos  $a(\theta)$  estrictamente decreciente y  $b(\theta)$  estrictamente creciente el desarrollo del modelo es totalmente análogo. Esta familia puede verse en Basulto, J. (1997) y en Kosmas, K. F. (1990).

Consideremos  $\theta \in (-\infty, \theta^*)$ .

1.  $h > 0$ :

$$\begin{aligned} I(\theta, \theta + h) &= -8 \log \int_{a(\theta+h)}^{b(\theta+h)} \frac{dx}{\sqrt{g(\theta)g(\theta+h)}} = -8 \log \frac{b(\theta+h) - a(\theta+h)}{\sqrt{g(\theta)}\sqrt{g(\theta+h)}} = \\ &= -8 \log \frac{g(\theta+h)}{\sqrt{g(\theta)g(\theta+h)}} = -8 \log \sqrt{\frac{g(\theta+h)}{g(\theta)}} = -4 \log \frac{g(\theta+h)}{g(\theta)} = \\ &= -4(\log g(\theta+h) - \log g(\theta)) \end{aligned}$$

Por lo tanto, teniendo en cuenta que la función  $g$  es derivable, vamos a obtener:

$$\lim_{h \rightarrow 0^+} \frac{I(\theta, \theta + h)}{|h|} = -4 \lim_{h \rightarrow 0^+} \frac{\log g(\theta+h) - \log g(\theta)}{h} = -4 \frac{\partial \log g(\theta)}{\partial \theta}$$

2.  $h < 0$ :

$$\lim_{h \rightarrow 0^-} \frac{I(\theta, \theta + h)}{|h|} = (\dots) = -4 \frac{\partial \log g(\theta)}{\partial \theta}$$

Teniendo en cuenta que en este caso  $g(\theta)$  es decreciente, la distribución a priori que tomaríamos sería:

$$\pi(\alpha) \propto \left| \frac{\partial \log g(\theta)}{\partial \theta} \right|$$

Esta distribución coincide con la obtenida a través de la propuesta de Basulto, J.(1997), que además tiene unas propiedades muy interesantes como veremos posteriormente.

**Nota:** Puede comprobarse que el resultado es el mismo si consideramos el modelo

$$f(x, \theta) = \frac{c(x)}{g(\theta)} \quad a(\theta) \leq x \leq b(\theta)$$

imponiendo las condiciones de antes y además que  $c(x) \geq 0$  y que exista  $C(x)$  primitiva de  $c(x)$ . En esta situación se tendrá:

1.  $h > 0$ :

$$\begin{aligned} I(\theta, \theta + h) &= -8 \log \int_{a(\theta+h)}^{b(\theta+h)} \frac{c(x)}{\sqrt{g(\theta)g(\theta+h)}} dx = \\ &= -8 \log \frac{C(b(\theta+h)) - C(a(\theta+h))}{\sqrt{g(\theta)}\sqrt{g(\theta+h)}} = -8 \log \frac{g(\theta+h)}{\sqrt{g(\theta)g(\theta+h)}} = \dots \end{aligned}$$

y desde ese punto todo el procedimiento se repite.

### 3.1.1 Invarianza ante reparametrizaciones.

El intento de demostrar invarianza en una familia cualquiera probablemente sea excesivo. No sólo esta cuestión, sino simplemente ver cuándo existe el límite que hemos definido. Para intentar ver cualquier tipo de planteamiento general en modelos no regulares, probablemente sea imprescindible exigir unas mínimas condiciones de regularidad en el modelo.

Acerca de estas consideraciones, podemos citar el artículo de Ghosal, S. and Samanta, T. (1997). En él se estudia una familia no regular que verifica una serie de condiciones y para ella se estudia un desarrollo asintótico para la distribución a posteriori y la distribución límite de la misma. La situación que se plantea en este artículo es la que sigue.

Sean  $X_1, \dots, X_n$  independientes e idénticamente distribuidas con distribución  $P_\theta$  y con densidad  $f(x, \theta)$  respecto a la medida de Lebesgue en  $\mathbb{R}$ , donde  $\theta \in \Theta \subseteq \mathbb{R}$  siendo  $\Theta$  abierto. Suponemos que  $\forall \theta \in \Theta$ ,  $f(\bullet, \theta)$  es estrictamente positiva en un intervalo cerrado (acotado o no)  $S(\theta) = [a_1(\theta), a_2(\theta)]$  y vale cero fuera de él. Está permitido que uno de los extremos sea constante y puede ser más o menos infinito. Vamos a exigir también que los conjuntos  $S(\theta)$  sean bien crecientes o bien decrecientes en  $\theta$ . En lo que sigue se supone que son decrecientes (si son crecientes es todo casi igual), es decir,  $a_1(\theta)$  es creciente y  $a_2(\theta)$  decreciente; aún más, vamos a suponer que estas funciones son estrictamente monótonas y continuamente diferenciables a menos que sean constantes o valgan más o menos infinito.

Vamos a asumir además las siguientes hipótesis:

- (1) En el conjunto  $\{(x, \theta) : x \in S(\theta)\}$ ,  $f(x, \theta)$  es conjuntamente continua en  $(x, \theta)$ .
- (2) Para cada  $x$ ,  $\log(f(x, \theta))$  es dos veces diferenciable respecto a  $\theta$  en el conjunto  $\{a_1(\theta) < x < a_2(\theta)\}$ .

$$(3) \quad \forall \theta \in \Theta, \quad E_\theta \left[ \frac{\partial \log f(x, \theta)}{\partial \theta} \right] \text{ es finita.}$$

**Nota 1:** Estas son algunas de las hipótesis que se asumen en la familia de Ghosal and Samanta (no todas). Es importante resaltar que, según se afirma en Ghosal, S. and Samanta, T. (1997), en Ghosal, S. et al. (1995) se demuestra que las

condiciones exigidas son suficientes para la existencia de un límite en probabilidad de la distribución a posteriori; Ghosh, J.K. et al. (1994) aplican sus resultados a diversas familias que no cumplen estas condiciones y obtienen que para ellas no existe un límite. Así, en vista de estos dos artículos, la familia considerada es esencialmente la única para la que existe el límite a posteriori.

Estamos viendo así que, en cierta forma, para poder obtener resultados generales sobre modelos no regulares hay que exigir algunas condiciones de regularidad.

**Nota 2:** Los modelos más importantes que pertenecen a esta familia son:

1. La familia de localización:  $f(x, \theta) = f_0(x - \theta)$ ,  $\theta \in \mathbb{R}$ , donde  $f_0(z)$  es una densidad en el intervalo  $[0, +\infty)$ . En este caso,  $a_1(\theta) = \theta$  y  $a_2(\theta) = +\infty$ .

2.  $f(x, \theta) = \frac{c(x)}{g(\theta)}$   $a(\theta) \leq x \leq b(\theta)$  (es decir, la familia tratada en el ejemplo 3.3).

Observemos que esta familia incluye a gran cantidad de modelos, entre los que podemos destacar:

a) Modelos uniformes con soporte en:

i)  $[0, \theta]$ ,  $\theta > 0$ .

ii)  $[-\theta, \theta]$ ,  $\theta > 0$

iii)  $[\theta, 1/\theta]$ ,  $0 < \theta < 1$

b) Familia truncada:  $f(x, \theta) = \frac{g(x)}{G(\theta)}$ ,  $x > \theta$ , donde  $g(\bullet)$  es una densidad en  $(0, +\infty)$

y  $G(x) = \int_x^{+\infty} g(t) dt$ .

Podemos observar que el modelo de Pareto pertenece a la familia truncada, ya que en este caso la función de densidad es

$$f(x, \theta) = \frac{\alpha \theta^\alpha}{x^{1+\alpha}} \quad x > \theta$$

y así estamos en la situación descrita tomando  $g(x) = \alpha x^{1-\alpha}$  y  $G(\theta) = \int_x^{+\infty} g(t) dt = \theta^{-\alpha}$ .

Señalemos también que modelos tales como los uniformes en  $[\theta-1/2, \theta+1/2]$  o en  $[\theta, 2\theta]$  no están en esta familia por no ser los soportes ni crecientes ni decrecientes en  $\theta$ .

Asumiendo las condiciones de regularidad (1), (2) y (3) vamos a establecer una proposición análoga a la que se desarrolla en Akahira, M. and Takeuchi, K. (1991) y que va a servir para asegurar la existencia del límite y decir cuánto vale éste.

**Proposición.** Bajo las condiciones señaladas anteriormente, se tiene:

$$\lim_{h \rightarrow 0} \frac{I(\theta, \theta + h)}{|h|} = E \left[ \frac{\partial \log f(x, \theta)}{\partial \theta} \right]$$

donde se entiende que la derivada se lleva a cabo en el conjunto  $\{a_1(\theta) < x < a_2(\theta)\}$  que es donde tenemos garantizado que  $\log(f(x, \theta))$  es dos veces diferenciable.

Como consecuencia de la proposición, la distribución a priori que elegiríamos en este caso sería:

$$\pi(\theta) \propto E \left[ \frac{\partial \log f(x, \theta)}{\partial \theta} \right]$$

**Nota:** Si mantenemos todas las condiciones anteriores, pero suponiendo ahora que  $a_1(\theta)$  es estrictamente decreciente y que  $a_2(\theta)$  es estrictamente creciente, se puede repetir toda la demostración, obteniéndose:

$$\pi(\theta) \propto -E\left[\frac{\partial \log f(x, \theta)}{\partial \theta}\right]$$

En definitiva, en la familia de modelos estudiada en Ghosal and Samanta, obtenemos como distribución a priori:

$$\pi(\theta) \propto \left| E\left[\frac{\partial \log f(x, \theta)}{\partial \theta}\right] \right|$$

que, evidentemente, será una distribución válida siempre que dicha esperanza no sea idénticamente nula. Así, siempre que se verifique  $\frac{\partial}{\partial \theta} \int f(x, \theta) dx = 0 = \int \frac{\partial}{\partial \theta} f(x, \theta) dx$ , tendremos  $E\left[\frac{\partial \log f(x, \theta)}{\partial \theta}\right] = 0$  y por tanto esta definición no será válida.

Observemos también antes de pasar a la demostración que en el caso de la familia (2) tratada anteriormente como ejemplo se tiene

$$\frac{\partial \log f(x, \theta)}{\partial \theta} = -\frac{\partial \log g(\theta)}{\partial \theta} \Rightarrow \left| E\left[\frac{\partial \log f(x, \theta)}{\partial \theta}\right] \right| = \left| E\left[\frac{\partial \log g(\theta)}{\partial \theta}\right] \right|$$

con lo que el resultado obtenido en el ejemplo concuerda con el obtenido a partir de la proposición. Puede verse también fácilmente que en el modelo  $U(0, \theta)$  ambos resultados coinciden.

**Dmt:**

- Consideremos primero  $h < 0$  ( Esto conllevará que  $S(\theta) \subseteq S(\theta+h)$  ).

$$\begin{aligned}
 I(\theta, \theta+h) &= -8 \log \int_{S(\theta)} f(x, \theta)^{1/2} f(x, \theta+h)^{1/2} dx = \\
 &= -8 \log \int_{S(\theta)} \exp \left\{ \frac{1}{2} \log f(x, \theta) + \frac{1}{2} \log f(x, \theta+h) \right\} dx = \\
 &= -8 \log \int_{S(\theta)} \exp \left\{ \frac{1}{2} \log f(x, \theta) + \frac{1}{2} \left( \log f(x, \theta) + h \frac{\partial \log f(x, \theta)}{\partial \theta} + o(h) \right) \right\} dx = \\
 &= -8 \log \int_{S(\theta)} \exp \left\{ \log f(x, \theta) + \frac{h}{2} \frac{\partial \log f(x, \theta)}{\partial \theta} + o(h) \right\} dx = \\
 &= -8 \log \int_{S(\theta)} \left\{ \exp(\log f(x, \theta) + \left( \frac{h}{2} \frac{\partial \log f(x, \theta)}{\partial \theta} + o(h) \right)) \exp(\log f(x, \theta)) + o(h) \right\} dx = \\
 &= -8 \log \left\{ \int_{S(\theta)} f(x, \theta) dx + \frac{h}{2} \int_{S(\theta)} \frac{\partial \log f(x, \theta)}{\partial \theta} f(x, \theta) dx + o(h) \right\} = \\
 &= -8 \log \left\{ 1 + \frac{h}{2} E \left[ \frac{\partial \log f(x, \theta)}{\partial \theta} \right] + o(h) \right\} = -8 \left\{ \log 1 + \frac{h}{2} E \left[ \frac{\partial \log f(x, \theta)}{\partial \theta} \right] + o(h) \right\} = \\
 &= -4h E \left[ \frac{\partial \log f(x, \theta)}{\partial \theta} \right] + o(h) = 4|h| E \left[ \frac{\partial \log f(x, \theta)}{\partial \theta} \right] + o(h)
 \end{aligned}$$

Por tanto, en el caso  $h < 0$  tenemos directamente:

$$\lim_{h \rightarrow 0^-} \frac{I(\theta, \theta+h)}{|h|} = 4 E \left[ \frac{\partial \log f(x, \theta)}{\partial \theta} \right]$$

- Consideremos ahora  $h > 0$  ( Esto conllevará que  $S(\theta) \supseteq S(\theta+h)$  ).

Ahora, la idea es la misma, pero los desarrollos en serie los hacemos en torno al punto  $\theta+h$  para cada  $h$  dado, en vez de hacerlos en torno a  $\theta$ .



$$\begin{aligned}
 & \text{Así, vamos a obtener: } I(\theta, \theta+h) = (\dots) = \\
 & = -8 \log \left\{ \int_{S(\theta+h)} f(\theta+h) dx - \frac{h}{2} \int_{S(\theta+h)} \frac{\partial \log f(x, \theta)}{\partial \theta} \Big|_{\theta=\theta+h} f(x, \theta+h) dx + o(h) \right\} = \\
 & = (\dots) = 4h \int_{S(\theta+h)} \frac{\partial \log f(x, \theta)}{\partial \theta} \Big|_{\theta=\theta+h} f(x, \theta+h) dx + o(h) = \\
 & = (\dots) = 4|h| \int_{S(\theta+h)} \frac{\partial \log f(x, \theta)}{\partial \theta} \Big|_{\theta=\theta+h} f(x, \theta+h) dx + o(h)
 \end{aligned}$$

Así, en este caso obtendremos (teniendo en cuenta que  $f(x, \theta)$  es conjuntamente continua, que  $\log(f(x, \theta))$  es dos veces diferenciable y que por tanto su primera derivada es continua, que tanto  $a(\theta)$  como  $b(\theta)$  son continuas y que  $E\left[\frac{\partial \log f(x, \theta)}{\partial \theta}\right] < +\infty$ ) que:

$$\begin{aligned}
 \lim_{h \rightarrow 0^+} \frac{I(\theta, \theta+h)}{|h|} & = 4 \lim_{h \rightarrow 0^+} \int_{S(\theta+h)} \frac{\partial \log f(x, \theta)}{\partial \theta} \Big|_{\theta=\theta+h} f(x, \theta+h) dx = (*) \\
 & = (*) \quad 4 \int_{S(\theta)} \frac{\partial \log f(x, \theta)}{\partial \theta} f(x, \theta) dx = 4 E \left[ \frac{\partial \log f(x, \theta)}{\partial \theta} \right]
 \end{aligned}$$

quedando demostrada de esta forma la proposición.

La igualdad señalada con (\*) es una consecuencia del teorema de la convergencia dominada de Lebesgue (TCD). Veámoslo con un poco más de detalle.

$$\text{Definamos } g(x, \theta) = \frac{\partial \log f(x, \theta)}{\partial \theta} f(x, \theta)$$

Al ser  $E\left[\frac{\partial \log f(x, \theta)}{\partial \theta}\right] < +\infty$ , tenemos que ha de existir una función  $H(x)$  no negativa e integrable en  $\mathbb{R}$  y tal que cumpla  $|g(x, \theta)|_{\chi_{S(\theta)}}(x) \leq H(x) \quad \forall x \in \mathbb{R}$ , lo que es equivalente a  $|g(x, \theta)| \leq H(x) \quad \forall x \in S(\theta)$ .

Como estamos en el caso  $S(\theta+h) \subseteq S(\theta)$ , entonces tendremos también  $|g(x, \theta+h)| \leq H(x) \quad \forall x \in S(\theta+h)$ .

Además, dada la continuidad en  $\theta$  de  $g(x, \theta)$ , sabemos que  $\lim_{h \rightarrow 0^+} g(x, \theta+h) = g(x, \theta)$ .

Tenemos por tanto que se verifican a la vez:

$$\begin{cases} |g(x, \theta+h)| \chi_{S(\theta+h)}(x) \leq H(x) \chi_{S(\theta)}(x) \\ \lim_{h \rightarrow 0^+} \{g(x, \theta+h) \chi_{S(\theta+h)}(x)\} = g(x, \theta) \chi_{S(\theta)}(x) \end{cases}$$

y podemos aplicar el TCD, obteniendo:

$$\begin{aligned} \lim_{h \rightarrow 0^+} \int_{S(\theta+h)} \frac{\partial \log f(x, \theta)}{\partial \theta} \Big|_{\theta=\theta+h} f(x, \theta+h) dx &= \lim_{h \rightarrow 0^+} \int g(x, \theta+h) \chi_{S(\theta+h)}(x) dx = \\ &= \int \lim_{h \rightarrow 0^+} g(x, \theta+h) \chi_{S(\theta+h)}(x) dx = \int g(x, \theta) \chi_{S(\theta)}(x) dx = E \left[ \frac{\partial \log f(x, \theta)}{\partial \theta} \right] \end{aligned}$$

con lo que queda demostrado (\*).

♣

Obsérvese que una de las diferencias de esta demostración y la del caso regular es el cambio de soporte en las densidades al considerar  $\theta$  y  $\theta+h$  (habiéndose de tener en cuenta cuál de ellos es el menor). La diferencia quizás más significativa es que al hacer desarrollos de Taylor de  $\log(f(x, \theta))$ , tomamos orden 1 mientras que en el caso regular se toma orden 2. Esto se debe a que en el caso regular se verifica

$$E_{\theta} \left[ \frac{\partial \log f(x, \theta)}{\partial \theta} \right] = \int \frac{\partial \log f(x, \theta)}{\partial \theta} f(x, \theta) dx = 0$$

quedando así el comportamiento del desarrollo que se hace marcado por el término de segundo orden, mientras que en el caso no regular dicho comportamiento lo marca el término de primer orden al no ser su esperanza nula.

Esta proposición conlleva a que en la familia de Ghosal and Samanta la regla de construcción de distribuciones a priori sea invariante ante reparametrizaciones. En efecto, trivialmente tenemos que:

$$\pi(\varphi) \propto \left| E \left[ \frac{\partial \log f(x, \varphi)}{\partial \varphi} \right] \right| = \left| E \left[ \frac{\partial \log f(x, \theta)}{\partial \theta} \frac{\partial \theta}{\partial \varphi} \right] \right| = \left| E \left[ \frac{\partial \log f(x, \theta)}{\partial \theta} \right] \right| \left| \frac{\partial \theta}{\partial \varphi} \right|$$

y, en definitiva, se obtiene:

$$\pi(\varphi) \propto \pi(\theta) \left| \frac{\partial \theta}{\partial \varphi} \right|$$

que es precisamente la propiedad de invarianza buscada.

**Ejemplo 3.4:** Consideremos la familia de localización:  $f(x, \theta) = f_0(x - \theta)$ ,  $\theta \in \mathbb{R}$ , donde  $f_0(z)$  es una densidad en el intervalo  $[0, +\infty)$ .

En este caso, al aplicar el resultado obtenido en la proposición tenemos lo siguiente.

$$\frac{\partial \log f(x, \theta)}{\partial \theta} = \frac{\partial f_0(x - \theta)}{f_0(x - \theta) \partial \theta} = \frac{-f_0'(x - \theta)}{f_0(x - \theta)} \Rightarrow$$

$$\begin{aligned} E \left[ \frac{\partial \log f(x, \theta)}{\partial \theta} \right] &= \int_0^{+\infty} \frac{-f_0'(x-\theta)}{f_0(x-\theta)} f_0(x-\theta) dx = - \int_0^{+\infty} f_0'(z) dz = \\ &= -[f_0(+\infty) - f_0(0)] = f_0(0). \end{aligned}$$

De esta forma, la distribución a priori sería  $\pi(\theta) \propto 1$  siempre que  $f_0(0) \neq 0$ , en cuyo caso lo que ocurre es que la convergencia de la información es de orden dos, con lo que tendríamos que calcular el límite dividiendo por  $h^2$ .

### **3.1.2 Propiedades frecuentistas de los intervalos Bayesianos.**

En esta sección vamos a ver a través de diversos ejemplos que la distribución propuesta para modelos no regulares verifica propiedades de coincidencia numérica entre los intervalos bayesianos y clásicos (o al menos coincidencia asintótica) lo cual, junto con la invarianza ante reparametrizaciones, constituye una propiedad importante y deseable para una distribución a priori.

En los ejemplos que vamos a tratar se aprecia cómo en aquellos modelos en los que no existe un estadístico suficiente de la misma dimensión que el parámetro, los intervalos bayesianos coinciden numéricamente con los intervalos obtenidos a través del principio de condicionar al estadístico auxiliar.

1.  $f(x, \theta) = \frac{c(x)}{g(\theta)}$        $a(\theta) \leq x \leq b(\theta)$ , donde  $a(\theta)$  es estrictamente creciente y  $b(\theta)$  estrictamente decreciente.

Para esta familia, podemos ver en Basulto, J. (1997) que el estadístico máximo-verosímil es suficiente y que si  $g(-\infty) = +\infty$ , entonces, siendo  $r(\pi, \theta)$  la función de distribución a posteriori usando la distribución a priori propuesta en el presente trabajo, se verifica que  $r(\pi, \theta)$  es una variable aleatoria cuya distribución en el

muestreo es  $U(0,1)$ , lo que asegura que un intervalo unilateral bayesiano de probabilidad  $\alpha$  construido a partir de  $r(\pi,\theta)$  tiene también nivel de confianza **exactamente**  $\alpha$  (el intervalo Bayesiano coincidiría con el intervalo clásico basado en la cantidad pivotal  $r(\pi,\theta)$ ).

Si  $g(-\infty) = c > 0$ , entonces el resultado se tiene asintóticamente.

2.  $f(x, \theta) = f_0(x - \theta)$ ,  $\theta \in \mathbb{R}$ , donde  $f_0(z)$  es una densidad en  $[0, +\infty)$ .

En este modelo, no existe un estadístico suficiente de dimensión unitaria para el parámetro. En Basulto (1997) puede verse que los intervalos bayesianos obtenidos con la distribución a priori  $\pi(\theta) \propto 1$  coinciden con los obtenidos a partir de  $\hat{\theta} | c_2, \dots, c_n$  donde  $c_k = x_{(k)} - x_{(1)}$ ,  $k = 1 \dots n$  siendo  $(c_2, \dots, c_n)$  un estadístico auxiliar.

### 3.1.3 Subfamilias uniparamétricas de la familia $U(\alpha, \beta)$

Dentro de la familia uniforme,  $U(\alpha, \beta)$ , con espacio paramétrico  $\Theta = \{(\alpha, \beta); \alpha < \beta, \alpha \in \mathbb{R}, \beta \in \mathbb{R}\}$ , vamos a considerar las subfamilias de la forma  $U(\alpha, \alpha + h(\alpha))$ , con  $h(\alpha) > 0$  y  $h'(\alpha)$  continua, que son de interés para ilustrar la coincidencia numérica de intervalos bayesianos y clásicos.

Cuando  $h'(\alpha) \leq -1$ , el estadístico máximo verosímil  $\hat{\alpha}$  es suficiente y los intervalos bayesianos coinciden (se aproximan) cuando  $h(-\infty) = \infty$  ( $h(-\infty) = c$  finito) a intervalos de confianza clásicos (Basulto, 1997), puesto que estamos en la situación 1. del epígrafe anterior.

Cuando  $h'(\alpha) > -1$ , el estimador máximo verosímil de  $\alpha$ ,  $\hat{\alpha}$ , es insuficiente, y en este caso veremos que los intervalos bayesianos coinciden numéricamente con los intervalos obtenidos al condicionar al estadístico auxiliar.

En efecto, dentro de la subfamilia  $U(\alpha, \alpha + h(\alpha))$ , con  $h'(\alpha) > -1$ , para una muestra aleatoria  $\bar{x}_n$  de tamaño  $n$ , la máxima reducción que se consigue es al estadístico mínimo suficiente  $(y_1, y_n)$ , mínimo y máximo, respectivamente. Para  $h'(\alpha) > 0$ , el estimador máximo verosímil de  $\alpha$ ,  $\hat{\alpha}$ , verifica la ecuación  $\hat{\alpha} + h(\hat{\alpha}) = y_n$ , siendo  $\hat{\alpha}$  un estimador insuficiente para estimar  $\alpha$ . En este caso, y siguiendo a R.A. Fisher, debemos utilizar un estadístico auxiliar  $v$  tal que al mantener “fijo” este estadístico auxiliar, el estimador máximo verosímil  $\hat{\alpha}$  logra la máxima reducción de la muestra  $\bar{x}_n$ . Es decir, la función de verosimilitud está condicionada al estadístico auxiliar  $v$  y sólo le afecta  $\hat{\alpha}$ .

Para la búsqueda de un estadístico auxiliar seguimos a O.E. Bardorff-Nielsen, O.E. and Cox, D.R. (1994). Sumergimos la subfamilia  $U(\alpha, \alpha + h(\alpha))$ ,  $h'(\alpha) > -1$ , en la familia Uniforme  $U(\alpha, \beta)$ , y calculamos el logaritmo del cociente de verosimilitudes como en un contraste de hipótesis entre el modelo  $U(\alpha, \alpha + h(\alpha))$ ,  $h'(\alpha) > -1$  frente al modelo  $U(\alpha, \beta)$ , obteniendo

$$\ell(\hat{\alpha}, \hat{\beta} | \bar{x}_n) - \ell(\hat{\alpha} | \bar{x}_n) = n \log \left[ \frac{h(\hat{\alpha})}{r} \right]$$

donde  $\hat{\alpha} = y_1$  y  $\hat{\beta} = y_n$  son los estimadores máximo verosímiles de  $\alpha$  y  $\beta$  en el modelo  $U(\alpha, \beta)$ , y  $r = y_n - y_1$  es el recorrido.

El estadístico  $v = h(\hat{\alpha})/r$  posee propiedades que son de gran interés para valorar la reducción de la muestra  $\bar{x}_n$  al estimador insuficiente  $\hat{\alpha}$ . El gráfico siguiente

que corresponde al caso de que  $h'(\alpha) > 0$  nos será de utilidad para señalar las propiedades.

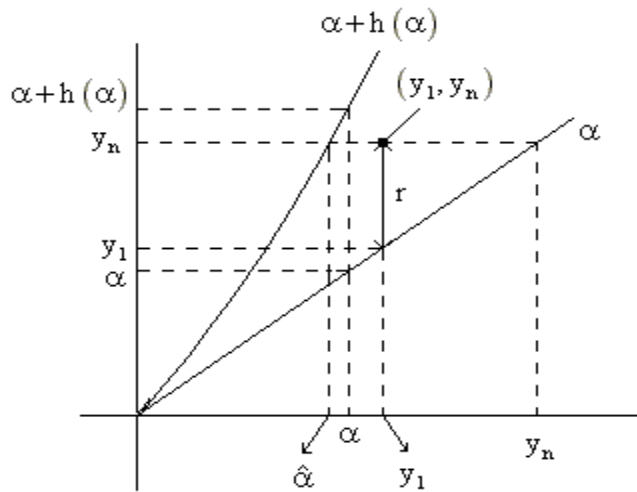


Figura 3.2: Modelo  $U(\alpha, \alpha + h(\alpha))$ ,  $h'(\alpha) > 0$

Siendo la función de verosimilitud del modelo  $U(\alpha, \alpha + h(\alpha))$ , con  $h'(\alpha) > 0$  igual a

$$L(\alpha | \bar{x}_n) \propto h^{-n}(\alpha) \quad , \quad \hat{\alpha} \leq \alpha \leq y_1$$

es claro, observando el gráfico que el estadístico posee las propiedades siguientes: **(i)**  $v \geq 1$ , **(ii)**  $v = 1$  siempre y cuando  $\hat{\alpha} = y_1$ , es decir, los datos  $(y_1, y_n)$  pertenecen a la curva  $\alpha + h(\alpha)$ . En consecuencia  $\hat{\alpha}$  es el único valor del parámetro  $\alpha$  compatible con la muestra  $\bar{x}_n$ , **(iii)** A medida de que  $v$  aumenta, la diferencia  $y_1 - \hat{\alpha}$  aumenta y la función de verosimilitud es más dispersa, es decir, la muestra es menos precisa, haciendo que la función de verosimilitud dependa de  $\hat{\alpha}$ , con lo que el estimador  $\hat{\alpha}$  aumenta en cantidad de suficiencia para estimar  $\alpha$ .

En la tabla 3.I recogemos las expresiones de la función de verosimilitud, el estimador máximo verosímil y la densidad del estadístico  $v$  para diferentes valores de  $h'(\alpha)$ .

	$L(\alpha   \bar{x}_n) \propto$	$\hat{\alpha} =$	$f(v   \alpha) =$
$h'(\alpha) > 0$	$\propto 1/h^n(\alpha)$ $\hat{\alpha} \leq \alpha \leq y_1$	$\hat{\alpha} + h(\hat{\alpha}) = y_n$	$= \frac{n(n-1)}{v^n h^n(\alpha)} \int_{\alpha^*}^{\alpha} h^{n-1}(z)(1+h'(z))dz$ $v = \frac{h(\alpha^*)}{\alpha^* + h(\alpha^*) - \alpha}$
$h'(\alpha) = 0$ $(h(\alpha) = a)$	$\propto 1/a^n$ $y_n - a \leq \alpha \leq y_1$	$= c(y_n - a) + (c-1)y$ $c \in [0, 1]$	$= \frac{n(n-1)}{v^n} \left[ 1 - \frac{1}{v} \right]$
$-1 < h'(\alpha) <$	$\propto 1/h^n(\alpha)$ $y_n^* \leq \alpha \leq y_1$ $y_n^* + h(y_n^*) = y_n$	$= y_1$	$= \frac{n(n-1)}{v^n h^n(\alpha)} \int_{\alpha}^{\alpha^{**}} h^{n-1}(z)dz$ $v = \frac{h(\alpha^{**})}{\alpha + h(\alpha) - \alpha^{**}}$

**Tabla 3.I**

**1. Modelos lineales:**  $h(\alpha) = a + b\alpha$

Cuando la función  $h(\alpha)$  es lineal, el estadístico  $v = h(\hat{\alpha})/r$  es auxiliar. En la tabla 3.II recogemos la función de densidad de  $v$  para  $h(\alpha) = a + b\alpha$  y diferentes valores de la pendiente  $b$ .

	$f(v) =$
$b > 0$ $a > -a/b$	$= \frac{(n-1)(b+1)}{b} \left[ \frac{1}{v^n} - \frac{1}{[(b+1)v - b]^n} \right]$
$b = 0$	$= \frac{(n-1)n}{v^n} \left[ 1 - \frac{1}{v} \right]$
$-1 < b < 0$	$= \frac{(n-1)}{ b v^n} \left[ 1 - \left( \frac{v(b+1)}{(b+v)} \right)^n \right]$

**Tabla 3.II**



Ahora, siguiendo a Fisher, R.A. (1956), debemos calcular la densidad condicionada del estimador  $\hat{\alpha}$  al estadístico auxiliar  $v$ . A partir de esta densidad condicionada hemos calculado la densidad condicionada de las cantidades pivotaes  $U = (a + b\hat{\alpha}) / (a + b\alpha)$  para los casos  $b > 0$  y  $-1 < b < 0$  y  $U = (\hat{\alpha} - (\alpha - ac + rc)) / (a - r)$  para el caso  $b=0$  a partir de las cuales se obtendrían los intervalos de confianza condicionados.

En la tabla 3.III recogemos las densidades condicionadas,  $f(\hat{\alpha} | \alpha, v)$  y las densidades de las cantidades pivotaes.

	$f(\hat{\alpha}   v, \alpha) =$	$f(u   v) =$
$b > 0$ $a > -a/b$	$= \left( \frac{a + b\hat{\alpha}}{a + b\alpha} \right)^{n-1} \frac{nb}{(a + b\alpha)} \frac{1}{1 - (Av)^n}$ $\alpha^* \leq \hat{\alpha} \leq \alpha$ $\alpha^* \text{ t.q. } v = \frac{a + b\alpha^*}{\alpha^* + a + b\alpha^* - \alpha}$ $A = [(b+1)v - b]^{-1}$	$= \frac{nu^{n-1}}{1 - (Av)^n}$ $Av \leq u \leq 1$
$b = 0$	$= (a - r)^{-1}$ $\alpha - ca + cr \leq \hat{\alpha} \leq \alpha + a - ca - (1 - c)r$	$= 1, \quad 0 \leq u \leq 1$
$-1 < b < 0$	$= \left( \frac{a + b\hat{\alpha}}{a + b\alpha} \right)^{n-1} \frac{n b }{(a + b\alpha)} \frac{1}{1 - \left( \frac{v(b+1)}{v+b} \right)^n}$ $\alpha \leq \hat{\alpha} \leq \alpha^{**}$ $\alpha^{**} \text{ t.q. } v = \frac{a + b\alpha^{**}}{\alpha + (a + b\alpha) - \alpha^{**}}$	$= \frac{nu^{n-1}}{1 - \left( \frac{v(b+1)}{v+b} \right)^n}$ $\frac{v(b+1)}{v+b} \leq u \leq 1$

Tabla 3.III

Desde el punto de vista bayesiano hemos calculado las distribuciones a priori no informativas según la propuesta del presente trabajo y las densidades a posteriori. Estos resultados están recogidos en la tabla 3.IV para diferentes valores de la pendiente  $b$ .

	$\pi(\alpha) \propto$	$\pi(\alpha   \bar{x}_n) =$
$b > 0$ $a > -a/b$	$\propto (a + b\alpha)^{-1}$ $\alpha > -a/b$	$= \frac{nb(a + b\hat{\alpha})^n}{1 - (Av)^n} (a + b\alpha)^{-(n+1)}$ $\hat{\alpha} \leq \alpha \leq y_1$
$b = 0$	$\propto 1$	$= (a - r)^{-1}$ $\hat{\alpha} \leq \alpha \leq y_1$
$-1 < b < 0$	$\propto (a + b\alpha)^{-1}$ $\alpha < -a/b$	$= \frac{n b (a + b\hat{\alpha})^n}{1 - M^n} (a + b\alpha)^{-(n+1)}$ $\hat{\alpha} \leq \alpha \leq y_1$ $M = \frac{a + b\hat{\alpha}}{a + by_n^*}$

**Tabla 3.IV**

Puede probarse en todos los casos que los intervalos bayesianos de probabilidad  $\gamma$  del tipo  $\{\alpha \leq \alpha(\bar{x})\}$  coinciden con los intervalos de confianza condicionados de nivel de confianza  $\gamma$ .

Veamos dicho resultado, por ejemplo, para el caso  $b > 0$ .

Sea  $\alpha(\bar{x})$  tal que  $(\hat{\alpha}, \alpha(\bar{x})]$  es un intervalo bayesiano de probabilidad  $\gamma$  obtenido a partir de la distribución a posteriori recogida en la tabla 3.IV. Entonces, se ha de cumplir

$$\gamma = P[\alpha \leq \alpha(\bar{x}) | \bar{x}] = \int_{\hat{\alpha}}^{\alpha(\bar{x})} \pi(\alpha | \bar{x}) d\alpha = \left[ \frac{(a + b\hat{\alpha})^n nb}{1 - (1/1 + b(1-1/v))^n} \frac{1}{b} \frac{(a + b\alpha)^{-n}}{-n} \right]_{\hat{\alpha}}^{\alpha(\bar{x})}$$

Llamando  $B = 1 - (1/1 + b(1-1/v))^n$ , tenemos

$$\gamma = B(a + b\hat{\alpha})^n \left\{ (a + b\hat{\alpha})^{-n} - (a + b\alpha(\bar{x}))^{-n} \right\} = B \left[ 1 - \left( \frac{a + b\hat{\alpha}}{a + b\alpha(\bar{x})} \right)^n \right]$$

Despejando  $\alpha(\bar{x})$  obtenemos finalmente

$$\alpha(\bar{x}) = \frac{(a + b\hat{\alpha})(1 - \gamma/B)^{-1/n} - a}{b}$$

Veamos ahora que el intervalo  $(\hat{\alpha}, \alpha(\bar{x}))$  obtenido tiene nivel de confianza (condicionando a  $v$ ) igual a  $\gamma$ . Del proceso que vamos a seguir puede deducirse también que el intervalo bayesiano coincide con el obtenido a partir de la cantidad pivotal  $U$  (siempre condicionando a  $v$ ).

$$\begin{aligned} P[\alpha \leq \alpha(\bar{x}) | v, \alpha] &= P[a + b\alpha \leq a + b\alpha(\bar{x}) | v, \alpha] = \\ &= P[a + b\alpha \leq (a + b\hat{\alpha}) \geq (1 - \gamma/B)^{-1/n} | v, \alpha] = \\ &= P\left[ \frac{a + b\hat{\alpha}}{a + b\alpha} \geq (1 - \gamma/B)^{1/n} \middle| v, \alpha \right] = P\left[ U \geq (1 - \gamma/B)^{1/n} \middle| v, \alpha \right] = 1 - F_{U|v} \left( (1 - \gamma/B)^{1/n} \right) \end{aligned}$$

Ahora, puesto que la función de densidad de  $U|v$  es

$$f_{U|v}(u | v) = \frac{nu^{n-1}}{1 - (Av)^n}, \quad Av \leq u \leq 1$$

la función de Distribución condicionada es

$$F_{U|v}(u | v) = \int_{Av}^u \frac{nt^{n-1}}{1-(Av)^n} dt = \frac{u^n - (Av)^n}{1-(Av)^n}$$

y por tanto,

$$\begin{aligned} P[\alpha \leq \alpha(\bar{x}) | v, \alpha] &= 1 - \frac{(1-\gamma/B) - (Av)^n}{1-(Av)^n} = \frac{1-(Av)^n - (1-\gamma/B) + (Av)^n}{1-(Av)^n} = \\ &= \frac{\gamma/B}{1-(Av)^n} = \gamma \frac{1}{B(1-(Av)^n)} \end{aligned}$$

Teniendo en cuenta que  $B = \frac{1}{1 - \left(\frac{1}{1+b(1-1/v)}\right)^n}$  y que  $A = \frac{1}{v(1+b)-b}$ ,

obtenemos:

$$B = \frac{1}{1 - \left(\frac{1}{1+b(1-1/v)}\right)^n} = \frac{1}{1 - \left(\frac{1}{v(1+b)-b}\right)^n} = \frac{1}{1-(Av)^n}$$

por lo que finalmente obtenemos

$$P[\alpha \leq \alpha(\bar{x}) | v, \alpha] = \gamma \frac{1-(Av)^n}{1-(Av)^n} = \gamma$$

que era el resultado buscado.

♣

## 2. Modelos no lineales.

Consideraremos ahora el caso no lineal, suponiendo, sin pérdida de generalidad que  $h'(\alpha) > 0$ .

En este caso, el resultado anterior se obtiene forma local, ya que la distribución del estadístico va a depender de  $\alpha$  a través de la función  $h'(\alpha)$ , de forma que en un entorno del valor  $\alpha_0$  en el cual podamos suponer que  $h'(\alpha)$  es constante, se verificarían los resultados del caso lineal. Vamos a enumerar a continuación los pasos que llevan a este resultado.

En primer lugar, observemos que la sucesión de vectores aleatorios

$$(t_{1,n}, t_{2,n}) = (n(\alpha - \hat{\alpha}), n(y_1 - \alpha))$$

converge en ley al vector aleatorio  $(t_1, t_2)$ , donde  $t_1$  y  $t_2$  siguen densidades exponenciales de parámetros  $(1+h'(\alpha))/h(\alpha)$  y  $1/h(\alpha)$  respectivamente, siendo además  $t_1$  y  $t_2$  independientes (ver Lehmann, E.L., 1999; pág. 282, Example 5.1.2.).

Observemos también que se verifica la igualdad  $n(t-1) = (t_{1,n} + t_{2,n})/r$ , puesto que

$$\begin{aligned} n(t-1) &= n\left(\frac{h(\hat{\alpha})}{r} - 1\right) = n\left(\frac{h(\hat{\alpha}) - r}{r}\right) = n\frac{y_n - \hat{\alpha} - r}{r} = \\ &= n\frac{y_1 - \hat{\alpha}}{r} = \frac{1}{r}(n(\alpha - \hat{\alpha}) + n(y_1 - \alpha)) = \frac{t_{1,n} + t_{2,n}}{r} \end{aligned}$$

Ahora, en virtud del teorema de Slutsky (ver Lehman, E.L., 1999; pág. 283), obtendremos que  $n(t-1)$  converge en ley a la variable aleatoria  $U = \frac{t_1 + t_2}{h(\alpha)}$ , con lo cual  $U$  es suma de  $\frac{t_1}{h(\alpha)} \sim \text{Exp}(1+h'(\alpha))$  y de  $\frac{t_2}{h(\alpha)} \sim \text{Exp}(1)$  que además son independientes. Así,  $n(t-1)$  converge en ley a una densidad que depende exclusivamente de  $h'(\alpha)$ .

De esta forma, si aproximamos  $h(\alpha)$  por una poligonal, en cada tramo de la misma, al ser la derivada constante, el estadístico  $t = h(\hat{\alpha})/r$  sería asintóticamente auxiliar. En dicho modelo poligonal, decimos que el estadístico es localmente auxiliar. De esta forma, los resultados obtenidos para el caso lineal se verificarían ahora de forma local.

### **3.2 Elección de la distribución a priori. Caso multiparamétrico.**

En el caso regular, cuando el parámetro  $\theta$  es un vector de  $\mathbb{R}^k$ , la relación de la medida de información que estamos estudiando con la matriz de Fisher se generaliza sin problemas. En concreto,

$$I(\theta, \theta + h) \propto h^t I(\theta) h + o(\|h\|^2)$$

Como  $I(\theta)$  es una matriz definida positiva, bajo distintos enfoques (Jeffreys, H.(1946), Kass, R.E. (1989), George, E.I. and McCulloch, R.(1993)), se adopta una distribución proporcional a  $\sqrt{|I(\theta)|}$  que asegura invarianza ante reparametrizaciones. Esta es la conocida regla general de Jeffreys que, como vimos en el capítulo anterior,

presenta deficiencias que hacen que el propio Jeffreys la modificase y que actualmente no sea la opción usada habitualmente.

La matriz de Información puede definirse en el caso no regular (aunque su significado no fuese el mismo), pues siempre podemos calcular

$$I(\theta) = E \left[ \left( \frac{\partial \ell}{\partial \theta} \right) \left( \frac{\partial \ell}{\partial \theta} \right)^t \right]$$

En el caso regular esta matriz también puede escribirse como

$$(I(\theta))_{ij} = -E \left[ \frac{\partial^2 \ell}{\partial \theta_i \partial \theta_j} \right]$$

Sin embargo, en el caso no regular no tenemos garantizado que  $I(\theta)$  sea definida positiva lo que conlleva a que sea posible que  $|I(\theta)| = 0$ , con lo que no vale la opción de tomar una distribución proporcional a  $\sqrt{|I(\theta)|}$  que asegura invarianza ante reparametrizaciones.

**Ejemplo 3.5:** Sea  $X \sim U(a,b)$ ,  $a < b$ .

En este ejemplo, tenemos:

$$f(x, a, b) = \frac{1}{b-a} \quad a \leq x \leq b$$

$$\ell(x | a, b) = \log f(x | a, b) = -\log(b-a)$$

$$\frac{\partial \ell}{\partial a} = \frac{1}{b-a}$$

$$\frac{\partial \ell}{\partial b} = -\frac{1}{b-a}$$

$$I(a, b) = E \left[ \begin{pmatrix} 1/(b-a) \\ -1/(b-a) \end{pmatrix} \begin{pmatrix} 1/(b-a) & -1/(b-a) \end{pmatrix} \right] = \begin{pmatrix} 1/(b-a)^2 & -1/(b-a)^2 \\ -1/(b-a)^2 & 1/(b-a)^2 \end{pmatrix}$$

y por tanto  $|I(a, b)| = 0$ .

No obstante, según vemos en el caso regular, parece que el criterio de búsqueda de invarianza general en el caso multiparamétrico no es la mejor opción.

Así, vamos a adoptar en el caso no regular también el criterio de que si la matriz de información es diagonal por bloques, consideraremos a los parámetros independientes a priori. En este caso si los parámetros no regulares unidimensionales son independientes del resto, podemos aplicar la regla anterior y luego obtener la distribución conjunta como producto de las marginales.

Un caso importante son los modelos de la forma  $\{f(x, \alpha, \theta), x > \theta\}$ , en los que  $\theta \in \mathbb{R}$ ,  $\alpha \in \mathbb{R}^k$  y para  $\theta$  fijo el modelo es regular (diremos que  $\alpha$  es un parámetro regular o que el modelo es regular con respecto a  $\alpha$ ) y para  $\alpha$  fijo,  $\theta$  es un parámetro no regular. En este caso, podemos incluir al modelo de Pareto y a los modelos truncados como ejemplos relevantes.

En esta situación,

$$I(\theta) = \begin{pmatrix} E\left[\left(\frac{\partial \ell}{\partial \alpha}\right)^2\right] & E\left[\frac{\partial \ell}{\partial \alpha} \frac{\partial \ell}{\partial \theta}\right] \\ E\left[\frac{\partial \ell}{\partial \alpha} \frac{\partial \ell}{\partial \theta}\right] & E\left[\left(\frac{\partial \ell}{\partial \theta}\right)^2\right] \end{pmatrix}$$

y así podremos considerar al parámetro  $\theta$  independiente de  $\alpha$  si  $E\left[\frac{\partial \ell}{\partial \alpha} \frac{\partial \ell}{\partial \theta}\right] = 0$ .

Una condición suficiente para ello es que  $\frac{\partial \ell}{\partial \theta}$  no dependa de  $x$ , ya que entonces, al ser



el modelo regular con respecto a  $\alpha$  se tiene  $E\left[\frac{\partial \ell}{\partial \alpha}\right] = 0$  y consecuentemente

$$E\left[\frac{\partial \ell}{\partial \alpha} \frac{\partial \ell}{\partial \theta}\right] = \frac{\partial \ell}{\partial \theta} E\left[\frac{\partial \ell}{\partial \alpha}\right] = 0.$$

**Ejemplo 3.6:** Sea  $X \sim P(\alpha, \theta)$ .

Para este modelo se tiene:

$$\begin{aligned} f(x, \alpha, \theta) &= \frac{\alpha \theta^\alpha}{x^{1+\alpha}} \quad x > \theta \\ \log f(x, \alpha, \theta) &= \log \alpha + \alpha \log \theta - (1 + \alpha) \log x \\ \frac{\partial \ell}{\partial \alpha} &= \frac{1}{\alpha} + \log \theta - \log x \\ \frac{\partial \ell}{\partial \theta} &= \frac{\alpha}{\theta} \end{aligned}$$

Por tanto, podemos considerar que los dos parámetros del modelo son independientes a priori.

Para  $\theta$  fijo, el modelo es regular y aplicamos entonces la regla de Jeffreys, con lo que

$$\pi(\alpha | \theta) \propto \sqrt{-E\left[\frac{\partial^2 \ell}{\partial \alpha^2}\right]} = \frac{1}{\alpha} c(\theta)$$

Para  $\alpha$  fijo, aplicamos la regla propuesta anteriormente para parámetros no regulares unidimensionales, obteniendo:

$$\pi(\theta | \alpha) \propto \left| E\left[\frac{\partial \ell}{\partial \theta}\right] \right| = \frac{1}{\theta} k(\alpha)$$

De esta forma, la distribución a priori conjunta compatible con estas condicionadas sería:

$$\pi(\alpha, \theta) \propto \frac{1}{\alpha \theta}$$

que coincide con la distribución a priori imparcial propuesta por Basulto, J.(1997).

**Ejemplo 3.7:** Consideremos  $X \sim f(x, \alpha)$  modelo regular. Sea  $F(x, \alpha)$  la función de distribución del modelo.

Definimos ahora el modelo truncado generado a partir de  $X$  como:

$$Y = X \quad \text{si} \quad x \leq \theta$$

Tendremos entonces:

$$\begin{aligned} g(y, \alpha, \theta) &= \frac{f(y, \alpha)}{F(\theta, \alpha)} \quad y \leq \theta \\ \log g(y, \alpha, \theta) &= \log f(y, \alpha) - \log F(\theta, \alpha) \\ \frac{\partial \log g}{\partial \theta} &= -\frac{\partial \log F(\theta, \alpha)}{\partial \theta} = -\frac{\frac{\partial F(\theta, \alpha)}{\partial \theta}}{F(\theta, \alpha)} = -\frac{f(\theta, \alpha)}{F(\theta, \alpha)} \end{aligned}$$

Como  $\frac{\partial \log g}{\partial \theta}$  no depende de  $x$ , podemos considerar ambos parámetros independientes a priori. Como además fijado  $\theta$  el modelo es regular, la distribución a priori conjunta la obtendremos como el producto de las marginales, que serían

$$\pi(\alpha | \theta) \propto \sqrt{-E \left[ \frac{\partial^2 \log g}{\partial \alpha^2} \right]} c(\theta)$$

$$\pi(\theta|\alpha) \propto \frac{f(\theta, \alpha)}{F(\theta, \alpha)} k(\alpha)$$

**Observación:** Si el truncamiento es del tipo  $X > \theta$ , entonces todo el procedimiento es muy similar, obteniéndose:

$$g(y, \alpha, \theta) = \frac{f(y, \alpha)}{1 - F(\theta, \alpha)} \quad y > \theta$$

$$\frac{\partial \log g}{\partial \theta} = \frac{f(\theta, \alpha)}{1 - F(\theta, \alpha)}$$

y la distribución marginal de  $\theta$  sería entonces

$$\pi(\theta|\alpha) \propto \frac{f(\theta, \alpha)}{1 - F(\theta, \alpha)} k(\alpha)$$

Notemos que el modelo de Pareto puede ajustarse a este esquema haciendo un truncamiento de la forma  $X > \theta$  en el modelo regular dado por  $X$ ,  $f(x|\alpha) = \alpha x^{-(1+\alpha)}$   $x > 0$ . En efecto, sea  $Y=X$  si  $X > \theta$ . Entonces,

$$F(\theta, \alpha) = \int_0^\theta f(t|\alpha) dt = \int_0^\theta \alpha t^{-(1+\alpha)} dt = 1 - \theta^{-\alpha}$$

y de esta forma

$$g(y, \alpha, \theta) = \frac{f(y, \alpha)}{1 - F(\theta, \alpha)} \quad \text{si } y > \theta \Rightarrow$$

$$\Rightarrow g(y, \alpha, \theta) = \frac{\alpha y^{-(1+\alpha)}}{\theta^{-\alpha}} \quad \text{si } y > \theta \Rightarrow Y \sim P(\alpha; \theta)$$

Aplicando la fórmula obtenida para los modelos truncados, tenemos

$$\left. \begin{aligned} - E \left[ \frac{\partial^2 \ell}{\partial \alpha^2} \right] &= \frac{1}{\alpha^2} \Rightarrow \pi(\alpha | \theta) \propto \frac{1}{\alpha} c(\theta) \\ \pi(\theta) &\propto \frac{f(\theta | \alpha)}{1 - F(\theta | \alpha)} = \propto \frac{1}{\theta} k(\alpha) \end{aligned} \right\} \Rightarrow \pi(\alpha; \theta) \propto \frac{1}{\alpha \theta}$$

que es la distribución obtenida anteriormente.

Así, una posible solución general para obtener distribuciones a priori en modelos con más de un parámetro de los cuales al menos uno es no regular sería buscar una reparametrización ortogonal adecuada que nos permitiese aplicar las propuestas unidimensionales y obtener la distribución conjunta como producto de las marginales. Este procedimiento general, como vimos en el capítulo 2, ha sido utilizado por varios autores en el caso regular, de los que podemos destacar a Tibshirani, R. (1989) y a Nicolau, A. (1993). Además, en el caso habitual de que el parámetro no regular  $\theta$  sea unidimensional, esto no supone ninguna restricción, pues siempre es posible una reparametrización en la que el parámetro regular  $\alpha$  sea ortogonal a  $\theta$  (Cox, D.R. and Reid, N. (1987)).

Como se vio en el capítulo anterior, en Jeffreys, H. (1961) podemos encontrar la ecuación obtenida por Huzurbazar, V.S. (1950) para el caso de dimensión 2 y bajo las condiciones de regularidad habituales. Siendo el parámetro del modelo  $\theta=(\theta_1, \theta_2)$  y tomando  $\varphi=(\varphi_1, \varphi_2)$  en la forma  $\varphi_1=\theta_1$  y  $\varphi_2=\varphi_2(\theta_1, \theta_2)$ , la condición de ortogonalidad quedaba:

$$E \left[ \frac{\partial^2 \ell}{\partial \theta_1 \partial \theta_2} \right] + \frac{\partial \theta_2}{\partial \varphi_1} E \left[ \frac{\partial^2 \ell}{\partial \theta_2^2} \right] = 0$$

Como señalamos, esta condición es aplicable a modelos regulares. No obstante, es fácil obtener una condición parecida en el caso de parámetros no regulares. En efecto:

$$\begin{aligned}
 E\left[\frac{\partial \ell}{\partial \varphi_1} \frac{\partial \ell}{\partial \varphi_2}\right] &= E\left[\left(\frac{\partial \ell}{\partial \theta_1} \frac{\partial \theta_1}{\partial \varphi_1} + \frac{\partial \ell}{\partial \theta_2} \frac{\partial \theta_2}{\partial \varphi_1}\right) \left(\frac{\partial \ell}{\partial \theta_1} \frac{\partial \theta_1}{\partial \varphi_2} + \frac{\partial \ell}{\partial \theta_2} \frac{\partial \theta_2}{\partial \varphi_2}\right)\right] = \\
 &= \frac{\partial \theta_1}{\partial \varphi_1} \frac{\partial \theta_1}{\partial \varphi_2} E\left[\left(\frac{\partial \ell}{\partial \theta_1}\right)^2\right] + \frac{\partial \theta_1}{\partial \varphi_1} \frac{\partial \theta_2}{\partial \varphi_2} E\left[\frac{\partial \ell}{\partial \theta_1} \frac{\partial \ell}{\partial \theta_2}\right] + \\
 &+ \frac{\partial \theta_2}{\partial \varphi_1} \frac{\partial \theta_1}{\partial \varphi_2} E\left[\frac{\partial \ell}{\partial \theta_1} \frac{\partial \ell}{\partial \theta_2}\right] + \frac{\partial \theta_2}{\partial \varphi_1} \frac{\partial \theta_2}{\partial \varphi_2} E\left[\left(\frac{\partial \ell}{\partial \theta_2}\right)^2\right]
 \end{aligned}$$

Como  $\frac{\partial \theta_1}{\partial \varphi_1} = 1$  y  $\frac{\partial \theta_1}{\partial \varphi_2} = 0$  y además suponemos  $\frac{\partial \theta_2}{\partial \varphi_2} \neq 0$ , la condición se

traduce en:

$$E\left[\frac{\partial \ell}{\partial \theta_1} \frac{\partial \ell}{\partial \theta_2}\right] + \frac{\partial \theta_2}{\partial \varphi_1} E\left[\left(\frac{\partial \ell}{\partial \theta_2}\right)^2\right] = 0$$

A continuación, vamos a aplicar este procedimiento a un caso especialmente relevante que es el modelo uniforme con ambos extremos desconocidos.

**Ejemplo 3.5: (Continuación)** Consideremos  $X \sim U(a,b)$ , donde  $-\infty < a < b < +\infty$ .

Ya hemos visto que en este ejemplo se tiene

$$\begin{aligned}f(x | a, b) &= \frac{1}{b-a} \quad a \leq x \leq b \\ \ell(x | a, b) &= \log f(x | a, b) = -\log(b-a) \\ \frac{\partial \ell}{\partial a} &= \frac{1}{b-a} \\ \frac{\partial \ell}{\partial b} &= -\frac{1}{b-a} \\ -E \left[ \frac{\partial \ell}{\partial a} \frac{\partial \ell}{\partial b} \right] &= E \left[ \left( \frac{\partial \ell}{\partial b} \right)^2 \right] = \frac{1}{(b-a)^2}\end{aligned}$$

Siendo los nuevos parámetros  $\varphi_1$  y  $\varphi_2$  donde  $\varphi_1 = a$ , la condición de ortogonalidad se traduce en:

$$\frac{1}{(b-a)^2} \left( -1 + \frac{\partial b}{\partial \varphi_1} \right) = 0 \Leftrightarrow \frac{\partial b}{\partial \varphi_1} = 1 \Leftrightarrow b = \varphi_1 + c(\varphi_2)$$

Usando la recomendación de Jeffreys, tomamos  $c(\varphi_2) = \varphi_2$  y de esta forma la reparametrización efectuada es

$$\begin{cases} a = \varphi_1 \\ b = \varphi_1 + \varphi_2 \end{cases} \Leftrightarrow \begin{cases} \varphi_1 = a \\ \varphi_2 = b - a \end{cases}$$

Vemos así que la elección  $c(\varphi_2) = \varphi_2$  es razonable en el sentido de que lleva a una reparametrización con interpretación clara: los parámetros pasan de ser los extremos del intervalo a ser el origen y la longitud del mismo. Observemos que  $\varphi_1$  responde a las características de un parámetro de localización y  $\varphi_2$  a las de uno de escala. A partir de ahora vamos a denotar a los parámetros  $\varphi_1$  y  $\varphi_2$  por  $\mu$  y  $\sigma$  respectivamente.

Una vez conseguida la reparametrización ortogonal, podemos obtener la distribución a priori conjunta bajo el supuesto de independencia a priori entre los parámetros.

Observemos que el jacobiano de la transformación efectuada vale 1. Así, la función de densidad del modelo con los nuevos parámetros será

$$\begin{aligned}f(x|\mu, \sigma) &= \frac{1}{\sigma} \quad \text{si } \mu \leq x \leq \mu + \sigma, \quad \mu \in \mathbb{R}, \quad \sigma > 0 \\f(x|\mu, \sigma) &= \frac{1}{\sigma} I_{(0,1)}\left(\frac{x-\mu}{\sigma}\right) \\ \ell &= \log f = -\log \sigma\end{aligned}$$

dándose la ortogonalidad por ser  $\frac{\partial \ell}{\partial \mu} = 0$ .

- Para  $\sigma$  fijo, por ejemplo  $\sigma = 1$ , tenemos  $X \sim U(\mu, \mu+1)$ . Sabemos que en este caso, la distribución a priori es  $\pi(\mu) \propto 1$  y así  $\pi(\mu|\sigma) \propto c(\sigma)$ .

- Para  $\mu$  fijo, por ejemplo  $\mu = 0$ , tenemos  $X \sim U(0, \sigma)$ . Sabemos que en este caso, la distribución a priori es  $\pi(\sigma) \propto 1/\sigma$  y así  $\pi(\sigma|\mu) \propto k(\mu)/\sigma$ .

Por tanto, la distribución conjunta será  $\pi(\mu, \sigma) \propto \frac{1}{\sigma}$  y, dado que el jacobiano es 1, la distribución a priori para los parámetros originales es

$$\pi(a, b) \propto \frac{1}{b-a}$$

que coincide con la propuesta de Basulto, J. (1997).

**Observación:** Siendo la elección  $c(\varphi_2)$  una función estrictamente creciente o decreciente, la distribución a priori  $\pi(a, b)$  siempre resulta ser la misma, independientemente de la elección efectuada.

En efecto, si

$$\begin{cases} a = \varphi_1 \\ b = \varphi_1 + c(\varphi_2) \end{cases}$$

entonces tenemos

$$\begin{cases} \varphi_1 = a \\ \varphi_2 = c^{-1}(b - a) \end{cases}$$

y por tanto, la función de densidad del modelo viene dada por

$$f(x | \varphi_1, \varphi_2) = \frac{1}{c(\varphi_2)} \quad \varphi_1 \leq x \leq \varphi_1 + c(\varphi_2)$$

De esta forma,  $\pi(\varphi_1 | \varphi_2) \propto k(\varphi_2)$  y  $\pi(\varphi_2 | \varphi_1) \propto \frac{|c'(\varphi_2)|}{c(\varphi_2)} \tilde{k}(\varphi_2)$ , con lo que la

distribución conjunta resulta ser  $\pi(\varphi_1, \varphi_2) \propto \frac{|c'(\varphi_2)|}{c(\varphi_2)}$ .

Aplicando el correspondiente cambios de variable para calcular  $\pi(a, b)$  obtenemos

$$\pi(a, b) \propto \pi(\varphi_1, \varphi_2) |J| = \frac{|c'(\varphi_2)|}{c(\varphi_2)} \frac{1}{|c'(\varphi_2)|} = \frac{1}{c(\varphi_2)} = \frac{1}{b - a}$$



### 3.2.1 Propiedades frecuentistas de los intervalos bayesianos.

Al igual que en el caso uniparamétrico, vamos a ver a través de algunos ejemplos cómo la distribución a priori que proponemos conduce a intervalos bayesianos coincidentes numéricamente con los clásicos cuando existe un estadístico suficiente y completo de la misma dimensión que el parámetro; cuando no existe dicho estadístico, los intervalos bayesianos coinciden con los obtenidos a través del principio de condicionar a un estadístico auxiliar.

**Ejemplo 3.5: (Continuación)** Con respecto a la coincidencia entre intervalos bayesianos y clásicos, observemos que en este modelo los estadísticos máximo verosímiles  $\hat{\mu} = X_{(1)}$  y  $\hat{\sigma} = R = X_{(n)} - X_{(1)}$  son conjuntamente suficientes y completos. Vamos a comprobar directamente que los intervalos bayesianos de probabilidad  $\alpha$  para  $\mu$  y  $\sigma$  coinciden con los intervalos clásicos obtenidos a partir de las cantidades pivotaes  $U = \frac{\hat{\mu} - \mu}{\hat{\sigma}}$  y  $V = \frac{\hat{\sigma}}{\sigma}$ .

- Inferencia Bayesiana con la distribución  $\pi(\mu, \sigma) \propto \frac{1}{\sigma}$ .

La función de verosimilitud viene dada por

$$\ell(\mu, \sigma | \mathbf{x}) \propto \frac{1}{\sigma^n} \prod_{i=1}^n I_{(0,1)} \left( \frac{x_i - \mu}{\sigma} \right) = \frac{1}{\sigma^n} \prod_{i=1}^n h \left( \frac{x_i - \mu}{\sigma} \right)$$

Al tomar  $\pi(\mu, \sigma) \propto \frac{1}{\sigma}$ , la distribución a posteriori conjunta será:

$$\pi(\mu, \sigma | \mathbf{x}) \propto \frac{1}{\sigma^{n+1}} \prod_{i=1}^n h \left( \frac{x_i - \mu}{\sigma} \right)$$

o bien

$$\pi(\mu, \sigma | x) = \frac{\frac{1}{\sigma^{n+1}} \prod_{i=1}^n h\left(\frac{x_i - \mu}{\sigma}\right)}{\int_0^{+\infty} \frac{1}{\sigma^{n+1}} \int_{-\infty}^{+\infty} \prod_{i=1}^n h\left(\frac{x_i - \mu}{\sigma}\right) d\mu d\sigma}$$

Antes de continuar, es conveniente aclarar cuál es el rango en el que está definida esta densidad, rango que a su vez determinará los límites de la doble integral del denominador.

$$\begin{aligned} \prod I_{(0,1)}\left(\frac{x_i - \mu}{\sigma}\right) = 1 &\Leftrightarrow I_{(0,1)}\left(\frac{x_i - \mu}{\sigma}\right) = 1 \quad \forall i \Leftrightarrow 0 \leq \frac{x_i - \mu}{\sigma} \leq 1 \quad \forall i \Leftrightarrow \\ &\Leftrightarrow \begin{cases} 0 \leq \frac{x_{(1)} - \mu}{\sigma} \\ \frac{x_{(n)} - \mu}{\sigma} \leq 1 \end{cases} \Leftrightarrow \begin{cases} 0 \leq x_{(1)} - \mu \\ x_{(n)} - \mu \leq \sigma \end{cases} \Leftrightarrow \begin{cases} \mu \leq x_{(1)} \\ x_{(n)} - \mu \leq \sigma \end{cases} \end{aligned}$$

Esta región es la representada en la figura que sigue.

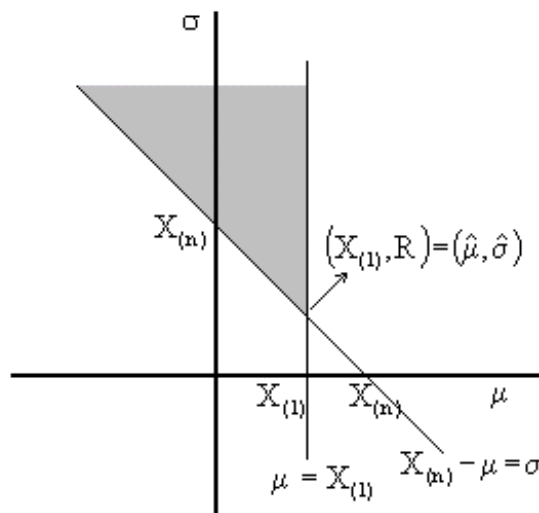


Figura 3.3: Rango de  $\pi(\mu, \sigma | x)$

Observemos en particular que  $\mu \in (-\infty, \hat{\mu})$  y que  $\sigma \in (\hat{\sigma}, +\infty)$ .

Antes de calcular las distribuciones a posteriori marginales, vamos a calcular la constante que hace que  $\pi(\mu, \sigma | \mathbf{x})$  sea función de densidad.

$$\begin{aligned} C &= \int_0^{+\infty} \frac{1}{\sigma^{n+1}} \int_{-\infty}^{+\infty} \prod_{i=1}^n h\left(\frac{x_i - \mu}{\sigma}\right) d\mu d\sigma = \int_{\mathbb{R}} \frac{1}{\sigma^{n+1}} \left( \int_{x_{(n)} - \sigma}^{x_{(1)}} d\mu \right) d\sigma = \\ &= \int_{\mathbb{R}} \frac{\sigma - R}{\sigma^{n+1}} d\sigma = \frac{1}{n(n-1)R^{n-1}}. \end{aligned}$$

1) Distribución marginal de  $\mu$ .

$$\pi(\mu | \mathbf{x}) = \int_0^{+\infty} \pi(\mu, \sigma | \mathbf{x}) d\sigma = \frac{1}{C} \int_{x_{(n)} - \mu}^{+\infty} \sigma^{-(n+1)} d\sigma = \frac{1}{C} \frac{(x_{(n)} - \mu)^{-n}}{n} = (n-1)R^{n-1} (x_{(n)} - \mu)^{-n}$$

donde el recorrido, como se señaló antes, es  $\mu \in (-\infty, \hat{\mu})$ .

La representación gráfica de esta función para  $n=10$ ,  $x_{(1)}=3$  y  $x_{(n)}=5$  es la siguiente:

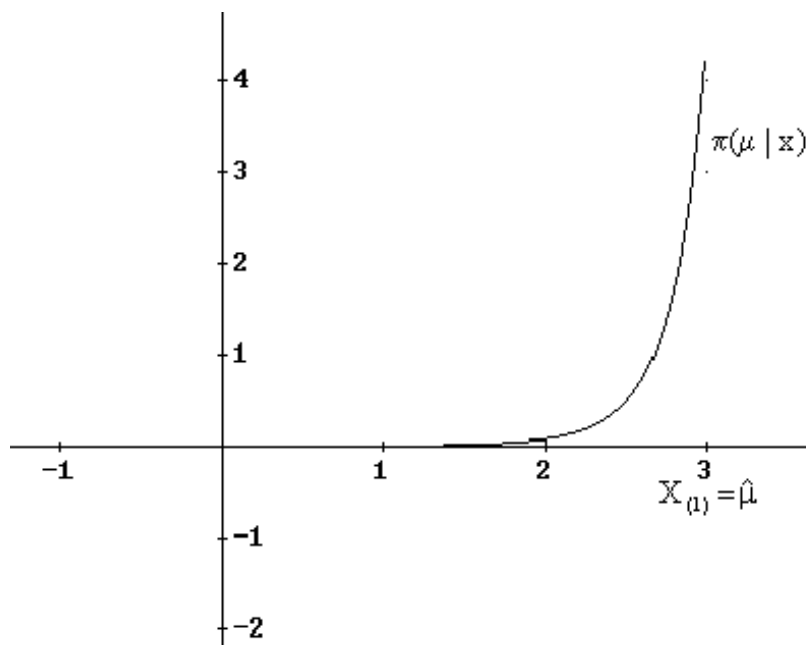
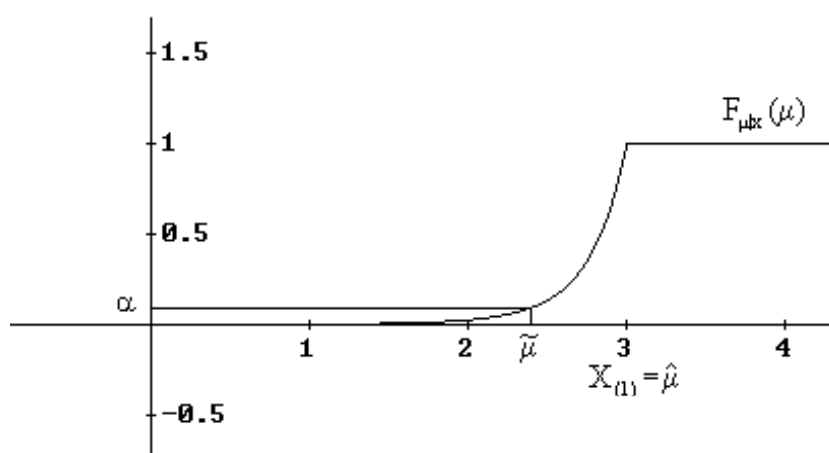


Figura 3.4: Función de densidad de  $\mu | \mathbf{x}$ ,  $n = 10$ ,  $x_{(1)} = 3$ ,  $x_{(n)} = 5$

En este caso podemos calcular explícitamente la función de distribución marginal a posteriori, que resulta ser

$$F_{\mu|x}(\mu) = \int_{-\infty}^{\mu} \pi(u|x) du = \left( \frac{R}{x_{(n)} - \mu} \right)^{n-1} \quad -\infty < \mu \leq x_{(1)} = \hat{\mu}.$$

Podemos obtener de esta forma una expresión explícita general para un intervalo de probabilidad  $1-\alpha$  de la forma  $(\tilde{\mu}, x_{(1)})$ .



**Figura 3.5:** Función de distribución de  $\mu | x$ ,  $n = 10$ ,  $x_{(1)} = 3$ ,  $x_{(n)} = 5$

$$1 - \alpha = P [\mu > \tilde{\mu} | x] = 1 - F_{\mu|x}(\tilde{\mu}) = 1 - \left( \frac{R}{x_{(n)} - \tilde{\mu}} \right)^{n-1} \Leftrightarrow (\dots) \Leftrightarrow \tilde{\mu} = x_{(n)} - R\alpha^{\frac{1}{n-1}}.$$

Por ejemplo, para  $n=10$ ,  $x_{(1)}=3$  y  $x_{(n)}=5$  ( $R=2$ ), y tomando  $\alpha=0.05$ , el intervalo de probabilidad 0.95 sería  $(2.2109, 3)$ . Para  $n=30$ , manteniendo los mismos valores anteriores, el intervalo resulta ser  $(2.7823, 3)$ .

## 2) Distribución marginal de $\sigma$ .

$$\pi(\sigma | \mathbf{x}) = \int_{-\infty}^{+\infty} \pi(\mu, \sigma | \mathbf{x}) d\mu = \frac{1}{C\sigma^{n+1}} \int_{x_{(n)}-\sigma}^{x_{(1)}} d\mu = \frac{(\sigma - R)n(n-1)R^{n-1}}{\sigma^{n+1}} \quad \sigma > R.$$

La representación gráfica de esta función para los datos anteriores es

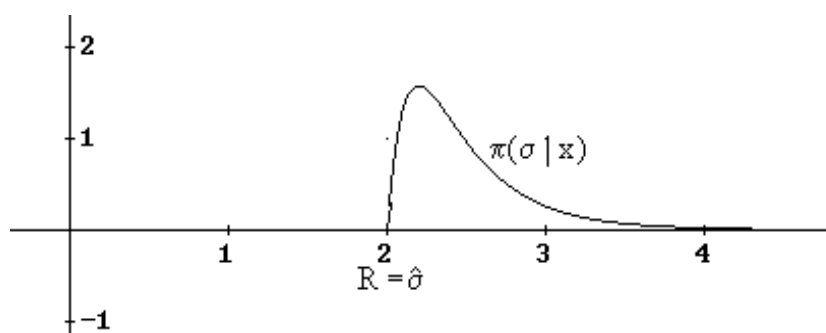


Figura 3.6: Función de densidad de  $\sigma | \mathbf{x}$ ,  $n = 10$ ,  $x_{(1)} = 3$ ,  $x_{(n)} = 5$

En este caso también puede obtenerse explícitamente la función de distribución. Concretamente, dicha función es

$$F_{\sigma|\mathbf{x}}(\sigma) = \int_R^\sigma \pi(s | \mathbf{x}) ds = 1 + R^{n-1} \sigma^{-n} (n(R - \sigma) - R) \quad \sigma > R.$$

Con esta función de distribución, no podemos obtener de forma explícita la expresión de un intervalo de probabilidad general de la forma  $(R, \tilde{\sigma})$ , aunque es fácil obtenerlo por métodos numéricos sin más que resolver la ecuación:

$$1 - \alpha = F_{\sigma|\mathbf{x}}(\tilde{\sigma}) = 1 + R^{n-1} \tilde{\sigma}^{-n} (n(R - \tilde{\sigma}) - R)$$

En el ejemplo anterior, es decir para  $n=10$ ,  $x_{(1)}=3$ ,  $x_{(n)}=5$  y  $\alpha=0.05$ , el intervalo que se obtiene es  $(2, 3.3012)$ ; si tomamos  $n=30$ , el intervalo sería  $(2, 2.3491)$ .

- Inferencia Clásica con las cantidades pivotaes  $U = \frac{\hat{\mu} - \mu}{\hat{\sigma}}$  y  $V = \frac{\hat{\sigma}}{\sigma}$ .

Dada una muestra aleatoria simple  $X_1, \dots, X_n$  de un modelo con función de densidad  $f_X(x)$  y función de distribución  $F_X(x)$ , es conocido que la distribución conjunta del estadístico  $(X_{(1)}, X_{(n)})$  viene dada por:

$$f_{X_{(1)}, X_{(n)}}(x, y) = n(n-1)(F_X(y) - F_X(x))^{n-2} f_X(x) f_X(y) \quad x \leq y$$

En nuestro modelo, esta densidad quedaría entonces

$$f_{X_{(1)}, X_{(n)}}(x, y) = n(n-1)(y-x)^{n-2} \left(\frac{1}{\sigma}\right)^n \quad \mu \leq x \leq y \leq \mu + \sigma$$

Consideremos ahora el cambio de variable  $T=X_{(1)}$ ,  $R = X_{(n)} - X_{(1)}$  de jacobiano unitario, obteniéndose que la función de densidad conjunta de  $(T, R)$  es

$$f_{T,R}(t, r) = n(n-1) \left(\frac{1}{\sigma}\right)^n r^{n-2} \quad \mu \leq t \leq t+r \leq \mu + \sigma$$

Por último, hacemos el cambio

$$\begin{cases} U = \frac{T-\mu}{R} & \left( = \frac{\hat{\mu}-\mu}{\hat{\sigma}} \right) \\ V = \frac{R}{\sigma} & \left( = \frac{\hat{\sigma}}{\sigma} \right) \end{cases}$$

cuyo jacobiano es  $\sigma^2 v$ , obteniéndose la siguiente función de densidad para  $(U, V)$ :

$$f_{(U,V)}(u, v) = n(n-1)v^{n-1} \quad \mu \leq \mu + \sigma uv \leq \mu + \sigma uv + \sigma v \leq \mu + \sigma$$


---

El soporte de esta función de densidad es la región que se representa en la figura que sigue.



Figura 3.7: Soporte de  $f_{u,v}(u,v)$

1) Distribución marginal de U.

$$f_U(u) = \int_0^{1/(1+u)} n(n-1)v^{n-1} dv = (n-1) \left( \frac{1}{1+u} \right)^n \quad u \geq 0$$

$$F_U(u) = \int_0^u (n-1)(1+t)^{-n} dt = 1 - \left( \frac{1}{1+u} \right)^{n-1} \quad u \geq 0$$

Observemos entonces que hay una clara relación entre la función de distribución de U y la función de distribución a posteriori para  $\mu$ , ya que, dado  $u_0 \geq 0$ , se verifica:

$$F_{\mu|x}(\hat{\mu} - \hat{\sigma}u_0) = F_{\mu|x}(x_{(1)} - Ru_0) = \left( \frac{R}{R + Ru_0} \right)^{n-1} = \left( \frac{1}{1+u_0} \right)^{n-1} = 1 - F_U(u_0).$$

2) Distribución marginal de V.

$$f_V(v) = \int_0^{(1-v)/v} n(n-1)v^{n-1} du = n(n-1) v^{n-2}(1-v) \quad 0 \leq v \leq 1$$

$$F_V(v) = \int_0^v n(n-1)t^{n-2}(1-t)dt = v^{n-1}(v-n(v-1)) \quad 0 \leq v \leq 1$$

Observemos que en este caso la relación que se produce entre las funciones de distribución de  $V$  y de  $\sigma$  es

$$F_{\sigma|x}(\hat{\sigma}/v_0) = F_{\sigma|x}(R/v_0) = 1 - v_0^{n-1}(v_0 - n(v_0 - 1)) = 1 - F_V(v_0).$$

- Coincidencia numérica de los intervalos obtenidos por ambos métodos.

La coincidencia de intervalos se basa en las relaciones encontradas para las funciones de distribución obtenidas por ambos métodos. Recordemos que dichas relaciones eran

$$F_{\mu|x}(\hat{\mu} - \hat{\sigma}u_0) = 1 - F_U(u_0).$$

$$F_{\sigma|x}(\hat{\sigma}/v_0) = 1 - F_V(v_0).$$

Sea ahora  $\mu_\alpha$  tal que  $(\mu_\alpha, \hat{\mu})$  es un intervalo bayesiano para  $\mu$  de probabilidad  $1-\alpha$ , es decir, que  $\mu_\alpha$  es el valor del intervalo  $(-\infty, \hat{\mu})$  que verifica  $1-\alpha = P[\mu \geq \mu_\alpha | x] = 1 - F_{\mu|x}(\mu_\alpha)$ . Ahora, consideramos el valor  $u_\alpha = (\hat{\mu} - \mu_\alpha)/\hat{\sigma}$ . Entonces, tenemos

$$\begin{aligned} 1-\alpha &= 1 - F_{\mu|x}(\mu_\alpha) = 1 - F_{\mu|x}(\hat{\mu} - \hat{\sigma}u_\alpha) = F_U(u_\alpha) = P[U \leq u_\alpha] = \\ &= P\left[\frac{\hat{\mu} - \mu}{\hat{\sigma}} \leq u_\alpha\right] = P[\mu \geq \hat{\mu} - \hat{\sigma}u_\alpha]. \end{aligned}$$



y de esta manera, al pivotar, el intervalo de confianza que se obtiene es  $(\hat{\mu} - \hat{\sigma}u_\alpha, \hat{\mu})$ , que coincide con el intervalo bayesiano ya que  $\hat{\mu} - \hat{\sigma}u_\alpha = \mu_\alpha$ .

Análogamente, sea  $\sigma_\alpha$  tal que  $(\hat{\sigma}, \sigma_\alpha)$  es un intervalo bayesiano para  $\sigma$  de probabilidad  $1-\alpha$ , es decir, que  $\sigma_\alpha$  es el valor del intervalo  $(\hat{\sigma}, +\infty)$  que verifica  $1-\alpha = P[\sigma \leq \sigma_\alpha | x] = F_{\sigma|x}(\sigma_\alpha)$ . Ahora, consideramos el valor  $v_\alpha = \hat{\sigma}/\sigma_\alpha$ . Entonces, tenemos

$$\begin{aligned} 1-\alpha &= F_{\sigma|x}(\sigma_\alpha) = F_{\sigma|x}(\hat{\sigma}/v_\alpha) = 1 - F_V(v_\alpha) = P[V \geq v_\alpha] = \\ &= P\left[\frac{\hat{\sigma}}{\sigma} \geq v_\alpha\right] = P\left[\sigma \leq \frac{\hat{\sigma}}{v_\alpha}\right] \end{aligned}$$

y de esta manera, al pivotar, el intervalo de confianza que se obtiene es  $(\hat{\sigma}, \hat{\sigma}/v_\alpha)$ , que coincide con el intervalo bayesiano ya que  $\hat{\sigma}/v_\alpha = \sigma_\alpha$ .

**Ejemplo 3.6:** Consideremos  $X$  siguiendo un modelo doble exponencial.

$$f(x | \mu, \sigma) = \frac{1}{2\sigma} e^{-\frac{|x-\mu|}{\sigma}} \quad ; x \in \mathbb{R}, \mu \in \mathbb{R}, \sigma > 0.$$

Como vemos, este modelo también es del tipo localización – escala. En este caso, no existe un estadístico suficiente de dimensión dos (la dimensión del espacio paramétrico) y existen estadísticos auxiliares. Los intervalos clásicos no condicionados no van a coincidir con los obtenidos a partir del principio de condicionar a los estadísticos auxiliares, y de nuevo vamos a ver cómo la utilización de la distribución a priori conjunta que se propone en el presente trabajo, conduce a intervalos bayesianos que coinciden numéricamente con los intervalos obtenidos condicionando a los estadísticos auxiliares. Este resultado no es propio del modelo doble exponencial, sino que es un resultado más general aplicable al caso de localización – escala y que puede verse, por ejemplo, en Welsh, A.H. (1996).

Observemos que el ejemplo anterior del modelo uniforme también es un modelo de localización – escala. En él, también existen estadísticos auxiliares, pero al ser el estadístico máximo – verosímil suficiente y completo, en virtud del teorema de Basu (ver, por ejemplo, Lehman, E.L. (1991)), dicho estadístico será independiente de cualquier estadístico auxiliar, con lo que la inferencia basada en el máximo – verosímil coincide con la efectuada al condicionar a los estadísticos auxiliares.

1) Obtención de la distribución a priori no informativa.

$$\begin{aligned} \ln f(x | \mu, \sigma) &= -\ln(2\sigma) - \frac{|x - \mu|}{\sigma} \\ \frac{\partial \ell}{\partial \mu} &= \frac{\text{signo}(x - \mu)}{\sigma} \quad ; \quad \frac{\partial \ell}{\partial \sigma} = -\frac{1}{\sigma} + \frac{|x - \mu|}{\sigma^2} \\ E \left[ \frac{\partial \ell}{\partial \mu} \frac{\partial \ell}{\partial \sigma} \right] &= \int_{-\infty}^{\mu} \left( -\frac{1}{\sigma} \right) \left( -\frac{1}{\sigma} - \frac{x - \mu}{\sigma^2} \right) \frac{1}{2\sigma} e^{\frac{x - \mu}{\sigma}} dx + \\ &+ \int_{\mu}^{+\infty} \frac{1}{\sigma} \left( -\frac{1}{\sigma} + \frac{x - \mu}{\sigma^2} \right) \frac{1}{2\sigma} e^{-\frac{x - \mu}{\sigma}} dx = 0 \end{aligned}$$

Así, como tenemos  $E \left[ \frac{\partial \ell}{\partial \mu} \frac{\partial \ell}{\partial \sigma} \right] = 0$ , podemos considerar que los parámetros son independientes a priori y obtener la distribución a posteriori conjunta como producto de las respectivas marginales.

En primer lugar, vamos a obtener la distribución a priori para  $\sigma$ . Dado un valor fijo de  $\mu$ , por simplicidad tomamos  $\mu=0$ , vamos a obtener directamente la distribución a priori para  $\sigma$ , aplicando el procedimiento descrito en el caso uniparamétrico, aunque para  $\mu$  fijo el modelo es regular y sabemos entonces que  $\pi(\sigma | \mu) \propto \sqrt{I(\sigma)} c(\mu)$ .

$$I(\sigma, \sigma + h) = -8 \log \int_{-\infty}^{+\infty} \frac{1}{2\sigma} e^{-\frac{|y|}{\sigma}} \frac{1}{2(\sigma + h)} e^{-\frac{|y|}{\sigma+h}} dy = -8 \log \left( \frac{2\sqrt{\sigma(\sigma+h)}}{h+2\sigma} \right)$$

$$\lim_{h \rightarrow 0} \frac{I(\sigma, \sigma + h)}{h} = 0 \quad ; \quad \lim_{h \rightarrow 0} \frac{I(\sigma, \sigma + h)}{h^2} = \frac{1}{\sigma^2}$$

Por tanto,  $I(\sigma, \sigma+h)$  es de orden 2 con respecto a  $h$  y así, la distribución a priori para  $\sigma$  es

$$\pi(\sigma | \mu) \propto \sqrt{\lim_{h \rightarrow 0} \frac{I(\sigma, \sigma + h)}{h^2}} = \frac{1}{\sigma} c(\mu)$$

Ahora, vamos a obtener la distribución a priori para  $\mu$ . Para ello, tomamos un valor fijo de  $\sigma$ , por simplicidad  $\sigma=1$ , obteniendo los resultados que siguen.

Sea  $h > 0$ .

$$\begin{aligned} A(\mu, \mu + h) &= \int_{-\infty}^{\mu} (1/2) e^{-\frac{x-\mu}{2}} e^{-\frac{x-\mu-h}{2}} dx + \int_{\mu}^{\mu+h} (1/2) e^{-\frac{x-\mu}{2}} e^{-\frac{x-\mu-h}{2}} dx + \\ &+ \int_{\mu+h}^{+\infty} (1/2) e^{-\frac{x-\mu}{2}} e^{-\frac{x-\mu-h}{2}} dx = \\ &= \frac{1}{2} e^{-\frac{h}{2}} + \frac{1}{2} h e^{-\frac{h}{2}} + \frac{1}{2} e^{-\frac{h}{2}} = \frac{1}{2} e^{-\frac{h}{2}} (h+2) \end{aligned}$$

$$I(\mu, \mu + h) = -8 \log \left( \frac{1}{2} e^{-\frac{h}{2}} (h+2) \right) = 4h - 8 \log \left( \frac{h+2}{2} \right)$$

$$\lim_{h \rightarrow 0^+} \frac{I(\mu, \mu + h)}{h} = 0 \quad ; \quad \lim_{h \rightarrow 0^+} \frac{I(\mu, \mu + h)}{h^2} = 1$$

El mismo resultado se obtiene operando análogamente si  $h < 0$ . Por tanto, la distribución a priori para  $\mu$  será

---

$$\pi(\mu | \sigma) \propto 1 \cdot k(\sigma)$$

**Observación:** Como hemos visto, para  $\sigma$  fijo,  $I(\mu, \mu + h)$  es de orden 2, lo que significa que el modelo también es regular respecto al parámetro  $\mu$ . De hecho, también puede verificarse sin dificultad que se tiene  $E\left[\frac{\partial \ell}{\partial \mu}\right] = 0$ , ya que

$$E\left[\frac{\partial \ell}{\partial \mu}\right] = \int_{-\infty}^{+\infty} \text{signo}(x - \mu) f(x; \mu) dx = -\frac{1}{2} \int_{-\infty}^{\mu} e^{x-\mu} dx + -\frac{1}{2} \int_{\mu}^{+\infty} e^{\mu-x} dx = -\frac{1}{2} + \frac{1}{2} = 0.$$

En conclusión, la distribución a priori conjunta para el modelo doble exponencial vendrá dada por

$$\pi(\mu, \sigma) \propto \frac{1}{\sigma}$$

## 2) Inferencia Bayesiana en el modelo doble exponencial.

$$\ell(\mu, \sigma | x) = \left(\frac{1}{2\sigma}\right)^n e^{-\frac{1}{\sigma} \sum |x_i - \mu|}$$

$$\pi(\mu, \sigma | x) \propto \left(\frac{1}{\sigma}\right)^{n+1} e^{-\frac{1}{\sigma} \sum |x_i - \mu|}$$

Así, llamando  $K = \int_{-\infty}^{+\infty} \int_0^{+\infty} \left(\frac{1}{\sigma}\right)^{n+1} e^{-\frac{1}{\sigma} \sum |x_i - \mu|} d\sigma d\mu$ , la distribución a posteriori conjunta será

$$\pi(\mu, \sigma | x) = \frac{\left(\frac{1}{\sigma}\right)^{n+1} e^{-\frac{1}{\sigma} \sum |x_i - \mu|}}{K}$$

Las funciones de distribución marginales vienen dadas entonces por

$$F_{\mu|x}(\mu) = \frac{\int_{-\infty}^{\mu} \left( \int_0^{+\infty} \left(\frac{1}{\sigma}\right)^{n+1} e^{-\frac{1}{\sigma} \sum |x_i - t|} d\sigma \right) dt}{K}$$

$$F_{\sigma|x}(\sigma) = \frac{\int_0^{\sigma} \left( \int_{-\infty}^{+\infty} \left(\frac{1}{s}\right)^{n+1} e^{-\frac{1}{s} \sum |x_i - \mu|} d\mu \right) ds}{K}$$

### 3) Inferencia condicionada en el modelo doble exponencial.

La función de densidad del modelo, como vimos antes, es

$$f(x | \mu, \sigma) = \frac{1}{\sigma} h\left(\frac{y - \mu}{\sigma}\right) \quad y \in \mathbb{R}, \quad \mu \in \mathbb{R}, \quad \sigma > 0, \quad \text{donde} \quad h(z) = \frac{1}{2} e^{-|z|}$$

En Kappenman, R.F. (1975), puede verse que los estadísticos máximo – verosímiles de este modelo vienen dados por:

$$\hat{\mu} = \begin{cases} x_{(m)} & \text{si } n \text{ es impar donde } m = (n+1)/2 \\ \frac{1}{2}x_{(m)} + \frac{1}{2}x_{(m+1)} & \text{si } n \text{ es par donde } m = n/2 \end{cases}$$

$$\hat{\sigma} = \frac{1}{n} \sum |x_i - \hat{\mu}|$$

Además, es muy fácil comprobar que estos estadísticos son equivariantes en el sentido de que  $\hat{\mu}(b+ax) = b+a\hat{\mu}(x)$  y  $\hat{\sigma}(b+ax) = |a| \hat{\sigma}(x)$ .

Ahora, definimos los estadísticos  $C_i = (X_{(i)} - \hat{\mu})/\hat{\sigma}$   $i=1, \dots, n$ , que constituyen una configuración de dimensión  $n-2$  pues se dan las relaciones

$$1. \sum |C_i| = 0$$

$$2. \begin{cases} C_m = 0 & \text{si } n \text{ es par} \\ C_m = -C_{m+1} & \text{si } n \text{ es impar} \end{cases}$$

En Welsh, A.H. (1996) puede verse que en esta situación general, si definimos las cantidades pivotaes  $U = \frac{\hat{\mu} - \mu}{\hat{\sigma}}$  y  $V = \frac{\hat{\sigma}}{\sigma}$ , la función de densidad de  $(U,V)$  condicionada a los estadísticos auxiliares  $\{C_i\}$  viene dada por

$$f_{(U,V)|C}(u, v) = \frac{v^{n-1} \prod h(vc_i + vu)}{\int_{-\infty}^{+\infty} \int_0^{+\infty} v^{n-1} \prod h(vc_i + vu) dv du}$$

La función de distribución conjunta, entonces, se obtiene como

$$F_{(U,V)|C}(u_0, v_0) = P [U \leq u_0, V \leq v_0 | C_i] = \frac{\int_{-\infty}^{u_0} \int_0^{v_0} v^{n-1} \prod h(vc_i + vu) dv du}{\int_{-\infty}^{+\infty} \int_0^{+\infty} v^{n-1} \prod h(vc_i + vu) dv du}$$

y haciendo en las integrales el cambio de variables  $U = (\hat{\mu} - T)/\hat{\sigma}$  y  $V = \hat{\sigma}/S$ , la expresión que obtenemos para la función de distribución es

$$F_{(U,V)C}(u_0, v_0) = \frac{\int_{\hat{\mu}-\hat{\sigma}u_0}^{+\infty} \int_{\hat{\sigma}/v_0}^{+\infty} \frac{1}{s^{n+1}} \prod h\left(\frac{x_i - t}{s}\right) ds dt}{\int_{-\infty}^{+\infty} \int_0^{+\infty} \frac{1}{s^{n+1}} \prod h\left(\frac{x_i - t}{s}\right) ds dt}$$

Aplicando esta expresión general al modelo doble exponencial, obtenemos la función de distribución

$$F_{(U,V)C}(u_0, v_0) = \frac{\int_{\hat{\mu}-\hat{\sigma}u_0}^{+\infty} \int_{\hat{\sigma}/v_0}^{+\infty} \frac{1}{s^{n+1}} e^{-\frac{1}{s}\sum|x_i-t|} ds dt}{\int_{-\infty}^{+\infty} \int_0^{+\infty} \frac{1}{s^{n+1}} e^{-\frac{1}{s}\sum|x_i-t|} ds dt} = \frac{\int_{\hat{\mu}-\hat{\sigma}u_0}^{+\infty} \int_{\hat{\sigma}/v_0}^{+\infty} \frac{1}{s^{n+1}} e^{-\frac{1}{s}\sum|x_i-t|} ds dt}{K}$$

A partir de esta expresión para la función de distribución conjunta, podemos obtener las funciones de distribución marginales, que serían

$$F_{U|C}(u_0) = \frac{\int_{\hat{\mu}-\hat{\sigma}u_0}^{+\infty} \left( \int_0^{+\infty} \frac{1}{s^{n+1}} e^{-\frac{1}{s}\sum|x_i-t|} ds \right) dt}{K}$$

$$F_{V|C}(v_0) = \frac{\int_{\hat{\sigma}/v_0}^{+\infty} \left( \int_{-\infty}^{+\infty} \frac{1}{s^{n+1}} e^{-\frac{1}{s}\sum|x_i-t|} dt \right) ds}{K}$$

#### 4) Coincidencia numérica de los intervalos obtenidos por ambos métodos.

Al igual que en el ejemplo de la distribución uniforme, se observan las siguientes relaciones entre las funciones de distribución condicionadas y las funciones de distribución a posteriori.

$$F_{\mu|x}(\hat{\mu} - \hat{\sigma}u_0) = 1 - F_{U|C}(u_0).$$

$$F_{\sigma|x}(\hat{\sigma}/v_0) = 1 - F_{V|C}(v_0).$$

A partir de estas igualdades, la coincidencia de los intervalos Bayesianos con los intervalos obtenidos aplicando el principio de condicionar se obtiene exactamente igual que se hizo en el ejemplo anterior. La única diferencia es que ahora sería quizás más lógico construir intervalos bilaterales centrados en los estadísticos máximo – verosímiles, pero la coincidencia se seguiría teniendo dada las relaciones entre las funciones de distribución.



## **CAPÍTULO 4: ANÁLISIS DE ALGUNOS MODELOS DE DURACIÓN.**

En este capítulo pretendemos ilustrar cómo resolver de forma práctica distintos modelos al usar la metodología descrita en los capítulos anteriores, cuáles son los problemas que surgen y cuáles son las herramientas disponibles para solventarlos, todo ello desde una perspectiva eminentemente práctica.

Vamos a centrarnos en los modelos de duración, de gran utilidad práctica en diversos campos de la ciencia, pero a la vez bastante complicados de estudiar debido a que la presencia de censura es bastante habitual.

El uso de métodos de simulación tipo Monte Carlo junto con el algoritmo de Gibbs (que permite simular muestras de distribuciones multidimensionales a partir de las densidades condicionadas unidimensionales) hace que dispongamos de procedimientos bastante mecánicos para resolver modelos de gran complejidad teórica.

Hemos analizado los modelos Exponencial, Exponencial con parámetro umbral y Weibull, ofreciendo en el apéndice B los listados de los programas utilizados para su resolución.

### **4.1 Algunos conceptos básicos de los modelos de duración.**

El análisis estadístico de los modelos de duración (también llamados modelos de tiempo de vida, modelos de supervivencia, etc.) ha sido aplicado en distintas especialidades de la ciencia. Las disciplinas en las que quizás han adquirido una

mayor importancia han sido la ingeniería (duración de componentes electrónicos) y la medicina (tiempo de vida de pacientes con ciertos tipos de enfermedades). No obstante, este tipo de modelos también han adquirido importancia en la economía, especialmente en estudios relativos al desempleo (tiempo que tarda una persona desempleada en encontrar trabajo). Otro campo en el que se está aplicando es el de la bibliometría, concretamente al estudiar la antigüedad de las citas (o referencias) de la literatura científica.

La búsqueda de modelos que se adapten a las situaciones anteriores conlleva una dificultad específica: han de reflejar que, generalmente, los datos son *censurados*. La censura es común en los modelos de duración, ya que en muchas ocasiones es necesario finalizar el experimento antes de que todos los elementos objeto de estudio hayan sido observados. En estos casos, disponemos de la información de que el tiempo de duración es superior al tiempo en que se ha detenido la observación, pero no conocemos el valor exacto de la duración del ítem. Así, aunque los modelos paramétricos que se suelen usar son bastante conocidos (Exponencial, Weibull, Gamma, Log-Normal, etc.), la presencia de censura hace que el análisis estadístico se complique en gran medida.

A veces, es adecuado introducir en los modelos un *parámetro umbral* o *de garantía* (el término en inglés es “*threshold parameter*”), es decir, un valor mínimo para el tiempo de duración de cualquier ítem o, en definitiva, una traslación del modelo original. Si adoptamos el enfoque Bayesiano para resolver un problema de este tipo y no tenemos ninguna creencia a priori especialmente relevante sobre los parámetros, tendremos que elegir una distribución a priori no informativa, dándose la circunstancia de que los modelos van a ser no regulares a causa de la presencia del parámetro umbral. En este capítulo, vamos a aplicar la propuesta de obtención de la distribución a priori no informativa a este tipo de situaciones y a partir de ella, veremos que los métodos de simulación permiten que el análisis bayesiano resulte relativamente simple.

Consideremos la variable aleatoria  $T$  que nos da el tiempo transcurrido hasta que ocurre un cierto suceso o tiempo de duración en una cierta población. Habitualmente, se supondrá que  $T$  es una variable absolutamente continua que toma valores en el intervalo  $[0, +\infty)$  (si el modelo tiene parámetro umbral, el recorrido de la variable será  $[\theta, +\infty)$ , donde  $\theta \geq 0$ ). Además de las funciones de distribución y densidad de la variable  $T$  (que representaremos por  $F(t)$  y  $f(t)$  respectivamente), en el marco de los modelos de duración adquieren especial importancia la función de supervivencia y la función de riesgo.

#### 4.1.1 La función de supervivencia y la función de riesgo.

La función de supervivencia se define como

$$S(t) = P[T > t]$$

que coincide con  $1-F(t)$ . Observemos que entonces  $S(t)$  es una función no creciente verificando  $S(0) = 1$  y  $\lim_{t \rightarrow +\infty} S(t) = 0$ .

La función de riesgo se define como

$$h(t) = \lim_{h \rightarrow 0} \frac{P[t \leq T < t+h \mid T \geq t]}{h}$$

y por tanto, siendo  $T$  absolutamente continua, obtenemos

$$h(t) = \frac{f(t)}{S(t)}$$

ya que

$$\lim_{h \rightarrow 0} \frac{P[t \leq T < t+h | T \geq t]}{h} = \frac{1}{P[T \geq t]} \lim_{h \rightarrow 0} \frac{P[t \leq T < t+h]}{h} = \frac{1}{S(t)} \lim_{h \rightarrow 0} \frac{F(t+h) - F(t)}{h}.$$

La función de riesgo tiene una interpretación clara, pues proporciona la tasa instantánea de "fallo" en cada instante  $t$  dado que el ítem ha durado hasta el instante  $t$ . Esta interpretación hace que en los datos de duración sea frecuente modelar en base a la función de riesgo en vez de buscar directamente una función de densidad que se ajuste a los datos. Así, según la situación, se busca un modelo con función de riesgo creciente, constante, decreciente, etc. más que una función de densidad con una forma determinada.

Es importante señalar que la especificación de las funciones de riesgo y de supervivencia determinan de forma única el modelo, al igual que lo hacen las funciones de densidad y de distribución. De hecho, el conocimiento de unas funciones a partir de otras es directo (salvo problemas de cálculo con las integrales que aparecen en las expresiones). Las siguientes relaciones (en el caso absolutamente continuo) son todas fáciles de verificar.

$$h(x) = -\frac{\partial \log(S(x))}{\partial x}$$
$$S(t) = \exp\left(-\int_0^t h(x) dx\right)$$
$$f(t) = h(t) \exp\left(-\int_0^t h(x) dx\right)$$

**Ejemplo:** Consideremos  $T \sim \text{Exp}(\lambda)$ . En este caso, la función de riesgo es constante, ya que, como es sabido,  $f(t) = \lambda e^{-\lambda t}$ ,  $F(t) = 1 - e^{-\lambda t}$ ,  $S(t) = e^{-\lambda t}$  y por tanto,  $h(t) = \lambda$ .

**Ejemplo:** Consideremos  $T \sim W(\lambda, \beta)$ , es decir,  $T$  sigue una distribución de Weibull con parámetros  $\lambda$  y  $\beta$  (donde ambos parámetros son positivos). Este modelo

---

es uno de los más utilizados en los estudios de duración. Uno de los motivos de ello, entre otros, es la flexibilidad de su función de riesgo dependiendo de los valores de los parámetros. En efecto, para este modelo se tiene:

$$f(t) = \lambda\beta(\lambda t)^{\beta-1} \exp\{-(\lambda t)^\beta\}, \quad t > 0$$
$$S(t) = \int_t^{+\infty} f(u)du = \exp\{-(\lambda t)^\beta\}, \quad t > 0$$

y por tanto, la función de riesgo vendrá dada por:

$$h(t) = \lambda\beta(\lambda t)^{\beta-1}$$

Ahora, teniendo en cuenta que  $h'(t) = \lambda\beta(\beta-1)(\lambda t)^{\beta-2}$  y que tanto  $\lambda$  como  $\beta$  como  $t$  son positivos, la función de riesgo es creciente para  $\beta > 1$ , es decreciente si  $\beta < 1$  y es constante si  $\beta = 1$ .

#### **4.1.2 Principales tipos de censura.**

Ya se comentó anteriormente, que en los experimentos concernientes a la observación de tiempos de vida o duración es frecuente la aparición de censura, en el sentido de que no se observan todos los tiempos de vida, sino que para algunos ítems sólo se conoce que su tiempo de duración es superior a un cierto valor, circunstancia que ocurre por la necesidad de detener el experimento antes de observar todos los tiempos de fallo. En este epígrafe vamos a definir los dos tipos de censura más importantes, conocidos como censura tipo I y tipo II.

- **Censura tipo I.**

La censura tipo I aparece cuando el período de observación está prefijado de antemano, de forma que el tiempo de vida de un individuo se conocerá si es inferior al tiempo de duración del experimento. En otro caso, sólo sabremos que el tiempo de vida es superior al valor prefijado que denotaremos por  $L$ . De forma un poco más general, podemos suponer que cada individuo tiene un período prefijado de observación distinto  $L_i$ , situación que puede darse si no todos los individuos han comenzado a ser observados en el mismo instante y se ha prefijado el momento en el que finalizará el experimento.

En realidad, podemos considerar una situación más general, en la que los tiempos de censura sean una muestra aleatoria simple de una variable aleatoria  $L$  con función de densidad  $f_L(l)$  y de supervivencia  $S_L(l)$ . En este caso, bajo los supuestos de que  $L$  y  $T$  son variables independientes y de que la distribución de  $L$  no depende del parámetro de interés  $\theta$  del que dependa la distribución de  $T$ , veremos que la verosimilitud que se obtiene es equivalente a la obtenida suponiendo que los tiempos de censura están prefijados y por tanto podemos hacer las inferencias condicionando a los valores observados de censura.

- **Censura tipo II.**

La censura tipo II se da cuando se prefija que sólo se observarán los  $r$  primeros tiempos de fallo ( $1 \leq r \leq n$ ). Así, si los tiempos de vida de los  $n$  individuos son  $T_1, \dots, T_n$ , nuestra observación se restringirá a  $T_{(1)} \leq T_{(2)} \leq \dots \leq T_{(r)}$ . Observemos que en este caso el número de tiempos de vida observados está prefijado, mientras que en la censura tipo I dicho número es una variable aleatoria.

### 4.1.3 Verosimilitudes asociadas a datos con censura.

En este apartado, vamos a obtener las funciones de verosimilitud asociadas a los dos tipos de censura mencionados anteriormente. Para ello, consideraremos las variables  $T_i$ ,  $i=1\dots n$  que nos proporcionan los tiempos de vida de los  $n$  individuos, que serán independientes e idénticamente distribuidas con función de densidad  $f(t, \theta)$  donde  $\theta \in \Theta$ . Comenzaremos con el caso de la censura tipo II, que es el que resulta más sencillo.

- **Censura tipo II.**

En este caso, la verosimilitud no es más que la función de densidad del vector  $T_{(1)}, T_{(2)}, \dots, T_{(r)}$ , y por tanto,

$$\begin{aligned} L(\theta | t_{(1)} \dots t_{(r)}) &= \frac{n!}{(n-r)!} f(t_{(1)}, \theta) \dots f(t_{(r)}, \theta) [1 - F(t_{(r)}, \theta)]^{n-r} = \\ &= \frac{n!}{(n-r)!} f(t_{(1)}, \theta) \dots f(t_{(r)}, \theta) [S(t_{(r)}, \theta)]^{n-r} \end{aligned}$$

- **Censura tipo I.**

En este caso, además de la variables  $T_i$ ,  $i=1\dots n$ , tenemos los tiempos prefijados de censura  $L_i$ ,  $i=1\dots n$ , de forma que el tiempo de vida del individuo  $i$ -ésimo será observado si  $T_i \leq L_i$ . Para obtener la verosimilitud, siguiendo a Lawless, J.F. (1982), vamos a definir las variables

$$U_i = \text{Min}(T_i, L_i) \quad \text{y} \quad \delta_i = \begin{cases} 1 & \text{si } T_i \leq L_i \\ 0 & \text{si } T_i > L_i \end{cases}$$

verificándose entonces que  $U_i = T_i$  si  $\delta_i = 1$  y  $U_i = L_i$  si  $\delta_i = 0$ .

Como veremos, la distribución de probabilidad de  $(U_i, \delta_i)$  viene dada por la expresión  $f(t_i, \theta)^{\delta_i} S(L_i, \theta)^{1-\delta_i}$  y, puesto que los pares se suponen independientes entre sí, obtenemos

$$L(\theta | u_i, \delta_i) = L(\theta | t_i, L_i, \delta_i) = \prod_{i=1}^n f(t_i, \theta)^{\delta_i} S(L_i, \theta)^{1-\delta_i}$$

Veamos a continuación cómo se obtiene esta verosimilitud (por comodidad, suprimiremos en la notación la dependencia del parámetro  $\theta$ ).

Observemos que  $U_i$  es una variable mixta, con una parte continua y otra discreta y que  $\delta_i$  es discreta (lo que crea dificultades de notación al tener que representar las probabilidades y las densidades conjuntamente).

A) Sea  $u < L$ :

1.  $P[U \leq u, \delta = 1] = P[T \leq u, T \leq L] = P[T \leq u] = F(u)$
2.  $P[U \leq u, \delta = 0] = P[L \leq u, T > L] = 0$

B) Sea  $u = L$ :

1.  $P[U = L, \delta = 1] = P[T > L, T \leq L] = 0$
2.  $P[U = L, \delta = 0] = P[T > L, T > L] = P[T > L] = S(L)$

Así, la distribución conjunta de  $(U_i, \delta_i)$  vendrá dada por:



$$\begin{aligned}
 h(L_i, 0) &= P[U_i = L_i, \delta_i = 0] = S(L_i) \\
 h(u_i, 1) &= \frac{\partial P[U \leq u_i, \delta = 1]}{\partial u_i} = \frac{\partial F(u_i)}{\partial u_i} = f(u_i), \quad u_i \leq L_i \\
 h(L_i, 1) &= 0 \\
 h(u_i, 0) &= \frac{\partial P[U \leq u_i, \delta = 0]}{\partial u_i} = 0, \quad u_i \leq L_i
 \end{aligned}$$

y estas cuatro expresiones pueden combinarse en

$$h(u_i, \delta_i) = f(u_i)^{\delta_i} S(u_i)^{1-\delta_i} = f(t_i)^{\delta_i} S(L_i)^{1-\delta_i}$$

como pretendíamos demostrar.

Como se dijo anteriormente, puede considerarse la situación de que los tiempos de censura sean aleatorios. Veamos que bajo las hipótesis de que la distribución de los tiempos de censura es independiente de la de los tiempos de vida y no depende de los parámetros de interés objeto de estudio, se obtiene una verosimilitud equivalente a la obtenida con anterioridad. En efecto, en esta situación tendríamos:

$$\begin{aligned}
 P[U \leq u, \delta = 1] &= P[T \leq u, T \leq L] = \int_0^u \int_t^{+\infty} f_T(t) f_L(l) dt dl = \\
 &= \int_0^u f_T(t) \left( \int_t^{+\infty} f_L(l) dl \right) dt = \int_0^u f_T(t) S_L(t) dt
 \end{aligned}$$

y así  $h(u, 1) = \frac{\partial P[U \leq u, \delta = 1]}{\partial u} = f_T(u) S_L(u)$ .

Análogamente, se obtiene  $P[U \leq u, \delta = 0] = \int_0^u f_L(t)S_T(t)dt$  y como consecuencia,  $h(u, 0) = \frac{\partial P[U \leq u, \delta = 0]}{\partial u} = f_L(u)S_T(u)$ .

Uniendo estas dos expresiones obtenemos

$$h(u_i, \delta_i) = (f_T(u_i)S_L(u_i))^{\delta_i} (f_L(u_i)S_T(u_i))^{1-\delta_i}$$

Por tanto, la verosimilitud para una muestra de tamaño  $n$  será

$$L = \prod_{i=1}^n (f_T(u_i)S_L(u_i))^{\delta_i} (f_L(u_i)S_T(u_i))^{1-\delta_i} = \left( \prod_{i=1}^n f_T(u_i)^{\delta_i} S_T(u_i)^{1-\delta_i} \right) \left( \prod_{i=1}^n S_L(u_i)^{\delta_i} f_L(u_i)^{1-\delta_i} \right)$$

y, puesto que la distribución de  $L$  no depende de los parámetros de interés,

$$L \propto \left( \prod_{i=1}^n f_T(u_i)^{\delta_i} S_T(u_i)^{1-\delta_i} \right) = \left( \prod_{i=1}^n f_T(t_i)^{\delta_i} S_T(L_i)^{1-\delta_i} \right)$$

que, como puede apreciarse, coincide con la verosimilitud obtenida considerando que los valores  $L_i$  están prefijados.

Para finalizar este epígrafe, señalemos que, en general, los procesos de inferencia son mucho más directos en el caso de censura tipo II, basándonos en las propiedades de los estadísticos de orden. Sin embargo, en presencia de censura tipo I los problemas son más complicados y, habitualmente, es imposible trabajar con procedimientos exactos.

A continuación vamos a analizar la resolución de dos modelos importantes para datos de supervivencia, como son el Exponencial y el Weibull. Podemos destacar

que el primero es un caso particular del segundo. No obstante, dada su mayor simplicidad, el modelo exponencial ha sido muy utilizado para describir datos de supervivencia. En ambos casos analizaremos los casos de datos sin censura y datos con censura tipo I, teniendo además en cuenta para el modelo exponencial la posibilidad de introducir un parámetro umbral.

## 4.2 El modelo exponencial.

### 4.2.1 El modelo exponencial sin censura.

Supongamos que los tiempos de vida  $T_i$ ,  $i=1..n$  son independientes e idénticamente distribuidos según un modelo exponencial de parámetro  $\lambda > 0$ . En este caso, la obtención de estimaciones e intervalos para  $\lambda$  sería inmediata, ya que la verosimilitud viene dada por

$$L(\lambda | t_1, \dots, t_n) \propto \lambda^n e^{-\lambda \sum t_i}$$

la distribución a priori sería

$$\pi(\lambda) \propto 1/\lambda$$

y por tanto, la distribución a posteriori para el parámetro es

$$\pi(\lambda | t_1, \dots, t_n) \propto \lambda^{n-1} e^{-\lambda \sum t_i}$$

es decir, obtenemos  $\lambda | t_1, \dots, t_n \sim \Gamma(n, \sum t_i)$  y a partir de esta distribución conocida podemos obtener las estimaciones y los intervalos de probabilidad correspondientes.

#### 4.2.2 El modelo exponencial con censura tipo I.

Supongamos que los tiempos de vida  $T_i$ ,  $i=1\dots n$  son independientes e idénticamente distribuidos según un modelo exponencial de parámetro  $\lambda > 0$  y que los datos son observados si  $T_i \leq L_i$ . La verosimilitud asociada a una muestra de tamaño  $n$ , como hemos descrito en el epígrafe 4.1.3 vendrá por

$$L(\lambda | t_i, L_i, \delta_i) = \prod_{i=1}^n f(t_i, \lambda)^{\delta_i} S(L_i, \lambda)^{1-\delta_i}$$

Al aplicar esta expresión a nuestro modelo, obtenemos:

$$L(\lambda | t_i, L_i, \delta_i) \propto \prod_{i=1}^n (\lambda e^{-\lambda t_i \delta_i} e^{-\lambda L_i (1-\delta_i)}) \propto \lambda^r e^{-\lambda T}$$

donde  $T = \sum (\delta_i t_i + (1-\delta_i) L_i)$  representa la suma de todos los tiempos de vida y todos los tiempos censurados.

Puesto que el modelo es regular, para la obtención de la distribución a priori aplicamos la regla de Jeffreys. Siendo  $\ell(\lambda | t_i, L_i, \delta_i) = \text{Log } L(\lambda | t_i, L_i, \delta_i)$ , obtenemos que

$$\frac{\partial^2 \ell}{\partial \lambda^2} = -\frac{r}{\lambda^2}$$

Para una muestra de tamaño 1 tendríamos

$$E[r] = P[r=1] = P[T \leq L] = 1 - e^{-\lambda L}$$

y así

$$E\left[-\frac{\partial^2 \ell}{\partial \lambda^2}\right] = \frac{1 - e^{-\lambda L}}{\lambda^2}$$

Al estar considerando valores de censura distintos entre sí, utilizamos el logaritmo de la verosimilitud para la muestra completa de tamaño  $n$  y así en este caso obtenemos

$$E[r] = E\left[\sum_i \delta_i\right] = \sum_i P[T_i \leq L_i] = \sum_i (1 - e^{-\lambda L_i})$$

y por tanto la distribución a priori para  $\lambda$  sería

$$\pi(\lambda) \propto \frac{\sqrt{\frac{1}{n} \sum_i (1 - e^{-\lambda L_i})}}{\lambda} \propto \frac{\sqrt{\sum_i (1 - e^{-\lambda L_i})}}{\lambda}$$

De esta forma, la distribución a posteriori viene dada por

$$\pi(\lambda | t_i, L_i, \delta_i) \propto \lambda^{r-1} e^{-\lambda T} \sqrt{\sum (1 - e^{-\lambda L_i})}$$

que en el caso relativamente habitual de que se tenga  $L_i = L \quad \forall i$ , quedaría reducida a

$$\pi(\lambda | t_i, L_i, \delta_i) \propto \lambda^{r-1} e^{-\lambda T} \sqrt{1 - e^{-\lambda L}}$$

Esta distribución no se ajusta a ningún modelo conocido y además la integración exacta del núcleo de la densidad no es posible, por lo que la resolución del modelo habría de llevarse a cabo a través de simulación. En este caso, podemos aplicar el método de aceptación-rechazo (ver, por ejemplo, Smith A.F. and Gelfand,

A.E., (1992)). Para ello, vamos a considerar la densidad de una variable Gamma de parámetros  $r$  y  $T$ , es decir,

$$g(\lambda) = \frac{T^r}{(r-1)!} \lambda^{r-1} e^{-\lambda T}$$

Se verifica que el cociente  $\pi(\lambda | t_i, L_i, \delta_i) / g(\lambda)$  está acotado  $\forall \lambda > 0$ , ya que

$$\frac{\pi(\lambda | t_i, L_i, \delta_i)}{g(\lambda)} = \frac{\lambda^{r-1} e^{-\lambda T} \sqrt{\sum (1 - e^{-\lambda L_i})}}{\frac{T^r}{(r-1)!} \lambda^{r-1} e^{-\lambda T}} \leq \frac{\sqrt{n}}{T^r} (r-1)! = M$$

Así, generamos dos valores  $\lambda_0 \sim \Gamma(r, T)$  y  $u_0 \sim U(0,1)$  y aceptamos el valor de  $\lambda_0$  si se verifica

$$u_0 \leq \frac{\pi(\lambda_0 | t_i, L_i, \delta_i)}{Mg(\lambda_0)} = \frac{\lambda_0^{r-1} e^{-\lambda_0 T} \sqrt{\sum (1 - e^{-\lambda_0 L_i})}}{\frac{\sqrt{n}}{T^r} (r-1)! \frac{T^r}{(r-1)!} \lambda_0^{r-1} e^{-\lambda_0 T}} = \sqrt{\frac{1}{n} \sum (1 - e^{-\lambda_0 L_i})}$$

Es importante hacer notar que la probabilidad de aceptar el valor  $\lambda_0$  generado en general va a ser bastante alta, pues los términos  $1 - e^{-\lambda_0 L_i}$  con frecuencia serán cercanos a la unidad. No obstante, la mejor cota que se puede conseguir para la probabilidad de aceptar es  $\sqrt{1/n}$  (que, por ejemplo, para  $n=30$  vale aproximadamente 0,1825).

En el epígrafe siguiente trataremos un modelo más general que éste (el modelo exponencial con parámetro umbral o modelo doble exponencial) y en él veremos con más detalle los procedimientos de simulación así como varios ejemplos que ilustran tanto el proceso como el tipo de resultado que podemos obtener.

### 4.2.3 El modelo exponencial con parámetro umbral y censura tipo I.

Como se indicó con anterioridad, el tratamiento de modelos con censura tipo I es complicado. Vamos a tratar en este epígrafe un ejemplo de modelo con este tipo de censura que ha generado mucha literatura debido precisamente a la complejidad que presenta, como es el modelo exponencial.

En efecto, este modelo es complicado de tratar desde el punto de vista clásico, aún sin la presencia de un parámetro umbral. Veremos que el tratamiento bayesiano usando como distribución a priori la propuesta en el presente trabajo es relativamente directo. A través de simulación, comprobaremos también que los intervalos bayesianos de probabilidad  $1-\alpha$  tienen nivel de confianza muy cercano a  $1-\alpha$ .

En este modelo, tendremos que los tiempos de vida  $T_i$ ,  $i=1\dots n$  son independientes e idénticamente distribuidos, pero además, introducimos un parámetro umbral o parámetro de garantía  $\theta > 0$ , de forma que  $T_i \geq \theta \quad \forall i$ , y entonces la distribución de los tiempos de vida es tal que  $T_i - \theta$  sigue un modelo exponencial de parámetro  $\lambda > 0$ , o dicho de otra forma, las variables  $T_i$  siguen una distribución exponencial trasladada. De esta forma, tendremos que la función de densidad que modela el tiempo de vida es

$$f(t, \lambda, \theta) = \lambda e^{-\lambda(t-\theta)} \quad , \quad t \geq \theta, \quad \lambda, \theta > 0$$

La verosimilitud asociada a una muestra de tamaño  $n$ , como hemos descrito en el epígrafe 4.1.3 viene dada por

$$L(\lambda, \theta | t_i, L_i, \delta_i) = \prod_{i=1}^n f(t_i, \lambda, \theta)^{\delta_i} S(L_i, \lambda, \theta)^{1-\delta_i}$$

Al aplicar esta expresión a nuestro modelo, y teniendo en cuenta que si  $L_i < \theta$  la información  $T_i \geq L_i$  está implícita en el hecho de que  $T_i > \theta$ , obtenemos:

$$L(\lambda, \theta | t_i, L_i, \delta_i) = \prod_{i=1}^n \lambda e^{-\lambda \delta_i (t_i - \theta)} e^{-\lambda (1 - \delta_i) L_i} = \lambda^r e^{-\lambda [\sum \delta_i (t_i - \theta) + \sum (1 - \delta_i) \max\{L_i - \theta, 0\}]} , \quad t_{(1)} \geq \theta$$

donde hemos denotado  $r = \sum_i \delta_i$ , es decir,  $r$  representa el número de tiempos de vida observados (no censurados).

El máximo que aparece en el exponente complica el tratamiento de este modelo. Para evitar este inconveniente, trabajaremos bajo la hipótesis de que en nuestra muestra se verifica  $L_i \geq t_{(1)} \quad \forall i$ . Si en la muestra se tuviese  $L_{i_0} < t_{(1)}$ , eliminaríamos esta observación.

La idea que subyace en esta hipótesis es que  $L_{i_0} < t_{(1)}$  lo único que significa es que hemos detenido la observación del individuo  $i_0$ -ésimo “antes de tiempo” y el valor censurado obtenido no aporta información en este modelo. Hemos de hacer notar que los estadísticos máximo-verosímiles de los parámetros son

$$\hat{\theta}_{MV} = t_{(1)} ; \quad \hat{\lambda}_{MV} = \frac{r}{\sum \delta_i (t_i - t_{(1)}) + \sum (1 - \delta_i) \max(L_i - t_{(1)}, 0)}$$

de forma que si  $L_{i_0} < t_{(1)}$ , entonces el dato  $L_{i_0}$  en realidad no es tenido en cuenta, pues los dos sumandos correspondientes al término  $i_0$  son nulos. Así, en todas las inferencias basadas en los estadísticos máximo-verosímiles podemos eliminar los tiempos de censura menores que el mínimo de los tiempos observados. Bajo este supuesto de que  $L_i \geq t_{(1)} \quad \forall i$ , el estimador máximo verosímil para  $\lambda$  es  $\hat{\lambda}_{MV} = r / (T - nt_{(1)})$ .



Bajo este supuesto, la verosimilitud será entonces

$$L(\lambda, \theta | t_i, L_i, \delta_i) = \prod_{i=1}^n \lambda e^{-\lambda \delta_i (t_i - \theta)} e^{-\lambda (1 - \delta_i) L_i} = \lambda^n e^{-\lambda [\sum \delta_i (t_i - \theta) + \sum (1 - \delta_i) L_i]}, \quad t_{(1)} \geq \theta$$

A partir de la expresión de la función de verosimilitud, puede verse como el estadístico mínimo suficiente es de dimensión tres (mientras que el espacio paramétrico es de dimensión dos); concretamente, el estadístico mínimo suficiente viene dado por

$$\left( r, \sum \delta_i t_i + \sum (1 - \delta_i) L_i, t_{(1)} \right)$$

Antes de pasar al análisis bayesiano, resaltemos que estamos ante un modelo complicado de tratar. De hecho, todas las soluciones propuestas hasta ahora son siempre aproximadas (una exposición de ellas puede verse en Lawless, J. F. (1982), págs. 108 y 132); el propio Lawless, F.J. dice: “Se han publicado muchos trabajos sobre la distribución exponencial con dos parámetros; (...). Sin embargo, muy pocos trabajos se han ocupado de datos con censura tipo I, por lo que permanecen abiertas cuestiones concernientes a la adecuación de los métodos aproximados en pequeñas muestras”.<sup>7</sup>

- **Inferencia bayesiana.**

Como hemos visto, la verosimilitud asociada a una muestra de tamaño  $n$  de nuestro modelo es

---

<sup>7</sup> Lawless, J.F. (1982), p.133. El texto original es: “*Much work has been published on the two-parameter exponential distribution; (...). Little work has been directed at Type I censored data, however, so that questions remain concerning the adequacy of approximate methods in small samples.*”

$$L(\lambda, \theta | t_i, L_i, \delta_i) = \prod_{i=1}^n \lambda e^{-\lambda \delta_i (t_i - \theta)} e^{-\lambda (1 - \delta_i) L_i} = \lambda^n e^{-\lambda [\sum \delta_i (t_i - \theta) + \sum (1 - \delta_i) (L_i - \theta)]}, \quad t_{(1)} \geq \theta$$

cuyo logaritmo (escrito para una muestra de tamaño 1 y omitiendo por tanto los subíndices) será

$$\ell(\lambda, \theta | t, L, \delta) = r \ln \lambda - \lambda [\delta(t - \theta) - (1 - \delta)(L - \theta)], \quad t \geq \theta$$

• **Obtención de la distribución a priori:**

En este modelo, el parámetro  $\lambda$  tiene un comportamiento regular pero el parámetro  $\theta$  es no regular (en el sentido en que se definió en el capítulo 3). Ambos parámetros pueden considerarse independientes a priori, por lo que aplicaremos la regla de obtención de las distribuciones unidimensionales propuesta en el presente trabajo obteniéndose la conjunta como el producto de las marginales.

En efecto, a la vista del logaritmo de la verosimilitud, tenemos:

$$\frac{\partial \ell}{\partial \theta} = \lambda \delta + \lambda (1 - \delta) = \lambda$$

Observamos así que  $E[\partial \ell / \partial \theta] \neq 0$ , lo que muestra el comportamiento no regular, pero además,  $\partial \ell / \partial \theta$  no depende de los datos, lo que unido al hecho de que  $E[\partial \ell / \partial \lambda] = 0$  por ser el modelo regular respecto al parámetro  $\lambda$ , lleva a que

$$E \left[ \frac{\partial \ell}{\partial \lambda} \frac{\partial \ell}{\partial \theta} \right] = \frac{\partial \ell}{\partial \theta} E \left[ \frac{\partial \ell}{\partial \lambda} \right] = 0$$

por lo que los parámetros pueden considerarse independientes a priori.

Ahora, calculamos las distribuciones condicionadas (que coinciden con las marginales dado el supuesto de independencia), aplicando la propuesta para el caso unidimensional.

El parámetro  $\theta$  es de localización y así, como vimos en el capítulo 3, obtendremos una distribución a priori constante, es decir,

$$\pi(\theta | \lambda) \propto 1$$

El parámetro  $\lambda$  tiene un comportamiento regular y por tanto utilizaremos la regla de Jeffreys para obtener la distribución a priori. Así, puesto que

$$\frac{\partial \ell}{\partial \lambda} = \frac{r}{\lambda} - [\delta(t - \theta) - (1 - \delta)(L - \theta)]$$

$$\frac{\partial^2 \ell}{\partial \lambda^2} = -\frac{r}{\lambda^2}$$

$$E[r] = P[r = 1] = P[T \leq L] = 1 - e^{-\lambda(L - \theta)}$$

obtenemos, para una muestra de tamaño 1, que

$$E\left[\frac{\partial^2 \ell}{\partial \lambda^2}\right] = \frac{1 - e^{-\lambda(L - \theta)}}{\lambda^2}$$

Al estar considerando valores de censura distintos entre sí, utilizamos el logaritmo de la verosimilitud para la muestra completa de tamaño  $n$  y así en este caso obtenemos

$$E[r] = E\left[\sum_i \delta_i\right] = \sum_i P[T_i \leq L_i] = \sum_i \left(1 - e^{-\lambda(L_i - \theta)}\right)$$

y por tanto la distribución a priori para  $\lambda$  sería

---

$$\pi(\lambda | \theta) \propto \frac{\sqrt{\frac{1}{n} \sum_i (1 - e^{-\lambda(L_i - \theta)})}}{\lambda} \propto \frac{\sqrt{\sum_i (1 - e^{-\lambda(L_i - \theta)})}}{\lambda}$$

La distribución a priori conjunta para este modelo será pues

$$\pi(\lambda, \theta) \propto \frac{\sqrt{\sum_i (1 - e^{-\lambda(L_i - \theta)})}}{\lambda}$$

• **Obtención de la distribución a posteriori:**

La distribución a posteriori de  $(\lambda, \theta)$  vendrá dada por

$$\begin{aligned} \pi(\lambda, \theta | t_i, L_i, \delta_i) &\propto L(\lambda, \theta | t_i, L_i, \delta_i) \pi(\lambda, \theta) \propto \\ &\propto \lambda^{r-1} \sqrt{\sum_i (1 - e^{-\lambda(L_i - \theta)})} e^{-\lambda[\sum_i \delta_i (t_i - \theta) + \sum (1 - \delta_i)(L_i - \theta)]}, \quad t_{(1)} \geq \theta \geq 0, \quad \lambda > 0 \end{aligned}$$

Por simplicidad de notación en los cálculos posteriores, consideremos

$$D(\theta) = \sum \delta_i (t_i - \theta) + \sum (1 - \delta_i)(L_i - \theta)$$

Observemos que

$$\begin{aligned} D(\theta) &= \sum (\delta_i t_i + (1 - \delta_i)L_i) - \theta \sum (\delta_i + (1 - \delta_i)) = T - n\theta \\ D'(\theta) &= -n \end{aligned}$$

donde  $T = \sum (\delta_i t_i + (1 - \delta_i) L_i)$  representa la suma de todos los tiempos de vida y todos los tiempos censurados.

Con estas notaciones, podemos escribir la distribución a posteriori como

$$\begin{aligned} \pi(\lambda, \theta | t_i, L_i, \delta_i) &\propto \lambda^{r-1} \sqrt{\sum_i (1 - e^{-\lambda(L_i - \theta)})} e^{-\lambda D(\theta)} = \\ &= \lambda^{r-1} \sqrt{\sum_i (1 - e^{-\lambda(L_i - \theta)})} e^{-\lambda(T - n\theta)}, \quad t_{(1)} \geq \theta, \quad \lambda > 0 \end{aligned}$$

Esta distribución a posteriori es propia, ya que

$$\int_0^{+\infty} \int_0^{t_{(1)}} \pi(\lambda, \theta | t_i, L_i, \delta_i) d\theta d\lambda \leq \sqrt{n} \int_0^{+\infty} \int_0^{t_{(1)}} \lambda^{r-1} e^{-\lambda D(\theta)} d\theta d\lambda$$

y además

$$\begin{aligned} K &= \int_0^{+\infty} \int_0^{t_{(1)}} \lambda^{r-1} e^{-\lambda D(\theta)} d\theta d\lambda = \int_0^{+\infty} \frac{\lambda^{r-1}}{-\lambda D'(\theta)} \int_0^{t_{(1)}} (-\lambda D'(\theta) e^{-\lambda D(\theta)}) d\theta d\lambda = \int_0^{+\infty} \frac{\lambda^{r-1}}{\lambda n} e^{-\lambda D(\theta)} \Big|_0^{t_{(1)}} d\lambda = \\ &= \frac{1}{n} \int_0^{+\infty} \lambda^{r-2} e^{-D(t_{(1)})} d\lambda - \frac{1}{n} \int_0^{+\infty} \lambda^{r-2} e^{-D(0)} d\lambda = \\ &= \frac{1}{n} \left[ \frac{\Gamma(r-1)}{D(t_{(1)})^{r-1}} - \frac{\Gamma(r-1)}{D(0)^{r-1}} \right] = \frac{1}{n} \left[ \frac{\Gamma(r-1)}{(T - nt_{(1)})^{r-1}} - \frac{\Gamma(r-1)}{T^{r-1}} \right] \end{aligned}$$

Hemos de resaltar que para la resolución de la integral se ha supuesto que  $r \geq 2$ , es decir, que hemos observado al menos dos tiempos de vida sin censurar. Para plantear la verosimilitud, sólo se necesita suponer  $r \geq 1$ , pero la forma de la distribución a priori hace que sea necesario restringir a  $r \geq 2$ . Obviamente, esto es un inconveniente desde el punto de vista teórico para esta metodología, aunque, evidentemente, desde el punto de vista práctico no supone ninguna restricción, pues es

de suponer que tenemos un número de observaciones “razonablemente” grande y que de ellas, sólo algunas son censuradas.

El tratamiento algebraico de esta distribución a posteriori no es viable dada la imposibilidad de obtener explícitamente la constante de integración. Por ello, utilizaremos métodos de simulación (concretamente muestreo de Gibbs) para obtener una muestra de tamaño grande (por ejemplo, 1000) de la distribución a posteriori de  $(\lambda, \theta)$ , a partir de la cual podremos aproximar cualquier característica de interés de dicha distribución.

En este modelo, existe una distribución a priori que permite trabajar la distribución a posteriori algebraicamente; dicha distribución es  $\pi^*(\lambda, \theta) \propto 1/\lambda$ , que sería la distribución obtenida para un modelo sin censura y también para un modelo con censura pero en el que el número de datos no censurados  $r$  fuese considerado fijo. Como comprobaremos posteriormente, los resultados obtenidos al usar  $\pi^*(\lambda, \theta)$  y  $\pi(\lambda, \theta)$  son prácticamente idénticos, lo que permite en este ejemplo obtener expresiones aproximadas para las distintas características de interés de las distribuciones a posteriori tales como la media, varianza, percentiles, intervalos probabilísticos, etc. Además, el uso de  $\pi^*(\lambda, \theta)$  permite simular con bastante rapidez el comportamiento frecuencialista de los intervalos bayesianos, lo que sería mucho más costoso usando la distribución a priori  $\pi(\lambda, \theta)$  ya que tendríamos que obtener un número grande de intervalos (por ejemplo 1000) cada uno de los cuales requeriría de la obtención de una muestra de tamaño grande (por ejemplo 1000) de la distribución a posteriori correspondiente. Como veremos con posterioridad, el excelente comportamiento frecuencialista de los intervalos bayesianos obtenidos usando  $\pi^*(\lambda, \theta)$  junto con la excelente aproximación entre ambas distribuciones a posteriori (y por tanto entre los intervalos obtenidos a partir de ambas distribuciones), permitirá

concluir que los intervalos bayesianos obtenidos a partir de  $\pi(\lambda, \theta)$  se comportan, aproximadamente, como intervalos clásicos.

A continuación, vamos a ver las expresiones de las distribuciones marginales de  $\lambda$  y  $\theta$  que se obtienen al tomar como distribución a priori  $\pi^*(\lambda, \theta)$ .

**A) Análisis de la distribución a posteriori obtenida a partir de  $\pi^*(\lambda, \theta)$ .**

• **Distribución a posteriori conjunta y constante de integración:**

Con las notaciones utilizadas con anterioridad, la distribución a posteriori en este caso sería

$$\pi^*(\lambda, \theta | t_i, L_i, \delta_i) \propto \lambda^{r-1} e^{-\lambda D(\theta)}, \quad 0 \leq \theta \leq t_{(1)}, \quad \lambda > 0$$

por lo que la constante de integración, anteriormente calculada, es

$$K = \frac{1}{n} \left[ \frac{\Gamma(r-1)}{D(t_{(1)})^{r-1}} - \frac{\Gamma(r-1)}{D(0)^{r-1}} \right] = \frac{1}{n} \left[ \frac{\Gamma(r-1)}{(T - nt_{(1)})^{r-1}} - \frac{\Gamma(r-1)}{T^{r-1}} \right]$$

• **Distribución marginal de  $\theta$  e intervalo probabilístico:**

La densidad marginal a posteriori de  $\theta$  se obtiene sin ninguna dificultad. En efecto,

$$\pi^*(\theta | t_i, L_i, \delta_i) = \frac{1}{K} \int_0^{+\infty} \lambda^{r-1} e^{-\lambda D(\theta)} d\lambda = \frac{1}{K} \frac{\Gamma(r)}{D(\theta)^r} = \frac{(r-1)!}{K} (T - n\theta)^{-r}, \quad 0 \leq \theta \leq t_{(1)}$$

En este caso, podemos obtener la función de distribución a posteriori explícitamente, pues

$$\begin{aligned} F(\theta | t_i, L_i, \delta_i) &= \int_0^\theta \pi^*(u | t_i, L_i, \delta_i) du = \frac{(r-1)!}{K} \int_0^\theta D(u)^{-r} du = \\ &= \frac{(r-1)!}{K} \frac{1}{D'(u)} \int_0^\theta D'(u) D(u)^{-r} du = \\ &= -\frac{(r-1)!}{nK} \frac{D(u)^{-r+1}}{-r+1} \Big|_0^\theta = \frac{(r-1)!}{nK(r-1)} D(u)^{-(r-1)} \Big|_0^\theta = \frac{(r-2)!}{nK} \left[ \frac{1}{D(\theta)^{r-1}} - \frac{1}{D(0)^{r-1}} \right] \end{aligned}$$

Teniendo en cuenta que  $D(\theta) = T - n\theta$ , en este caso podemos resolver algebraicamente la ecuación  $F(\tilde{\theta}_\alpha | t_i, L_i, \delta_i) = \alpha$  y entonces  $(\tilde{\theta}_\alpha, t_{(1)})$  es un intervalo unilateral de probabilidad  $1-\alpha$ .

Para obtener el extremo del intervalo, basta pues resolver la ecuación

$$\frac{(r-2)!}{nK} \left[ \frac{1}{(T - n\tilde{\theta}_\alpha)^{r-1}} - \frac{1}{T^{r-1}} \right] = \alpha$$

Definiendo  $\tilde{u}_\alpha = T - n\tilde{\theta}_\alpha$ , (con lo que  $\tilde{\theta}_\alpha = (T - \tilde{u}_\alpha)/n$ ), tendríamos

$$\frac{(r-2)!}{nK} \left[ \frac{1}{\tilde{u}_\alpha^{r-1}} - \frac{1}{T^{r-1}} \right] = \alpha \Rightarrow \tilde{u}_\alpha = T \left( \frac{(r-2)!}{\alpha n K T^{r-1} + (r-2)!} \right)^{1/(r-1)}$$

y por tanto

$$\tilde{\theta}_\alpha = \frac{T - \tilde{u}_\alpha}{n} = \frac{T}{n} \left( 1 - \left( \frac{(r-2)!}{\alpha n K T^{r-1} + (r-2)!} \right)^{1/(r-1)} \right)$$



siendo, como se mencionó anteriormente, el intervalo unilateral de probabilidad  $1-\alpha$   $(\tilde{\theta}_\alpha, t_{(1)})$ .

• **Distribución marginal de  $\lambda$  e intervalo probabilístico:**

La función de densidad marginal a posteriori de  $\lambda$  viene dada por

$$\begin{aligned} \pi^*(\lambda | t_i, L_i, \delta_i) &= \frac{1}{K} \int_0^\theta \lambda^{r-1} e^{-\lambda D(\theta)} d\theta = \frac{1}{K} \frac{\lambda^{r-1}}{-\lambda D'(\theta)} \int_0^{t_{(1)}} -\lambda D'(\theta) e^{-\lambda D(\theta)} d\theta = \\ &= \frac{1}{K} \frac{\lambda^{r-2}}{n} e^{-\lambda D(\theta)} \Big|_0^{t_{(1)}} = \frac{\lambda^{r-2}}{nK} \left[ e^{-\lambda D(t_{(1)})} - e^{-\lambda D(0)} \right], \quad \lambda > 0 \end{aligned}$$

Para el parámetro  $\lambda$  no podemos obtener la función de distribución explícitamente (ni tampoco los extremos del intervalo probabilístico). No obstante, es conveniente hacer algunas manipulaciones que permiten expresarla a partir de las funciones de distribución de dos variables Gamma de distintos parámetros, ya que ello permite la resolución numérica de las ecuaciones que definen los extremos del intervalo con el programa Mathematica 3.0. en un tiempo "razonable" para tamaños muestrales grandes (la resolución directa de una ecuación del tipo  $\int_0^{\tilde{\lambda}_\alpha} \pi_\lambda^*(u | t_i, L_i, \delta_i) du = a$  requiere un tiempo excesivo sobre todo a medida que aumenta  $n$ , lo que hace inviable los procedimientos de simulación con un número suficiente de replicaciones).

Concretamente, la relación mencionada anteriormente es:

$$F(\lambda | t_i, L_i, \delta_i) = \int_0^\lambda \frac{1}{nK} u^{r-2} \left[ e^{-uD(t_{(1)})} - e^{-uD(0)} \right] du =$$

$$= \frac{1}{nK} \left[ \int_0^\lambda u^{r-2} e^{-uD(t_{(1)})} du - \int_0^\lambda u^{r-2} e^{-uD(0)} du \right] = \frac{1}{nK} \left[ \frac{\Gamma(r-1)}{D(t_{(1)})^{r-1}} F_{G1}(\lambda) - \frac{\Gamma(r-1)}{D(0)^{r-1}} F_{G2}(\lambda) \right]$$

donde  $G1 \sim \gamma(r-1, D(t_{(1)}))$  y  $G2 \sim \gamma(r-1, D(0))$ .

Para el parámetro  $\lambda$  vamos a plantearnos obtener el intervalo bilateral de probabilidad  $1-\alpha$  ( $\tilde{\lambda}_{\alpha/2}, \tilde{\lambda}_{1-\alpha/2}$ ) definido por las ecuaciones  $F(\tilde{\lambda}_{\alpha/2} | t_i, L_i, \delta_i) = \alpha/2$  y  $F(\tilde{\lambda}_{1-\alpha/2} | t_i, L_i, \delta_i) = 1-\alpha/2$ , que, usando para  $F(\lambda | t_i, L_i, \delta_i)$  la expresión dada anteriormente, son resueltas sin dificultad por el programa Mathematica 3.0.

• **Simulación del comportamiento frecuencialista de los intervalos bayesianos.**

A través de simulación, hemos estudiado el comportamiento frecuencialista de los intervalos bayesianos obtenidos anteriormente. Para ello, hemos utilizado el programa Mathematica 3.0. La simulación se ha efectuado para  $\alpha \in \{0.1, 0.05\}$ ,  $\theta \in \{5, 10\}$ ,  $\lambda \in \{0.25, 0.5, 1, 3, 5\}$  y  $n \in \{10, 20, 30, 40, 50\}$  haciendo en todos los casos  $m=1000$  replicaciones para simular la proporción de veces que los intervalos contienen a los parámetros respectivos. El listado correspondiente se aporta en el apéndice B.1.

Los resultados obtenidos se presentan en las tablas siguientes, donde  $NC(\theta)$  y  $NC(\lambda)$  son el nivel de confianza simulado de los intervalos bayesianos de probabilidad  $1-\alpha$  para los parámetros  $\theta$  y  $\lambda$  respectivamente.

$\theta=5 ; \alpha=0.1$										
$\lambda=1/4$		$\lambda=1/2$		$\lambda=1$		$\lambda=3$		$\lambda=5$		
	NC( $\theta$ )	NC( $\lambda$ )	NC( $\theta$ )	NC( $\lambda$ )	NC( $\theta$ )	NC( $\lambda$ )	NC( $\theta$ )	NC( $\lambda$ )	NC( $\theta$ )	NC( $\lambda$ )
n=10	0.925	0.877	0.908	0.892	0.916	0.877	0.919	0.895	0.920	0.879
n=20	0.890	0.897	0.914	0.904	0.903	0.904	0.914	0.877	0.914	0.904
n=30	0.895	0.895	0.904	0.903	0.902	0.896	0.904	0.911	0.922	0.891
n=40	0.912	0.893	0.907	0.909	0.898	0.902	0.898	0.899	0.904	0.894
n=50	0.896	0.906	0.909	0.912	0.914	0.895	0.899	0.875	0.910	0.894

**Tabla 4.I.a:** Nivel de confianza simulado de los intervalos Bayesianos de  $\theta$  y  $\lambda$

$\theta=10 ; \alpha=0.1$										
$\lambda=1/4$		$\lambda=1/2$		$\lambda=1$		$\lambda=3$		$\lambda=5$		
	NC( $\theta$ )	NC( $\lambda$ )	NC( $\theta$ )	NC( $\lambda$ )	NC( $\theta$ )	NC( $\lambda$ )	NC( $\theta$ )	NC( $\lambda$ )	NC( $\theta$ )	NC( $\lambda$ )
n=10	0.916	0.896	0.915	0.886	0.927	0.897	0.904	0.892	0.918	0.880
n=20	0.899	0.883	0.923	0.906	0.899	0.879	0.896	0.893	0.917	0.894
n=30	0.896	0.901	0.899	0.897	0.911	0.904	0.893	0.891	0.912	0.892
n=40	0.886	0.895	0.910	0.908	0.903	0.903	0.911	0.900	0.916	0.892
n=50	0.916	0.904	0.893	0.903	0.900	0.897	0.917	0.896	0.895	0.906

**Tabla 4.I.b:** Nivel de confianza simulado de los intervalos Bayesianos de  $\theta$  y  $\lambda$

$\theta=5 ; \alpha=0.05$										
$\lambda=1/4$		$\lambda=1/2$		$\lambda=1$		$\lambda=3$		$\lambda=5$		
	NC( $\theta$ )	NC( $\lambda$ )	NC( $\theta$ )	NC( $\lambda$ )	NC( $\theta$ )	NC( $\lambda$ )	NC( $\theta$ )	NC( $\lambda$ )	NC( $\theta$ )	NC( $\lambda$ )
n=10	0.945	0.926	0.952	0.947	0.957	0.936	0.957	0.950	0.958	0.943
n=20	0.946	0.952	0.951	0.932	0.958	0.951	0.956	0.944	0.945	0.942
n=30	0.957	0.959	0.957	0.955	0.946	0.962	0.949	0.940	0.958	0.948
n=40	0.944	0.957	0.952	0.948	0.957	0.949	0.955	0.952	0.966	0.953
n=50	0.947	0.949	0.963	0.949	0.956	0.938	0.956	0.946	0.962	0.943

**Tabla 4.I.c:** Nivel de confianza simulado de los intervalos Bayesianos de  $\theta$  y  $\lambda$

$\theta=10 ; \alpha=0.05$										
$\lambda=1/4$		$\lambda=1/2$		$\lambda=1$		$\lambda=3$		$\lambda=5$		
	NC( $\theta$ )	NC( $\lambda$ )	NC( $\theta$ )	NC( $\lambda$ )	NC( $\theta$ )	NC( $\lambda$ )	NC( $\theta$ )	NC( $\lambda$ )	NC( $\theta$ )	NC( $\lambda$ )
n=10	0.959	0.937	0.952	0.944	0.957	0.958	0.961	0.946	0.954	0.943
n=20	0.947	0.943	0.949	0.951	0.960	0.951	0.967	0.957	0.946	0.952
n=30	0.957	0.942	0.958	0.947	0.940	0.944	0.955	0.937	0.955	0.941
n=40	0.965	0.953	0.964	0.939	0.956	0.948	0.957	0.937	0.958	0.942
n=50	0.951	0.935	0.958	0.951	0.954	0.948	0.947	0.938	0.948	0.957

**Tabla 4.1.d:** Nivel de confianza simulado de los intervalos Bayesianos de  $\theta$  y  $\lambda$ .

Como podemos ver, la aproximación es excelente aún para tamaño muestral pequeño, no apreciándose diferencias significativas en cuanto al comportamiento para los distintos valores considerados de  $\alpha$ ,  $\theta$ ,  $\lambda$  y  $n$ .

**B) Análisis de la distribución a posteriori obtenida a partir de  $\pi(\lambda, \theta)$ .**

En este caso, ya hemos señalado cuál sería la distribución a posteriori y cómo en esta situación el tratamiento algebraico no es viable. Por ello, vamos a aplicar muestreo de Gibbs para simular una muestra de la distribución a posteriori. Comenzaremos proporcionando una breve descripción del procedimiento de simulación que vamos a utilizar.

Sea  $\theta = (\theta_1, \dots, \theta_k)^t$  el vector de parámetros y denotemos por  $\pi(\theta)$  a la distribución a posteriori de la que se pretende simular una muestra. Supongamos que tenemos procedimientos para simular muestras de todas las distribuciones condicionales completas  $\pi(\theta_i | \theta_{\{i\}})$ ,  $i=1..k$ , donde  $\theta_{\{i\}}$  representa el vector  $(\theta_1, \dots, \theta_{i-1}, \theta_{i+1}, \dots, \theta_k)^t$ . Entonces, partiendo de valores iniciales  $\theta_{\{i\}}^{(0)}$  se genera  $\theta_1^{(1)}$  a partir de  $\pi(\theta_1 | \theta_{\{1\}}^{(0)})$ ,  $\theta_2^{(1)}$  a partir de  $\pi(\theta_2 | \theta_1^{(1)}, \theta_3^{(0)}, \dots, \theta_k^{(0)})$ ,  $\theta_3^{(1)}$  a partir de

$\pi(\theta_3 | \theta_1^{(1)}, \theta_2^{(1)}, \theta_4^{(0)}, \dots, \theta_k^{(0)})$ , y así sucesivamente hasta obtener  $\theta_k^{(1)}$  a partir de  $\pi(\theta_k | \theta_{\{k\}}^{(1)})$  con lo que se obtiene  $\theta^{(1)} = (\theta_1^{(1)}, \dots, \theta_k^{(1)})^t$ , momento en el cual puede volverse a repetir todo el procedimiento. Bajo condiciones muy generales, la sucesión  $\{\theta^{(i)}\}$  así construida es una realización de una cadena de Markov con la distribución a posteriori como distribución estacionaria (Gelfand, A.E. and Smith, A.F. (1990)).

La muestra así obtenida no es aleatoria, por lo que para estudiar las características de la distribución a posteriori habríamos de aplicar técnicas de estimación en cadenas de Markov (Hastings, W.K., (1970)). Para obtener una muestra aleatoria pueden ser utilizados dos tipos de procedimientos. El primero consiste en construir distintas cadenas partiendo de valores iniciales elegidos independientemente y tomar los valores obtenidos en cada una de las cadenas tras un número suficientemente grande de pasos como para garantizar que se ha alcanzado el estado estacionario (Gelfand, A.E. and Smith, A.F., (1990)). La otra opción consiste en construir una única cadena y, una vez alcanzado el estado estacionario, elegir los valores obtenidos cada  $m$  etapas donde  $m$  es suficientemente grande como para poder considerar que  $\theta^{(k)}$  y  $\theta^{(k+m)}$  son independientes (Geman, S. and Geman, D., (1984)). Si se pretende obtener una muestra de tamaño  $n$  y se considera alcanzado el estado estacionario tras  $k$  etapas en cada cadena, el primer método requiere la obtención de  $kn$  valores de  $\theta$ , mientras que el segundo método requeriría  $k+nm$ . Teniendo en cuenta que el valor de  $k$  requerido suele ser bastante superior al de  $m$  (suele bastar  $m=10$  ó  $15$ ), el primer método requiere de un número muy superior de operaciones y por consiguiente de tiempo de computación que el segundo.

Hasta ahora, no existen resultados teóricos generales sobre cuándo se puede considerar estacionaria la cadena de Markov; lo habitual es basarse en representaciones gráficas de diversas características de la cadena que permitan visualizar si se está o no próximo al estado estacionario. Sigue siendo un tema de investigación la búsqueda de métodos que, de forma automática, sugieran el número

de etapas que hemos de desechar. Tampoco está claro si es preferible la utilización de una o de  $n$  cadenas independientes.

Según lo descrito anteriormente, hemos de obtener procedimientos para simular ambas distribuciones condicionadas; dichos procedimientos son los que se especifican a continuación.

• **Generación de datos de la distribución  $\lambda|\theta$ .**

El núcleo de la densidad de la distribución condicionada que pretendemos simular vendrá dada por

$$f(\lambda|\theta) = \lambda^{r-1} \sqrt{\sum_i (1 - e^{-\lambda(L_i - \theta)})} e^{-\lambda C}, \quad \lambda > 0$$

donde se ha denotado  $C = T - n\theta$ .

Vamos a usar el método de aceptación-rechazo (ver, por ejemplo, Smith, A.F. and Gelfand, A.E., (1992)). Para ello, consideremos la densidad de una variable Gamma de parámetros  $r$  y  $C$ , es decir,

$$g(\lambda) = \frac{C^r}{\Gamma(r)} \lambda^{r-1} e^{-\lambda C}$$

Se verifica que el cociente  $f(\lambda|\theta)/g(\lambda)$  está acotado  $\forall \lambda > 0$ , ya que

$$\frac{f(\lambda|\theta)}{g(\lambda)} = \frac{\lambda^{r-1} e^{-\lambda C} \sqrt{\sum_i (1 - e^{-\lambda(L_i - \theta)})}}{\frac{C^r}{\Gamma(r)} \lambda^{r-1} e^{-\lambda C}} \leq \frac{\sqrt{n}}{C^r} \Gamma(r) = M$$

Así, generamos dos valores  $\lambda_0 \sim \Gamma(r, C)$  y  $u_0 \sim U(0,1)$  y aceptamos el valor de  $\lambda_0$  si se verifica

$$u_0 \leq \frac{f(\lambda_0 | \theta)}{Mg(\lambda_0)} = \frac{\lambda_0^{r-1} e^{-\lambda_0 C} \sqrt{\sum (1 - e^{-\lambda_0(L_i - \theta)})}}{\frac{\sqrt{n}}{C^r} \Gamma(r) \frac{C^r}{\Gamma(r)} \lambda_0^{r-1} e^{-\lambda_0 C}} = \sqrt{\frac{1}{n} \sum (1 - e^{-\lambda_0(L_i - \theta)})}$$

Es importante hacer notar que la probabilidad de aceptar el valor  $\lambda_0$  generado en general va a ser bastante alta, pues los términos  $1 - e^{-\lambda_0(L_i - \theta)}$  con frecuencia serán cercanos a la unidad. No obstante, la mejor cota que se puede conseguir para la probabilidad de aceptar es  $\sqrt{1/n}$  (que, por ejemplo, para  $n=30$  vale aproximadamente 0,1825).

- **Generación de datos de la distribución  $\theta|\lambda$ .**

El núcleo de la densidad de la distribución condicionada que pretendemos simular vendrá dado por

$$f(\theta | \lambda) \propto \sqrt{\sum_i (1 - e^{-\lambda(L_i - \theta)})} e^{n\lambda\theta}, \quad 0 < \theta < t_{(1)}$$

De nuevo, usaremos el método de aceptación-rechazo, considerando para ello la densidad  $g(\theta) \propto e^{n\lambda\theta}$ ,  $0 \leq \theta \leq t_{(1)}$ . El cálculo directo de la constante de integración lleva a

$$g(\theta) = \frac{n\lambda}{e^{n\lambda t_{(1)}} - 1} e^{n\lambda\theta} \quad 0 \leq \theta \leq t_{(1)}$$

Esta densidad, puede ser simulada usando el teorema de la transformación integral, ya que podemos obtener sin dificultad tanto la función de distribución de la variable como su inversa. En efecto,

$$G(\theta) = \int_0^\theta \frac{n\lambda}{e^{n\lambda t_{(1)}} - 1} e^{n\lambda z} dz = \frac{e^{n\lambda\theta} - 1}{e^{n\lambda t_{(1)}} - 1}$$

y además,

$$U = \frac{e^{n\lambda\theta} - 1}{e^{n\lambda t_{(1)}} - 1} \Leftrightarrow 1 + U(e^{n\lambda t_{(1)}} - 1) = e^{n\lambda\theta} \Leftrightarrow \theta = \frac{1}{n\lambda} \text{Ln}\left(1 + U(e^{n\lambda t_{(1)}} - 1)\right)$$

Así, generamos un valor  $u_0 \sim U(0,1)$  y a partir de él obtenemos, una realización de la densidad  $g(\theta)$  que será  $\theta_0 = \frac{1}{n\lambda} \text{Ln}\left(1 + u_0(e^{n\lambda t_{(1)}} - 1)\right)$ .

Como en el caso anterior, hemos de comprobar que el cociente  $f(\theta|\lambda)/g(\theta)$  está acotado  $\forall \theta \in [0, t_{(1)}]$ . En efecto,

$$\frac{f(\theta|\lambda)}{g(\theta)} = \frac{e^{n\lambda\theta} \sqrt{\sum (1 - e^{-\lambda(L_i - \theta)})}}{\frac{n\lambda}{e^{n\lambda t_{(1)}} - 1} e^{n\lambda\theta}} \leq \frac{\sqrt{n} (e^{n\lambda t_{(1)}} - 1)}{n\lambda} = M$$

y por tanto, una vez generado  $\theta_0$  hemos de obtener otro valor  $u_0 \sim U(0,1)$  y entonces el valor  $\theta_0$  será aceptado si

$$u_0 \leq \frac{f(\theta_0|\lambda)}{Mg(\theta_0)} = \frac{e^{n\lambda\theta_0} \sqrt{\sum (1 - e^{-\lambda(L_i - \theta_0)})}}{\frac{\sqrt{n} (e^{n\lambda t_{(1)}} - 1)}{n\lambda} \frac{n\lambda}{e^{n\lambda t_{(1)}} - 1} e^{n\lambda\theta_0}} = \sqrt{\frac{1}{n} \sum (1 - e^{-\lambda(L_i - \theta_0)})}$$



Para implementar el proceso de simulación, se ha usado el programa Mathematica 3.0, construyendo una única cadena de Markov, de la que se han desechado 1000 realizaciones, momento a partir del cual hemos elegido los valores obtenidos cada 15 etapas, obteniendo una muestra de la distribución final de tamaño 1000. Como punto de partida se ha elegido  $\theta^{(0)} = 0,9 \cdot t_{(1)}$ . El listado correspondiente se aporta en el apéndice B.2.

**C) Comparación de los resultados obtenidos usando  $\pi(\lambda, \theta)$  y  $\pi^*(\lambda, \theta)$**

Un histograma de la muestra simulada así como varias características de la misma han sido obtenidos para varios valores de  $\theta$  y  $\lambda$ , comparándolos con los resultados proporcionados por las fórmulas correspondientes usando la distribución a priori  $\pi^*(\lambda, \theta)$ , según se muestra en las tablas y gráficos que siguen.

	Int. probab. usando $\pi$	Int. probab. usando $\pi^*$
$\theta = 5; \lambda = 1; n = 10;$	$\theta \in (4.83359, 5.04808)$	$\theta \in (4.83685, 5.04808)$
	$\lambda \in (0.77966, 2.97176)$	$\lambda \in (0.76946, 2.94729)$
$\theta = 5; \lambda = 1; n = 30;$	$\theta \in (4.95339, 5.04205)$	$\theta \in (4.95662, 5.04205)$
	$\lambda \in (0.76185, 1.70094)$	$\lambda \in (0.80363, 1.77380)$
$\theta = 5; \lambda = 4; n = 30;$	$\theta \in (4.97698, 5.00339)$	$\theta \in (4.97815, 5.00339)$
	$\lambda \in (2.52517, 6.15752)$	$\lambda \in (2.65622, 6.18443)$
$\theta = 5; \lambda = 0.25; n = 30;$	$\theta \in (4.57938, 5.00644)$	$\theta \in (4.60071, 5.00644)$
	$\lambda \in (0.16305, 0.37497)$	$\lambda \in (0.16796, 0.37694)$

**Tabla 3.II:** Comparación de los intervalos probabilísticos obtenidos con  $\pi(\lambda, \theta)$  y  $\pi^*(\lambda, \theta)$

A continuación mostramos las gráficas de las funciones de densidad marginales obtenidas usando  $\pi^*(\lambda, \theta)$  y los histogramas correspondientes a las simulaciones usando  $\pi(\lambda, \theta)$  para los mismos valores de  $\theta$ ,  $\lambda$  y  $n$  para los que se han

obtenido los intervalos, observándose en todos los casos que la aproximación es excelente. El listado correspondiente se aporta en el apéndice B.2.

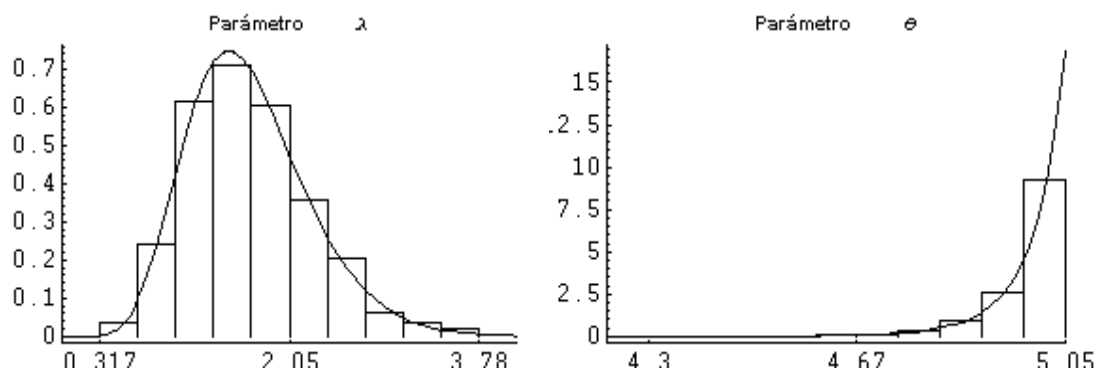


Figura 4.1.a: Func. de densidad e histograma para  $\theta = 5$ ;  $\lambda = 1$ ;  $n = 10$ ;

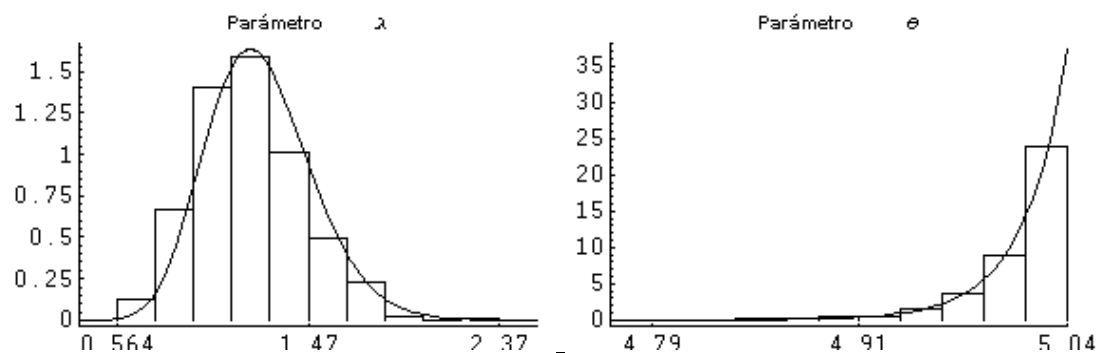


Figura 4.1.b: Func. de densidad e histograma para  $\theta = 5$ ;  $\lambda = 1$ ;  $n = 30$ ;

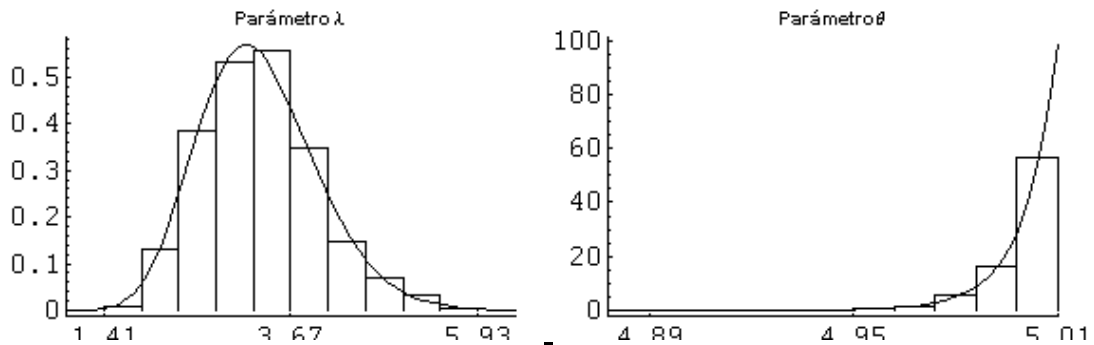


Figura 4.1.c: Func. de densidad e histograma para  $\theta = 5$ ;  $\lambda = 4$ ;  $n = 30$ ;

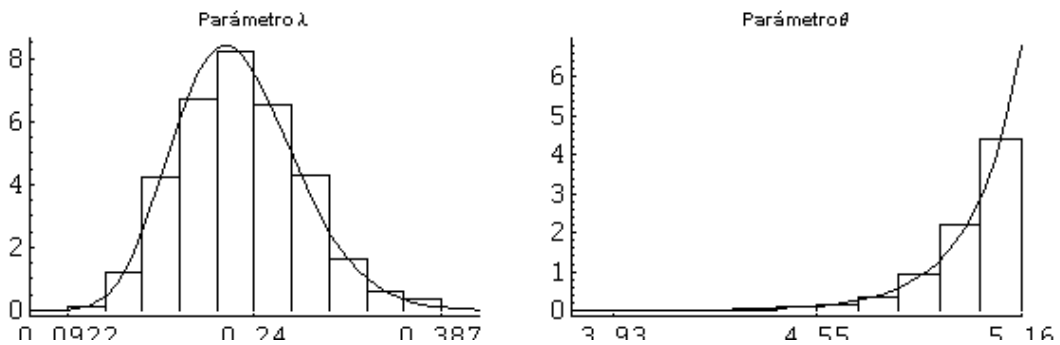


Figura 4.1.d: Func. de densidad e histograma para  $\theta = 5$ ;  $\lambda = 0.25$ ;  $n = 30$ ;

### 4.3 El modelo Weibull .

Uno de los modelos más utilizados en los datos de supervivencia es el modelo Weibull dada su flexibilidad. Como ya se ha indicado, el modelo exponencial es más sencillo de analizar. No obstante, es restrictivo en el sentido de que sólo posee un parámetro y se caracteriza por su función de riesgo constante. El modelo Weibull es más general, incluye al modelo Exponencial como caso particular y, según los valores de los parámetros, se adapta a situaciones de riesgo creciente, decreciente o constante. Otra de las características importantes es que la función de supervivencia podemos

obtenerla de forma explícita, lo que facilita la manipulación de la verosimilitud en presencia de censura tipo I.

Como se indicó con anterioridad, decimos que la variable  $T$  sigue un modelo Weibull de parámetros  $\lambda > 0$  y  $\beta > 0$  si su función de densidad es de la forma

$$f(t) = \lambda\beta(\lambda t)^{\beta-1} \exp\{-(\lambda t)^\beta\}, \quad t > 0$$

Otra ventaja que presenta el modelo Weibull es que podemos trabajar en vez de con los datos habituales de los tiempos de supervivencia con sus logaritmos, obteniendo de esta forma un modelo de localización-escala. En efecto, si definimos  $X = \text{Log}T$ , la función de densidad de  $X$  será

$$f(x) = \frac{1}{b} \exp\left\{\frac{x-u}{b} - \exp\left[\frac{x-u}{b}\right]\right\}, \quad x \in \mathbb{R}$$

donde la relación entre los parámetros viene dada por  $u = -\log \lambda$  y  $b = 1/\beta$ . De hecho, la variable  $X$  sigue una distribución de valor extremo de parámetros  $u \in \mathbb{R}$  y  $b > 0$ .

A continuación, vamos a analizar la resolución de este modelo en los casos de datos sin censura y de presencia de censura tipo I.

#### **4.3.1 El modelo Weibull sin censura.**

Supongamos que los tiempos de vida  $T_i$ ,  $i=1\dots n$  son independientes e idénticamente distribuidos según un modelo Weibull de parámetros  $\lambda$  y  $\beta$ . Consideremos los logaritmos de dichos tiempos de vida,  $X_i = \text{Log}T_i$  que serán

independientes e idénticamente distribuidos según un modelo de valor extremo de parámetros  $u$  y  $b$ . Así, la verosimilitud correspondiente vendrá dada por

$$L(u, b | x_i) \propto \left(\frac{1}{b}\right)^n \exp\left\{\sum_{i=1}^n \left[\frac{x_i - u}{b} - e^{\frac{x_i - u}{b}}\right]\right\} = \left(\frac{1}{b}\right)^n \exp\left\{\sum_{i=1}^n \frac{x_i - u}{b} - \sum_{i=1}^n e^{\frac{x_i - u}{b}}\right\}$$

Haciendo un cambio de origen en los datos de forma que se tenga  $\sum_{i=1}^n x_i = 0$  (es decir considerando los datos en desviaciones respecto a la media), podemos simplificar algo la verosimilitud, quedando ahora

$$L(u, b | x_i) \propto \left(\frac{1}{b}\right)^n \exp\left\{-\sum_{i=1}^n e^{\frac{x_i - u}{b}}\right\} e^{-\frac{nu}{b}}$$

La distribución a priori para este modelo, puesto que es de localización-escala y regular vendrá dada por  $\pi(u, b) \propto 1/b$ , con lo que la distribución a posteriori para el parámetro es

$$\pi(u, b | x_i) \propto \left(\frac{1}{b}\right)^{n+1} \exp\left\{-\sum_{i=1}^n e^{\frac{x_i - u}{b}}\right\} e^{-\frac{nu}{b}}$$

Según se indicó en el capítulo 3, sección 2.1, al ser este modelo de localización-escala y no existir un estadístico suficiente de la misma dimensión que el espacio paramétrico, los intervalos bayesianos obtenidos a partir de la distribución a priori  $\pi(u, b) \propto 1/b$  coincidirán numéricamente con los intervalos obtenidos a partir de la Inferencia Condicional (Welsh, A.H. (1996)). Así, para obtener los intervalos podríamos aplicar las expresiones correspondientes a los intervalos condicionados. No obstante, dichas expresiones conllevan la resolución de ecuaciones del tipo

$$\int_{-\infty}^u \int_0^{+\infty} g(v, t) dt dv = \alpha; \quad \int_0^v \int_{-\infty}^{+\infty} g(v, t) dv dt = \alpha$$

donde las integrales a su vez hemos de resolverlas por métodos numéricos. Además, para obtener las estimaciones de los parámetros a través de las medias marginales habríamos de resolver nuevas integrales numéricas. Por ello, optaremos para la obtención de resultados en este modelo por la simulación a través del muestreo de Gibbs.

El proceso de simulación para este caso es bastante parecido al que hay que desarrollar en presencia de censura tipo I (aunque algo más simple). Por ello, la explicación detallada del mismo se hará en el epígrafe siguiente y esta descripción será válida para el caso de datos sin censura con la única diferencia de que en el modelo con censura la verosimilitud posee un término adicional. Este procedimiento ha sido implementado usando el programa Mathematica 3.0. El listado correspondiente se aporta en el apéndice B.3.

#### **4.3.2 El modelo Weibull con censura tipo I.**

Supongamos que los tiempos de vida  $T_i$ ,  $i=1\dots n$  son independientes e idénticamente distribuidos según un modelo Weibull de parámetros  $\lambda > 0$  y  $\beta > 0$  y que los datos son observados si  $T_i \leq L_i$ . Consideremos  $X_i = \text{Log}T_i$  y  $\eta_i = \text{Log}L_i$ , con lo que ahora los datos serán independientes e idénticamente distribuidos según un modelo de valor extremo de parámetros  $u = -\text{Log}\lambda$  y  $b = \beta^{-1}$ , siendo observados los datos que verifiquen  $X_i \leq \eta_i$ . La verosimilitud asociada a esta muestra de tamaño  $n$ , como hemos descrito en el epígrafe 4.1.3 vendrá por

$$L(u, b | x_i, \eta_i, \delta_i) = \prod_{i=1}^n f(x_i, u, b)^{\delta_i} S(\eta_i, u, b)^{1-\delta_i}$$

que aplicada a nuestro caso concreto queda

$$L(u, b | x_i, \eta_i, \delta_i) \propto \prod_{i=1}^n \left( \frac{1}{b} \exp \left\{ \frac{x_i - u}{b} - e^{\frac{x_i - u}{b}} \right\} \right)^{\delta_i} \left( \exp \left\{ -e^{\frac{\eta_i - u}{b}} \right\} \right)^{1 - \delta_i} \propto \left( \frac{1}{b} \right)^r \exp \left\{ \sum_{i=1}^n \left( \frac{x_i - u}{b} \right) \delta_i \right\} \exp \left\{ - \sum_{i=1}^n e^{\frac{x_i - u}{b}} \delta_i - \sum_{i=1}^n e^{\frac{\eta_i - u}{b}} (1 - \delta_i) \right\}$$

donde  $r = \sum_{i=1}^n \delta_i$  es el número de datos no censurados. Por comodidad en la notación, podemos suponer que los  $r$  primeros datos son los no censurados y además vamos a definir

$$T(u, b) = \sum_{i=1}^n e^{\frac{x_i - u}{b}} \delta_i + \sum_{i=1}^n e^{\frac{\eta_i - u}{b}} (1 - \delta_i) = \sum_{i=1}^n e^{\frac{z_i - u}{b}}$$

donde  $z_i = \text{Min}\{x_i, \eta_i\}$ , es decir,  $\{z_i\}_{i=1}^n$  representa el conjunto de todas las observaciones, sean estas censuradas o no. Con la suposición anterior y la notación introducida, la verosimilitud puede escribirse como

$$L(u, b | x_i, \eta_i, \delta_i) \propto \left( \frac{1}{b} \right)^r \exp \left\{ \sum_{i=1}^r \frac{x_i - u}{b} \right\} e^{-T(u, b)}$$

Además, vamos a efectuar un cambio de origen en los datos, restando a todos ellos (censurados y sin censurar) la media de los datos no censurados, con lo que obtendremos que  $\sum_{i=1}^r x_i = 0$  y la verosimilitud será entonces

$$L(u, b | x_i, \eta_i, \delta_i) \propto \left( \frac{1}{b} \right)^r e^{-\frac{ur}{b}} e^{-T(u, b)}$$

Para la obtención de la distribución a priori hemos de observar, como en el caso exponencial, que el número de datos no censurados es aleatorio y que la esperanza correspondiente es

$$E[r] = \sum_{i=1}^n P[\delta_i = 1] = \sum_{i=1}^n P[X_i \leq \eta_i] = \sum_{i=1}^n \left( 1 - \exp \left\{ -e^{-\frac{\eta_i - u}{b}} \right\} \right)$$

y por tanto, al tratarse de un modelo de localización-escala la distribución a priori sería

$$\pi(u, b) \propto \frac{1}{b} \sqrt{\sum_{i=1}^n \left( 1 - \exp \left\{ -e^{-\frac{\eta_i - u}{b}} \right\} \right)}$$

Como consecuencia, la distribución a posteriori conjunta viene dada por

$$\begin{aligned} \pi(u, b | x_i, \eta_i, \delta_i) &\propto \left( \frac{1}{b} \right)^{r+1} e^{-\frac{ur}{b}} e^{-T(u,b)} \sqrt{\sum_{i=1}^n \left( 1 - \exp \left\{ -e^{-\frac{\eta_i - u}{b}} \right\} \right)} \\ &\propto \left( \frac{1}{b} \right)^{r+1} e^{-\frac{ur}{b}} e^{-T(u,b)} \sqrt{\sum_{i=1}^n F_X(\eta_i, u, b)} \end{aligned}$$

donde  $F_X(x, u, b)$  es la función de distribución del modelo de valor extremo de parámetros  $u$  y  $b$ .

Usaremos muestreo de Gibbs para obtener una muestra de la densidad conjunta a posteriori de  $(u, b)$ , a partir de la cual podremos obtener cualquier característica de interés del modelo. Para ello, necesitamos poder simular las distribuciones  $u | b$  y  $b | u$ . Dichas distribuciones condicionadas vienen dadas por:



$$\pi(u | b, x_i, \eta_i, \delta_i) = \pi(u | b) \propto e^{-\frac{ur}{b}} e^{-T(u,b)} \sqrt{\sum_{i=1}^n F_X(\eta_i, u, b)}, \quad u \in \mathbb{R}$$

$$\pi(b | u, x_i, \eta_i, \delta_i) = \pi(b | u) \propto \left(\frac{1}{b}\right)^{r+1} e^{-\frac{ur}{b}} e^{-T(u,b)} \sqrt{\sum_{i=1}^n F_X(\eta_i, u, b)}, \quad b > 0$$

Observemos en primer lugar que  $0 \leq \sqrt{\sum_{i=1}^n F_X(\eta_i, a, b)} \leq \sqrt{n}$ . Así, una vez simuladas las distribuciones

$$\pi_1(u | b) \propto e^{-\frac{ur}{b}} e^{-T(u,b)}, \quad u \in \mathbb{R}$$

$$\pi_2(b | u) \propto \left(\frac{1}{b}\right)^{r+1} e^{-\frac{ur}{b}} e^{-T(u,b)}, \quad b > 0$$

podemos obtener muestras de  $\pi(u | b)$  y  $\pi(b | u)$  aplicando el método de aceptación-rechazo como vimos en el caso del modelo exponencial con censura tipo I. Además, la proporción de rechazo es suficientemente pequeña como para que el método resulte operativo.

Teniendo en cuenta que  $T(u, b) = \sum_{i=1}^n \exp\{(z_i - u)/b\} \geq 0$ , obtenemos que  $|\exp\{-T(u, b)\}| \leq 1$ , por lo que podríamos pensar en llevar a cabo la simulación sin tener en cuenta este término y posteriormente aplicar aceptación-rechazo. No obstante, en este caso la proporción de rechazos sería tan elevada que el método no resultaría operativo.

Así, hemos de simular muestras de  $\pi_1(u|b)$  y  $\pi_2(b|u)$  que como puede apreciarse no responden a ningún modelo conocido. Tanto en uno como en otro caso hemos optado por obtener numéricamente el máximo ( $M$ ) de los núcleos de las densidades y un rango  $(a,b)$  suficientemente amplio que contenga “casi toda” la masa de probabilidad. Luego, tras generar un par de valores  $(z_0, v_0) \sim U((a, b) \times (0, M))$  el valor  $z_0$  es aceptado como una realización de la densidad, por ejemplo,  $\pi_1(u|b)$  si  $\pi_1(z_0|b) \leq v_0$  (análogamente, para  $\pi_2(b|u)$  habría de cumplirse  $\pi_2(z_0|u) \leq v_0$ ). Como puede observarse, estamos aplicando el método de aceptación-rechazo tomando como distribución de apoyo la uniforme (lo que obliga a trabajar en un rango finito y de ahí la necesidad de buscar un intervalo finito en el que esté un porcentaje alto de la masa de probabilidad de la distribución simulada).

El procedimiento descrito se ha implementado usando el programa Mathematica 3.0. El listado correspondiente se aporta en el apéndice B.4.

## **CAPÍTULO 5: APLICACIÓN A LA EVALUACIÓN DE LAS REVISTAS CIENTÍFICAS.**

En este capítulo vamos a aplicar la metodología descrita con anterioridad al campo de la Bibliometría, centrándonos en el estudio de la distribución de la antigüedad de las citas en la literatura científica.

El indicador más comúnmente utilizado para evaluar y clasificar a las revistas científicas es el Factor de Impacto, que para cada año sólo tiene en cuenta las citas hechas a artículos publicados en los dos años anteriores. En general, existe tendencia a considerar sólo citas a artículos recientes. Esto se pone de manifiesto, por ejemplo, en el hecho de que en las bases de datos del ISI (“Institute Scientific Information”) las citas con 10 o más años de antigüedad aparecen agregadas.

En nuestra aplicación, uno de los objetivos es efectuar una evaluación o comparación entre las revistas que tenga en cuenta “toda la historia” en cuanto a las citas recibidas y no sólo “la historia reciente”. Para ello hemos buscado modelos probabilísticos adecuados para describir la variable que nos proporciona la antigüedad de las citas recibidas por una determinada revista científica.

Si utilizamos los datos que proporciona directamente el ISI, necesitamos introducir modelos censurados. Dado el alto porcentaje de censura que presentan las distribuciones, hemos elaborado estrategias de búsqueda para poder disponer de la distribución completa correspondiente a las revistas que hemos analizado. Estos últimos datos proceden de un número menor de revistas fuente pero permiten utilizar modelos sin censura.

### **5.1 El uso de las citas para la evaluación de las revistas científicas.**

La característica fundamental que hace que un determinado trabajo pueda calificarse como documento científico ha de ser el hecho de que haya sido evaluado de forma positiva por otros científicos. En efecto, el investigador debe someterse a la crítica de sus colegas para que el resultado de sus investigaciones pueda difundirse y sea considerado creíble. Es la comunidad científica la que debe decidir sobre el interés y la solidez de los enunciados teóricos o de los resultados experimentales presentados.

Para poder someter a prueba los trabajos elaborados, se hace necesaria la existencia de un espacio público de debate. Éste se ha ido constituyendo progresivamente, pero podemos citar como momento crucial la creación de las academias de ciencias en varios países en el siglo XVII. Así se constituyó y se organizó poco a poco lo que actualmente llamamos “comunidad científica” que permite a los investigadores debatir entre ellos sus respectivos trabajos a través fundamentalmente de las revistas científicas.

Cuando un investigador escribe un documento científico, su contenido suele apoyarse en la mayoría de los casos en los contenidos de otros trabajos anteriores. Estos otros trabajos aparecen recogidos en la sección del documento denominada Bibliografía. Este comportamiento de apoyarse en otros investigadores produce las denominadas *citas* a trabajos escritos anteriormente por otros científicos.

La evaluación de la importancia de un determinado documento científico, generalmente, se hace sobre la base del principio de que un trabajo será considerado tanto más importante cuantas más veces sea citado por el resto de la comunidad científica. A partir de aquí, parece razonable que por agregación, una revista científica que reciba muchas citas sea considerada como “muy importante”. Bajo este principio subyace la idea de que si una revista recibe muchas citas será porque aporta unos conocimientos que son útiles a los investigadores.

Es evidente que la evaluación sobre la base de las citas recibidas presenta varios inconvenientes, derivados principalmente del hecho de que no siempre una cita a un trabajo está asociada a la utilidad del contenido del mismo. La explicación de muchas de estas objeciones puede encontrarse en la respuesta a la siguiente pregunta: ¿qué motivos empujan a un autor a citar trabajos ya existentes?.

Aparte de la que se supone causa principal y más frecuente (es decir, citamos un determinado trabajo porque los resultados en él expuestos han sido útiles para nuestra investigación), en Callon, M. Et al. (1995) se citan otros motivos:

- i) Se tratará de mencionar resultados publicados anteriormente con el fin de convencer a nuestros lectores y hacer más creíbles nuestros asertos.
- ii) Podemos hacer referencia a un trabajo para ponerlo en tela de juicio, para matizar su alcance o para rebatirlo (es lo que llamaríamos una “cita negativa”).
- iii) Podemos efectuar una cita para poner en evidencia que nuestro interés no se limita a los trabajos expuestos o para demostrar la amplitud de nuestros conocimientos (es lo que podríamos denominar una “cita formal”).
- iv) Podemos efectuar una cita para complacer a algún colega o una “autocita” para complacernos a nosotros mismos (es lo que podríamos llamar una “cita complaciente”).

Es evidente que los motivos ii), iii) y iv) pueden distorsionar cualquier análisis hecho sobre la base de las citas recibidas. Podemos indicar también el inconveniente derivado del hecho de que un trabajo muy importante puede dejar de ser citado porque el contenido del mismo ya se dé por conocido y admitido, si bien esto suele ocurrir a largo plazo.

Para analizar el alcance de una cita determinada habría que evaluar también el contexto de la cita, es decir, la o las frases que en el texto acompañan a la cita. Podemos determinar si es positiva (los resultados del artículo citado son admitidos sin discusión) o bien si su presencia obedece simplemente a una cuestión formal, la cita

tiene connotaciones negativas, etc. No obstante, como parece obvio, desde un punto de vista práctico esto no es viable, ya que los datos de que dispondremos sólo pondrán de relieve que la cita se ha producido. Por tanto, la evaluación se lleva a cabo sobre la base de las citas recibidas, teniendo presente que pueden darse ciertas distorsiones. Parece comúnmente aceptado que este tipo de análisis y los indicadores en él utilizados son más una medida de la **visibilidad** que alcanzan los trabajos o revistas que una medida de su calidad.

Para estudiar el comportamiento de la edad o la antigüedad de las citas o referencias, como puede verse en Gupta, B.M. (1997,1998), podemos utilizar dos enfoques diferentes: el sincrónico y el diacrónico.

Los estudios sincrónicos se llevan a cabo usando las referencias que aparecen en un conjunto fuente de revistas o trabajos en un instante fijado de tiempo (generalmente, en un año prefijado). A partir de ellas, se obtiene la distribución de la edad o antigüedad de los materiales citados. En otras palabras, se fija la literatura que cita, es decir, que recoge en sus documentos referencias o bibliografía, y la edad de los trabajos citados se computa **hacia atrás**.

Los estudios diacrónicos están relacionados con la observación de un ítem particular a través de sucesivos períodos de tiempo. En el caso diacrónico, los recuentos de citaciones se efectúan **hacia delante**.

Habitualmente, suele usarse el estudio sincrónico ya que la obtención de datos es más directa. En el presente trabajo, usaremos esta forma de obtener los datos.

Para finalizar este epígrafe, indiquemos que los análisis bibliométricos basados en las citas pueden clasificarse en dos grandes categorías, según que conduzcan a indicadores de actividad (o calidad, visibilidad, etc.) o a indicadores de relación. Los primeros se centran en análisis individualizados (bien sea de autores, revistas, temáticas, instituciones, países, etc.) mientras que los segundos se ocupan de las

interrelaciones existentes entre las distintas unidades de un determinado grupo. En el caso de las revistas científicas, con el primer tipo de actividad el propósito sería construir indicadores asociados a cada revista o ítem; en este caso necesitaríamos conocer cuántas citas han recibido cada una de las revistas estudiadas. Con el segundo tipo de actividad nos ocuparíamos del intercambio de citas entre las revistas del grupo objeto de análisis; necesitaríamos entonces conocer no sólo el número de citas recibidas por una revista A, sino cuántas de ellas proceden de la propia revista A (autocitas), de la B, de la C, etc. En el presente trabajo nos ocuparemos sólo de los análisis individualizados.

## **5.2 El Factor de Impacto y el proceso de obsolescencia.**

Actualmente, la relevancia de una determinada revista científica se establece fundamentalmente sobre la base del llamado factor de impacto (FI) de la misma. El FI de una revista A en el año t se define como el número de citas que recibe dicha revista en el año t correspondientes a artículos publicados en los años t-1 y t-2 dividido por el número total de artículos publicados en los años t-1 y t-2 (ver, Stigler, S.M. (1994)). Basándose en dichos coeficientes se establecen ordenaciones entre las revistas, de forma que se considera de mayor **prestigio** aquellas publicaciones efectuadas en revistas con un mayor FI.

Es evidente la necesidad de disponer de indicadores para evaluar la producción científica, puesto que las autoridades de política científica necesitan distinguir entre la ciencia que debe ser financiada y la que no sobre la base de que se alcance o no la calidad necesaria. El FI es también el índice de calidad utilizado para llevar a cabo los citados repartos. Es éste un modo muy útil y cómodo para los evaluadores de la ciencia, pues ahorra análisis más profundos y basta con mirar en las bases de datos los correspondientes factores de impacto. El FI también es uno de los baremos utilizados a la hora de la financiación de proyectos de investigación, de la selección de candidatos para plazas de profesor universitario, etc.

Si definimos la variable T que mide la antigüedad de las citas recibidas por una revista, el FI tiene en cuenta tan sólo las frecuencias de dos de los valores de esta variable (relativizando por el número de artículos publicados). Vemos así que el FI es una medida de la “visibilidad a corto plazo” que alcanzan las publicaciones. Desde nuestro punto de vista, veremos que se trata de una medida con limitaciones para describir los distintos comportamientos que sigue la variable T.

En efecto, hay otra característica a tener en cuenta que es importante. Es bien conocido el proceso de obsolescencia o caída en desuso de la literatura científica. Así, una característica común que se aprecia al analizar la variable T es que a unos primeros años de aumento del número de citas le sigue un proceso de caída que culmina con la desaparición de citas tras un periodo más o menos largo de tiempo (ver Gupta, B.M. (1997,1998), Moed, H.F. et al. (1999), Glänzel, W. and Shoepflin, U. (1994,1995), etc.).

Entre los argumentos que se han esgrimido para explicar el fenómeno de la obsolescencia, hemos recogidos los puntos siguientes: i) Como consecuencia del crecimiento del número de trabajos publicados, hay más literatura reciente que antigua disponible para ser citada, ii) El campo de estudio ha adquirido su interés recientemente, iii) Muchos trabajos antiguos que contienen información válida y útil tienden a ser incorporados en trabajos posteriores o los resultados en ellos expresados son generalizados en otros trabajos más recientes.

De acuerdo con Price, D. (1986), la distribución de referencias en los artículos de una especialidad (o las citas recibidas) muestra la naturaleza de la investigación que se lleva a cabo en la misma. De forma genérica, una tasa alta de obsolescencia (es decir, un período de “desaparición” corto de los trabajos) en un determinado campo, indicaría que los participantes conocen bien un planteamiento general del problema así como un conjunto de subproblemas en el que se está interesado en un momento determinado. Tan pronto como se encuentra la solución a este subconjunto de situaciones particulares, se produce un movimiento a interesarse por otro bloque de



subproblemas distinto. Sería este el caso de los trabajos más aplicados, en los que se presupone que los procedimientos teóricos necesarios ya son conocidos y se van aplicando a diversas situaciones, con lo que el “avance” en la obtención de resultados es más rápido. Por el contrario, una tasa baja de decaimiento indicaría que los participantes tratan de resolver problemas más generales, tal vez no claramente especificados y utilizando enfoques o herramientas distintas, lo que hace que los avances significativos se produzcan con mayor lentitud. Sería éste, de forma genérica, el caso de los trabajos más teóricos.

De todo lo dicho con anterioridad, la conclusión fundamental es que cualquier estudio encaminado a la valoración y clasificación de las revistas ha de tener en cuenta tanto la fase inicial de aumento de citas como el período de decaimiento, ya que es la “historia completa” la que nos muestra las características de los trabajos publicados.

Si observamos el fenómeno desde el punto de vista del investigador, sabemos que aquellos científicos que son más frecuentemente citados por el resto de colegas son los que gozan de un mayor prestigio. Es razonable pensar que el único objetivo de una persona que publica un trabajo no sea sólo ser citado de forma inmediata; el hecho de comprobar que su aportación a la comunidad sigue siendo útil transcurrido un largo período de tiempo también será motivo de satisfacción y de reconocimiento por parte de sus colegas. Podríamos preguntarnos entonces: ¿qué es preferible, publicar un artículo que tendrá una alta visibilidad inmediata y que “será olvidado” a los 10 años o conseguir un menor impacto a corto plazo pero seguir siendo citado a los 20 años de la publicación?. Además, desde la perspectiva del número de veces que son citados los artículos, un bajo factor de impacto acompañado de una vida larga puede suponer en términos absolutos un mayor número de citas que un alto factor de impacto acompañado de un período de vida corto.

Hemos de tener en cuenta también que, como ya se ha mencionado, el análisis bibliométrico a través de las citas conlleva el problema inherente de que el hecho de que un artículo sea muy citado no significa necesariamente que el trabajo sea de

calidad. Se puede citar un artículo en términos negativos, es decir, para mostrar desacuerdo con los resultados establecidos. En Callon, M. et al (1995), se dice textualmente: “*Contar el número de citas recibidas por un artículo consiste, pues, en valorar a la vez tanto su visibilidad como su impacto, pero en absoluto en medir su calidad*”<sup>8</sup>. Parece lógico pensar que este problema puede suavizarse en buena parte a medida que las citas son de mayor antigüedad y que si un artículo perdura en el tiempo siendo citado esto es una muestra de su calidad. Esta idea refuerza la necesidad de tener en cuenta la fase de descenso de la distribución de T.

Una consecuencia negativa del uso generalizado del F.I. es que el hecho de evaluar sólo una parte de la distribución está provocando cambios importantes en la orientación de la investigación, ya que se está primando un tipo de actividad científica que provoca muchas citas pero durante un período corto de tiempo, es decir, tanto los autores como las propias revistas buscan publicar trabajos que por sus características sean más propensos a ser citados a corto plazo, lo cual no tiene por qué significar que sean de mayor calidad. Un dato interesante es que en las bases de datos del ISI, sólo se consideran citas a artículos de menos de 10 años de antigüedad; el resto aparecen agregadas, como si carecieran de importancia. Pues bien, en algunas revistas, las citas de más de 10 años de antigüedad pueden suponer hasta un 70 ó 75% del total.

Como dijimos anteriormente, la variable T suele mostrar un comportamiento común de una fase creciente seguida por otra decreciente, llevando a distribuciones campaniformes y asimétricas a la derecha. Podemos distinguir inmediatamente grandes diferencias en función de las temáticas o áreas de conocimiento (Callon, M. et al. (1995)). Así, por ejemplo, se puede apreciar que, en general, las revistas de medicina tienen FI superiores a las de matemáticas, pero por el contrario la tasa de decaimiento es muy superior en las primeras, llevando a que los artículos caen en desuso mucho antes. Este fenómeno de comportamiento diferente según las temáticas es bien conocido, por lo que también suele ser habitual asignar a cada revista factores de impacto relativos dentro de su temática. No obstante, seguimos pensando que, aún

---

<sup>8</sup> Callon et al. (1995), pág. 47

centrándonos en una temática concreta, el FI es una medida con limitaciones para representar a la distribución de citas correspondiente, ya que se ocupa sólo de la fase de crecimiento de la misma y que por tanto lo mínimo que hemos de exigir a cualquier índice de calidad es que tenga en cuenta también de alguna forma la tasa de decaimiento de las citas.

### **5.3 Datos utilizados y procedimiento seguido.**

En el presente trabajo se han usado los datos elaborados por el ISI (“Institute Scientific Information”). En las bases “Science Journal Citation Reports” y “Social Science Journal Citation Reports” se ofrece información del número de citas recibidas por las revistas y de la antigüedad de las mismas, si bien en ella sólo se contempla el número de citas de antigüedad inferior o igual a 9 años, apareciendo todas las citas de mayor antigüedad agregadas en un solo dato. Así, la utilización de estos datos implica la búsqueda de modelos pero en presencia de censura, ya que de ciertas observaciones sólo sabremos que son mayores que 10. Estos datos se elaboran teniendo en cuenta las referencias de 5550 revistas.

En las bases “Science Citation Index” y “Social Science Citation Index” se ofrece información detallada, artículo por artículo, de un número de revistas más reducido (3300) que el que se tiene en cuenta en la base de datos anterior, incluyéndose las referencias que aparecen en cada trabajo. A través de un proceso de búsqueda automatizado en las referencias, podemos elaborar la distribución completa de la antigüedad de las citas de una revista. Concretando un poco más, el procedimiento ha sido el siguiente. Consideremos, por ejemplo, la base de datos correspondiente al año 1999. En las referencias de este conjunto fuente hemos buscado todas las citas correspondientes a una determinada revista A. Posteriormente, sobre este conjunto de resultados, hemos hecho un proceso de recuento para obtener cuántas de estas citas correspondían a artículos publicados en 1999, 1998, 1997, etc. (que hemos hecho corresponder, como indicaremos más adelante, a antigüedades de

0.5, 1.5, 2.5, etc.). De esta forma se obtiene la distribución de la edad de las citas hechas a la revista A en el año 1999. El proceso se ha repetido para las distintas revistas consideradas y para distintos años.

De esta forma, las opciones que tenemos son considerar datos censurados pero teniendo en cuenta un mayor número de revistas o datos sin censurar procedentes de un conjunto de fuentes más reducido. Indiquemos que en ambos casos, estamos utilizando un enfoque sincrónico en la elaboración de los datos.

Como se verá posteriormente, un modelo que se adapta bastante bien a nuestros datos es el modelo Weibull. En el presente trabajo, se tendrán en cuenta las dos posibilidades, de forma que las conclusiones se establecerán tanto sobre la base de los datos censurados como sobre la base de los no censurados. Dado el elevado porcentaje de datos censurados en algunas revistas, las conclusiones más fiables van a ser las obtenidas sin censura, a pesar de que la base de datos correspondiente es menos completa. No obstante, hemos preferido proporcionar también la metodología necesaria para obtener resultados en el caso de que la única posibilidad fuese disponer de datos con censura que son los que proporciona directamente el ISI.

Para la obtención de la distribución de la antigüedad de las citas de una revista dada, hemos agregado los datos correspondientes de las bases de datos de 1999, 1998, 1997 y 1996. Así, en cierta forma, estamos introduciendo un supuesto de estacionariedad en el comportamiento de la variable en estos cuatro años. Al ser un período de tiempo tan corto, el supuesto es bastante razonable. De hecho, la simple observación de los histogramas correspondientes a cada año nos muestra un comportamiento estable. Con esta agregación se consigue un mayor número de datos lo que conlleva una notable suavización de las irregularidades y una mayor precisión en las estimaciones.

Aunque la variable  $T$  es de naturaleza continua ( y por ello para la modelización hemos usado modelos continuos), a efectos prácticos el valor otorgado

de antigüedad a una cita en el año  $t$  correspondiente a un artículo publicado en el año  $s$  ha sido  $(t-s)+0,5$ , puesto que lo que sabemos es que la antigüedad de la cita se sitúa en el intervalo  $[t-s, t-s+1]$ . Conviene señalar que en la literatura existente suele ser más habitual tomar  $t-s+1$  o bien  $t-s$  (aunque esta última opción conlleva la dificultad de que aparezca el valor 0), si bien a efecto comparativo las opciones son equivalentes pues la diferencia entre las mismas se reduce a efectuar cambios de origen.

Los datos obtenidos (tanto con censura como sin censura) así como los correspondientes histogramas para el caso de ausencia de censura se ofrecen en el apéndice A1.

Como ya hemos indicado, el modelo que hemos utilizado por adaptarse bastante bien a los datos obtenidos ha sido el Weibull. En la literatura de biliometría se ha usado con más frecuencia el modelo Log-Normal (Gupta, B.M. (1997), Matriccioni, E. (1994), Egghe, L. And Rao, I.K. (1992), etc.). Este modelo es más fácilmente manejable en ausencia de censura, aunque en presencia de datos censurados el modelo Weibull resulta más versátil ya que disponemos de una expresión explícita para la función de supervivencia. También ha sido utilizado el modelo exponencial dada sus simplicidad (Rao, I.K. and Meera, B.M. (1991), Brookes, B.C. (1974), Ruiz, R. y Jiménez, E. (1996), etc.). Este modelo tiene el inconveniente de no adaptarse bien en la fase inicial de ascenso. En nuestros datos, los análisis gráficos efectuados (no incluidos en el trabajo) indicaban que el modelo exponencial no era adecuado para muchas de las revistas estudiadas. Otra alternativa que aparece en Gupta, B.M. (1997) es usar dos modelos exponenciales distintos, uno para cada fase de comportamiento.

La elección del modelo la hemos basado en el estudio gráfico de la función de supervivencia a lo largo del tiempo. En el modelo Weibull la función de supervivencia viene dada por  $S(t) = \exp\{-(\lambda t)^\beta\}$ . De este forma tenemos

$$\text{Ln}[S(t)] = -(\lambda t)^\beta \Rightarrow \text{Ln}[\text{Ln}[S(t)]] = \beta \text{Ln}[\lambda] + \beta \text{Ln}[t]$$

Ahora, representamos  $\hat{S}(t)$  (que será una estimación no paramétrica de  $S(t)$ ) frente a  $\text{Ln}[t]$  y comprobamos si los puntos se ajustan a una recta. Este procedimiento puede verse en Nelson, W.B. (1972) y en Lawless, J.F. (1982). La diferencia entre los casos con y sin censura estaría en la obtención de  $\hat{S}(t)$ . En el caso de ausencia de censura hemos usados la estimación usual  $\hat{S}(t) = n_t/n$  (donde  $n_t$  es el número de observaciones que son mayores o iguales que  $t$ ) y en presencia de censura hemos utilizado la estimación usual de Kaplan-Meier (ver, por ejemplo, Lawless, J.F., 1982). Las gráficas correspondientes a cada revista pueden verse en el apéndice A2.

Para la obtención de resultados hemos aplicado los métodos de simulación descritos con anterioridad, obteniendo una muestra de tamaño 1000 de la distribución conjunta  $(u, b | x_i)$  trabajando a partir del modelo de valor extremo (es decir, con los logaritmos de los datos originales) tras realizar el cambio de origen ya descrito correspondiente en cada caso. A partir de esta muestra, podemos obtener una muestra de cualquier otra característica del modelo, interesándonos principalmente en los percentiles. El percentil de orden  $p$ ,  $t_p$ , se puede poner como función de los parámetros  $u$  y  $b$  como sigue.

En el modelo de valor extremo tenemos

$$F(x) = 1 - \exp\left\{-e^{(x-u)/b}\right\}$$

Así, siendo  $\tilde{x}_p$  el percentil de orden  $p$  de este modelo (con el cambio de origen introducido), tenemos:

$$F(\tilde{x}_p) = 1 - \exp\left\{-e^{(\tilde{x}_p - u)/b}\right\} = p \Rightarrow \tilde{x}_p = b \operatorname{Ln}[-\operatorname{Ln}[1 - p]] + u$$

Deshaciendo el cambio de origen, con  $x_p = \tilde{x}_p + c$ , es decir,

$$x_p = b \operatorname{Ln}[-\operatorname{Ln}[1 - p]] + u + c$$

con lo que el percentil correspondiente en la distribución Weibull correspondiente a los datos originales será

$$t_p = \exp\{b \operatorname{Ln}[-\operatorname{Ln}[1 - p]] + u + c\}$$

Existen diversas características relacionadas con el tiempo que tardan los trabajos en dejar de ser citados. Hemos pensado que la medida más intuitiva que podemos elegir es un percentil alto de la distribución y concretamente hemos tomado como referencia  $P_{90}$ . En efecto, esta característica indicaría el período de tiempo durante el cual los trabajos de una revista recibirían el 90% de sus citas, momento a partir del cual consideraremos que los trabajos “caen en el olvido”. Evidentemente, la elección del 90% puede ser discutible e igualmente podríamos pensar en elegir percentiles más bajos como el 75, 80, 85%, o percentiles más altos como 95 ó 99%.

El objetivo fundamental del análisis será la comparación entre el período inicial de la distribución a través del FI con la velocidad de la fase descendente de la misma a través del  $P_{90}$ , poniendo de manifiesto que la evaluación de las revistas exclusivamente basándose en el FI deja de lado cierta información y que es necesario tener en cuenta, como mínimo, un indicador de cada fase de la distribución, la parte de crecimiento y de la caída. Una idea hasta cierto punto similar puede encontrarse en Moed, H.F. et al. (1999), pues en dicho trabajo también se plantea una clasificación de las revistas teniendo en cuenta ambas fases de la distribución de T, si bien en él los

criterios son distintos y la conclusión a la que se llega es de tipo cualitativo, pues se divide a las revistas en distintos grupos a la vista de los resultados.

En primer lugar, para poner de manifiesto la disparidad de comportamientos en función de las temáticas hemos analizado a título meramente ilustrativo las revistas “Journal of Applied Probability” (J-APPL-PROBAB), “Bird Study” (BIRD-STUDY) y “Medicina Clínica” (MED-CLIN-BARCELONA).

No obstante, como ya se ha mencionado, dentro de las temáticas también encontramos comportamientos muy diversos. Nuestro propósito ha sido analizar las revistas del ámbito de Economía Aplicada. Para ello, hemos hecho una selección de 12 revistas de dicha temática y que aparecen en las dos bases de datos del ISI mencionadas. Las revistas analizadas han sido:

<b>Nº</b>	<b>Nombre de la revista</b>	<b>Abreviatura</b>
1	APPLIED ECONOMICS	APPL-ECON
2	ECONOMETRIC THEORY	ECONOMET-THEOR
3	ECONOMETRICA	ECONOMETRICA
4	INSURANCE MATHEMATICS&ECONOMICS	INSUR-MATH-ECON
5	JOURNAL OF APPLIED ECONOMETRICS	J-APPL-ECONOM
6	JOURNAL OF APPLIED PROBABILITY	J-APPL-PROBAB
7	JOURNAL OF APPLIED STATISTICS	J-APPL-STAT
8	JOURNAL OF ECONOMETRICS	J-ECONOMETRICS
9	JOURNAL OF MATHEMATICAL ECONOMICS	J-MATH-ECON
10	OXFORD BULLETIN OF ECONOMICS AND STATISTICS	OXFOR-B-ECON-STAT
11	REVIEW OF ECONOMICS AND STATISTICS	REV-ECON-STAT
12	TECHNOMETRICS	TECHNOMETRICS



## 5.4 Resultados y conclusiones.

### 5.4.1 Tres temáticas distintas.

Como acabamos de mencionar, hemos seleccionado tres revistas a título ilustrativo para poner de manifiesto las importantes diferencias de comportamiento que se dan entre temáticas distintas. En este apartado el análisis será muy breve, ya que nuestro objetivo es el estudio de las revistas del ámbito de Economía Aplicada.

Recogemos en la tabla 5.I el FI de cada revista en el año 1999 así como la mediana y los percentiles de orden 75 y 90, todos estimados sobre la base de un modelo Weibull.

FI99	REVISTA	DATOS SIN CENSURA			DATOS CON CENSURA		
		Me	Q3	P90	Me	Q3	P90
<b>0,592</b>	BIRD-STUDY	11,0191	17,9537	<b>25,6672</b>	11,1608	18,6251	27,1578
<b>0,435</b>	J-APPL-PROBAB	12,5901	19,2505	<b>26,2692</b>	13,4511	21,7343	30,9075
<b>0,811</b>	MED-CLIN-BARCEL.	4,4948	7,1941	<b>10,1510</b>	4,8138	7,8117	11,1363

Tabla 5.I: Resultados de tres temáticas distintas

En este apartado, sólo pondremos de manifiesto cómo la revista correspondiente al ámbito de medicina posee un factor de impacto mayor que las otras dos, pero sin embargo la tasa de decaimiento en el número de citas a medida que aumenta la antigüedad es muy superior, lo que lleva a que en un período de aproximadamente 10 años los trabajos prácticamente dejan de ser citados (como podemos apreciar en el valor del percentil 90, que es 10,1510 años) mientras que en las revistas del ámbito de biología y de probabilidad este período se sitúa, respectivamente, en 25,6672 y 26,2692 años.

Así, en cualquier clasificación sobre la base del factor de impacto exclusivamente, las revistas de medicina saldrían beneficiadas, lo que no parece

razonable puesto que la aportación de los trabajos, si bien muy abundante en un principio, es bastante efímera, mientras que en los otros casos la aportación es mucho más duradera.

Esta idea puede simplemente ser observada en base a los histogramas de cada revista para la variable antigüedad de las citas (que aparecen en el apéndice A1). Puede apreciarse cómo en el caso de MED-CLIN-BARCELONA el descenso en el número de citas es muchísimo más elevado que en las otras dos revistas.

#### **5.4.2 Doce revistas del ámbito de Economía Aplicada.**

En la tabla 5.II, recogemos los FI de cada revista en el año 1999 así como los percentiles de orden 50, 75 y 90 estimados en base al modelo Weibull usando los datos con y sin censura (como hemos indicado, los datos junto con los correspondientes histogramas pueden verse en el apéndice A1).

FI99	REVISTA	DATOS SIN CENSURA			DATOS CON CENSURA		
		Me	Q3	P90	Me	Q3	P90
<b>0,235</b>	APPL-ECON	6,0423	9,0191	<b>12,0923</b>	6,9132	9,9170	12,9201
<b>2,206</b>	ECONOMETRICA	14,5620	22,3780	<b>30,6491</b>	15,1413	22,0763	29,0972
<b>0,404</b>	ECONOMET-THEORY	5,9404	8,7748	<b>11,6753</b>	6,1157	8,7730	11,4312
<b>0,103</b>	INSUR-MATH-ECON	5,3431	8,0846	<b>10,9510</b>	6,0850	9,8800	14,1661
<b>0,794</b>	J-APPL-ECONOM	5,4843	7,3178	<b>9,0382</b>	6,2753	8,8544	11,3980
<b>0,435</b>	J-APPL-PROBAB	12,5901	19,2505	<b>26,2692</b>	13,4511	21,7343	30,9075
<b>0,257</b>	J-APPL-STAT	4,9473	7,1367	<b>9,3335</b>	5,4474	8,1125	10,8765
<b>0,829</b>	J-ECONOMETRICS	9,3227	13,7294	<b>18,2270</b>	9,3119	14,0177	18,9145
<b>0,320</b>	J-MATH-ECON	10,9385	16,1823	<b>21,5557</b>	12,0268	19,4896	27,8081
<b>0,508</b>	OXFORD-B-ECON-STAT	7,6830	11,0271	<b>14,3664</b>	7,5616	10,3689	13,0692
<b>0,973</b>	REV-ECO-STAT	12,9510	21,3279	<b>30,7286</b>	12,3630	19,2315	26,5843
<b>1,250</b>	TECHNOMETRICS	17,8798	25,7315	<b>33,5896</b>	18,2249	28,4272	39,3862

**Tabla 5.II:** Resultados para las revistas del ámbito de Economía Aplicada.

Pretendemos establecer una clasificación de las revistas que tenga en cuenta por una parte la incidencia a corto plazo de los trabajos que publica cuando se mide con el FI y por otra parte la perdurabilidad de los mismos en el tiempo medida a través del percentil de orden 90, estimado sobre la base de la hipótesis de que la antigüedad de las citas se ajusta a un modelo Weibull. Hemos utilizado los datos sin censura, pues como se explicó con anterioridad creemos que aún procediendo de un menor número de revistas, la ausencia de censura hace que los resultados sean más fiables.

En primer lugar, hemos efectuado un simple gráfico de dispersión en el que representamos para cada revista un punto cuya coordenada en el eje X es su FI del año 1999 y cuya coordenada Y es su  $P_{90}$ . Las líneas en vertical y en horizontal representan los valores medios de los factores de impacto y de los percentiles, respectivamente, para el grupo de revistas analizadas. El rótulo de cada punto representa a las distintas revistas sobre la base de la numeración establecida junto al gráfico.

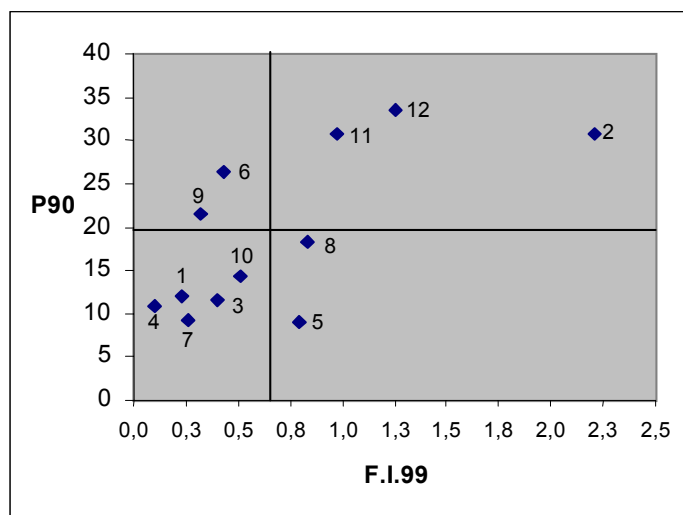


Figura 5.1: F.I. y  $P_{90}$  para revistas del ámbito de Economía Aplicada.

Vemos en este gráfico que las revistas etiquetadas con 2, 11 y 12 (es decir, *ECONOMETRICA*, *REV-ECON-STAT* y *TECHNOMETRICS*) destacan dentro del grupo tanto en un criterio como en el otro. Análogamente, las revistas 1, 3, 4, 7 y 10

(APPL-ECONOM, ECONOMET-THEOR, INSUR-MATH-ECON, J-APPL-STAT Y OXFORD-B-ECON-STAT) están por debajo de la media en ambos aspectos.

Podemos observar también que las revistas 5 y 8 (J-APPL-ECONOM y J-ECONOMETRICS) se benefician del criterio del FI en las clasificaciones, puesto que si bien destacan en este aspecto con respecto a la media, están por debajo de ella en el criterio del P<sub>90</sub>, o lo que es lo mismo, en la longitud del período de vigencia de sus publicaciones. Análogamente, las revistas 6 y 7 (J-APPL-PROBAB y J-APPL-STAT) salen perjudicadas en las clasificaciones sobre la base del FI, pues están por debajo de la media con respecto a este criterio, pero por encima en cuanto al de durabilidad.

Vamos a continuación a establecer una comparación entre la clasificación a través del FI y la que se obtendría sobre la base del criterio del P<sub>90</sub>. Las clasificaciones son las que se detallan en la tabla 5.III.a.

CLASIFICACIÓN POR F.I.			CLASIFICACIÓN POR P <sub>90</sub>		
Orden	Revista	F.I.	Orden	Revista	P <sub>90</sub> .
1	ECONOMETRICA	2,206	1	TECHNOMETRICS	33,5896
2	TECHNOMETRICS	1,250	2	REV-ECO-STAT	30,7286
3	REV-ECO-STAT	0,973	3	ECONOMETRICA	30,6491
4	J-ECONOMETRICS	0,829	4	J-APPL-PROBAB	26,2692
5	J-APPL-ECONOM	0,794	5	J-MATH-ECON	21,5557
6	OXFORD-B-ECON-STAT	0,508	6	J-ECONOMETRICS	18,2270
7	J-APPL-PROBAB	0,435	7	OXFORD-B-ECON-STAT	14,3664
8	ECONOMET-THEORY	0,404	8	APPL-ECON	12,0923
9	J-MATH-ECON	0,320	9	ECONOMET-THEORY	11,6753
10	J-APPL-STAT	0,257	10	INSUR-MATH-ECON	10,9510
11	APPL-ECON	0,235	11	J-APPL-STAT	9,3335
12	INSUR-MATH-ECON	0,103	12	J-APPL-ECONOM	9,0382

**Tabla 5.III.a:** Clasificación en base al F.I. y en base al P<sub>90</sub>.

En la tabla 5.III.b hemos combinado ambas clasificaciones para apreciar mejor las diferencias existentes en función del criterio establecido.

REVISTA	Orden Por FI	Orden Por P <sub>90</sub>	Difer.
APPL-ECON	11	8	3
ECONOMETRICA	1	3	-2
ECONOMET-THEORY	8	9	-1
INSUR-MATH-ECON	12	10	2
J-APPL-ECONOM	5	12	-7
J-APPL-PROBAB	7	4	3
J-APPL-STAT	10	11	-1
J-ECONOMETRICS	4	6	-2
J-MATH-ECON	9	5	4
OXFORD-B-ECON-STAT	6	7	-1
REV-ECO-STAT	3	2	1
TECHNOMETRICS	2	1	1

**Tabla 5.III.b:** Diferencia entre las clasificaciones en base al F.I. y al P<sub>90</sub>.

Observemos que en un grupo de doce revistas, atendiendo a los dos criterios distintos encontramos diferencias de hasta siete lugares entre ambas clasificaciones, que es el caso de la revista J-APPL-ECONOM; esta revista, por tanto, actualmente resulta muy beneficiada por el hecho de atender exclusivamente al FI como índice de calidad de las revistas. En el otro extremo estaría J-MATH-ECON que es la novena sobre la base del FI pero sin embargo es de las revistas con mayor perdurabilidad en sus publicaciones.

Evidentemente, el paso siguiente sería preguntarse de qué forma podemos tener en cuenta ambos criterios para elaborar un único índice, teniendo en cuenta que queremos contemplar magnitudes tan dispares como un FI definido a través del cociente del total de citas entre el total de publicaciones y un percentil de la

distribución de la antigüedad de las citas. Podemos encontrar múltiples soluciones “razonables” al problema planteado, posiblemente, cada una de ellas con ventajas e inconvenientes.

Pero antes de buscar esa solución se hace necesario puntualizar que al igual que vamos a pretender establecer una clasificación sobre la base de los criterios de impacto y de durabilidad de las publicaciones, también debemos plantearnos el tener en cuenta más factores y la necesidad de decidir qué criterios resultan relevantes y han de ser tenidos en cuenta y cuáles no. No obstante, nuestro propósito principal es poner de manifiesto nuestra opinión de que el período de vigencia de las publicaciones es un criterio imprescindible y que el FI, si bien es una característica a tener en cuenta, no es suficiente en sí mismo para valorar a las revistas.

No creemos que sea posible establecer el criterio “perfecto” para una tarea tan complicada como intentar medir la “calidad” de las revistas científicas. Sin embargo, pensamos que la clasificación que propondremos más adelante es más razonable que las actuales clasificaciones basadas en el FI, ya que tiene en cuenta otra característica que consideramos muy importante en la literatura científica como es el período de vigencia de los trabajos.

Volviendo al problema de la agregación de las características FI y  $P_{90}$ , es evidente que se hace necesaria una relativización de los valores a través de algún procedimiento, ya que no podemos pensar en agregar o promediar dos magnitudes tan dispares. Como ya se ha mencionado, podemos pensar en varias posibilidades. La idea consistiría en comparar los valores individuales de cada revista con los del conjunto en cada uno de los criterios. En las clasificaciones sobre la base del FI se ha utilizado el llamado Factor de Impacto Relativo, consistente en dividir el FI de una revista por el promedio de la temática a la que pertenezca. Una posibilidad, podría ser obtener este cociente para el FI y para el  $P_{90}$ , pasando posteriormente a hacer un promedio de ambos. No obstante, por tratarse en este caso de dos características de promedios tan distintos, el procedimiento que vamos a elegir es el obtener los valores tipificados

tanto de los factores de impacto como de los percentiles dentro del grupo objeto de análisis. Los valores tipificados de ambas características se ofrecen en la tabla 5.IV para las doce revistas objeto de estudio.

REVISTA	FI <sub>99</sub> (tip.)	P <sub>90</sub> (tip.)
APPL-ECON	-0,8160	-0,7907
ECONOMETRICA	2,6968	1,3213
ECONOMET-THEORY	-0,5148	-0,8382
INSUR-MATH-ECON	-1,0512	-0,9206
J-APPL-ECONOM	0,1803	-1,1383
J-APPL-PROBAB	-0,4595	0,8228
J-APPL-STAT	-0,7768	-1,1047
J-ECONOMETRICS	0,2427	-0,0925
J-MATH-ECON	-0,6645	0,2864
OXFORD-B-ECON-STAT	-0,3294	-0,5319
REV-ECO-STAT	0,4993	1,3304
TECHNOMETRICS	0,9930	1,6560

Tabla 5.IV: Valores de FI y P<sub>90</sub> tipificados.

Estos valores tipificados indican de forma mucho más homogénea cuál es la situación de cada revista dentro del grupo sobre la base de cada uno de los criterios de clasificación, pudiendo además ser agregados o promediados.

Un promedio entre estos valores tipificados establecería la clasificación dentro del grupo sobre la base de los criterios mencionados de impacto y durabilidad. Surge ahora la posibilidad de asignar ponderaciones a cada una de las características, ponderaciones que incluso podrían variar en función de las temáticas, del objetivo concreto de la clasificación que se está efectuando, etc. Nosotros hemos asignado igual ponderación a los dos criterios, obteniendo la clasificación que se ofrece a continuación en la tabla 5.V.

REVISTA	FI <sub>99</sub> (tip.)	P <sub>90</sub> (tip.)	Media
ECONOMETRICA	2,6968	1,3213	<b>2,0091</b>
TECHNOMETRICS	0,9930	1,6560	<b>1,3245</b>
REV-ECO-STAT	0,4993	1,3304	<b>0,9148</b>
J-APPL-PROBAB	-0,4595	0,8228	<b>0,1817</b>
J-ECONOMETRICS	0,2427	-0,0925	<b>0,0751</b>
J-MATH-ECON	-0,6645	0,2864	<b>-0,1891</b>
OXFORD-B-ECON-STAT	-0,3294	-0,5319	<b>-0,4307</b>
J-APPL-ECONOM	0,1803	-1,1383	<b>-0,4790</b>
ECONOMET-THEORY	-0,5148	-0,8382	<b>-0,6765</b>
APPL-ECON	-0,8160	-0,7907	<b>-0,8033</b>
J-APPL-STAT	-0,7768	-1,1047	<b>-0,9407</b>
INSUR-MATH-ECON	-1,0512	-0,9206	<b>-0,9859</b>

Tabla 5.V: Promedio de FI y P<sub>90</sub>.

Las diferencias existentes entre una clasificación en base exclusivamente al FI y la que acabamos de establecer pueden apreciarse en la tabla 5.VI.

REVISTA	Orden Por FI	Orden Por FI y P <sub>90</sub>	Difer.
APPL-ECON	11	10	1
ECONOMETRICA	1	1	0
ECONOMET-THEORY	8	9	-1
INSUR-MATH-ECON	12	12	0
J-APPL-ECONOM	5	8	-3
J-APPL-PROBAB	7	4	3
J-APPL-STAT	10	11	-1
J-ECONOMETRICS	4	5	-1
J-MATH-ECON	9	6	3
OXFORD-B-ECON-STAT	6	7	-1
REV-ECO-STAT	3	3	0
TECHNOMETRICS	2	2	0

Tabla 5.VI: Diferencia entre la clasificación en base al FI y la clasificación bicriterio.



Puede verse cómo las diferencias son de hasta tres posiciones para un grupo de tan sólo doce componentes, siendo estos los casos de las revistas J-APPL-ECONOM, J-APPL-PROBAB y J-MATH-ECON, si bien la primera de ellas se beneficia del actual criterio del FI mientras que para las otras dos el criterio es negativo.

Como ya hemos puesto de manifiesto a lo largo del trabajo, creemos que, si bien este método de clasificación podría ser mejorado, es más justo que el actual sistema basado sólo en el FI.



## **CUESTIONES ABIERTAS Y CONCLUSIONES GENERALES.**

En esta sección pretendemos resaltar las principales cuestiones que quedan abiertas en este trabajo (y que por tanto constituyen posibles líneas de investigación futuras) así como algunas conclusiones generales a las que hemos podido llegar durante su realización. Par ello, distinguiremos entre el bloque metodológico y el práctico.

En el primer bloque teórico, destaquemos en primer lugar que sería deseable obtener resultados generales sobre el comportamiento frecuencialista (al menos asintótico) de los intervalos bayesianos, sobre todo en el caso multiparamétrico. Este tipo de resultados entraña gran dificultad, dado que el abandono de las condiciones de regularidad habituales hace que los desarrollos asintóticos necesarios no sean válidos en muchos casos. Algunos resultados de comportamiento asintótico en modelos no regulares (aunque desde otras perspectivas) pueden verse en Akahira, M. and Takeuchi, K. (1995) y en Ghosal, S. and Samanta, T. (1997), aunque referidos al caso uniparamétrico. Es de destacar como conclusión importante que en el caso de localización – escala, la distribución a priori propuesta en el presente trabajo conduce a intervalos bayesianos que coinciden numéricamente de forma exacta con los intervalos obtenidos al condicionar al estadístico auxiliar. Esta propiedad la hemos observado en todos los casos en los que no existe un estadístico suficiente de la misma dimensión que el parámetro lo cual nos lleva a la conclusión de la idoneidad del uso de la inferencia condicional en este tipo de modelos frente al uso de la inferencia incondicional que puede llevar a resultados absurdos (y que además se construye en base a estadísticos no suficientes). Por ello, también es necesario disponer de métodos de obtención de estadísticos auxiliares, ya que no siempre es fácil. Este problema está generando gran cantidad de literatura y existen diversos caminos para obtener estadísticos, al menos, asintóticamente auxiliares (Bardoff-Nielsen, O.E. and Cox, D.R. (1994)).

Enlazando con la idea anterior, obtenemos también la conclusión de que en la actualidad el criterio que más se ha impuesto para decidir la idoneidad de una distribución a priori no informativa es la del estudio del comportamiento frecuencialista de los intervalos con ella obtenidos, siendo también relevante las propiedades de invarianza. Este hecho, nos indica a su vez el giro que ha sufrido el concepto de distribución a priori no informativa hacia la idea de distribución a priori de referencia, imparcial o, según algunos autores, incluso objetiva.

La propuesta que efectuamos para obtener las distribuciones a priori no informativas debe entenderse como un procedimiento global, aplicable a cualquier modelo, ya sea este regular o no. La coincidencia de resultados con la regla de Jeffreys en el caso regular uniparamétrico es entonces una propiedad positiva de nuestro método, ya que en la situación de regularidad el acuerdo sobre la distribución a utilizar es prácticamente unánime. Esta conexión junto con las buenas propiedades observadas en cuanto a la invarianza y al comportamiento frecuencialista de los intervalos bayesianos hace que tengamos una opción razonable para obtener distribuciones a priori y que no nos obliga a “detenernos” ante la presencia de un modelo no regular que puede surgir, por ejemplo, al efectuar un truncamiento en un parámetro desconocido, al necesitar introducir un parámetro de garantía, etc.

En el caso multiparamétrico, quedaría también pendiente la generalización del método de obtención de las distribuciones a priori en cualquier modelo. En efecto, en el trabajo se resuelve el problema a través de la ortogonalización para un conjunto importante de modelos, aunque, como sabemos, la ortogonalización no es siempre posible. No obstante, señalemos que, aun en el caso regular, que ha sido mucho más ampliamente analizado en la literatura, la situación es básicamente la misma, pues aunque se han propuesto algunos métodos (Berger, J. and Bernardo, J.M. (1992)), la única situación en la que prácticamente hay acuerdo a la hora de elegir una distribución a priori no informativa es en el caso ortogonal.

Otro aspecto que queremos destacar es que, como se desprende del capítulo 1, el hecho de que los rangos de las variables en una familia paramétrica dependan del parámetro desconocido, no significa necesariamente que el modelo no sea regular. Evidentemente, el carácter de regularidad dependerá de la definición que se adopte, pero como hemos visto, es posible que aún dependiendo el recorrido del parámetro, puede cumplirse  $E[\partial\ell/\partial\theta]=0$  y que esta es en realidad la condición esencial de regularidad (siempre que  $f(x, \theta)$  sea una función suficientemente “suave”). Vamos a remarcar también que el término “información” debería interpretarse como medida de discriminación o diferencia entre dos distribuciones distintas (es decir para dos valores distintos del parámetro); de esta forma, como dice Pitman, E.J. (1979), el término Información de Fisher no es adecuado, pues hace referencia a una tasa puntual de discriminación y por tanto es una medida de “sensibilidad” del modelo en cada valor del espacio paramétrico. Hemos seguido usando la terminología Información de Fisher dado que ya está ampliamente extendida y el cambio de nombre podría provocar confusiones.

En cuanto al bloque práctico, la principal cuestión que permanecería abierta sería la búsqueda de indicadores adecuados para medir la calidad de las revistas científicas, ya que en base a dichos indicadores se deciden cuestiones muy importantes y el actual sistema del factor de impacto no resulta completamente satisfactorio. La propuesta que hemos hecho debe servirnos fundamentalmente para observar cómo la inclusión de un criterio adicional al del factor de impacto puede provocar cambios importantes en las valoraciones finales. Pensamos que es indispensable tener en cuenta la fase de descenso en la distribución de la antigüedad de las citas, si bien podemos cuestionarnos qué otras características debemos considerar como relevantes y ser por tanto tenidas en cuenta a la hora de elaborar un índice de calidad. También podemos cuestionarnos si sería adecuado utilizar características distintas del percentil 90 para recoger el comportamiento de la fase de “envejecimiento”, como por ejemplo otros percentiles, la entropía de la distribución (Matriccioni, E. (1994)), los propios parámetros estimados de los modelos, un

coeficiente de asimetría, etc. De esta forma, somos conscientes de que en la aplicación práctica lo que hemos hecho fundamentalmente, más que aportar una solución cerrada, es poner de relieve el problema que supone el uso generalizado y exclusivo del factor de impacto para establecer clasificaciones entre las revistas.

En relación a lo anterior, queremos también destacar que unos artículos científicos son más propensos a ser citados a corto plazo que otros por la propia naturaleza de los contenidos de los mismos. Así, de forma genérica, podemos indicar a modo de ejemplos que los trabajos aplicados suelen ser citados más a corto plazo que los trabajos teóricos que necesitan de un mayor tiempo para su asimilación por parte de la comunidad científica; que los artículos de medicina tienden a recibir en poco tiempo todas las citas; que en biología los trabajos taxonómicos suelen recibir pocas citas pero son muy duraderos mientras que un artículo que haga una predicción a corto plazo de la evolución de un determinado ecosistema tenderá a recibir pronto citas corroborando o refutando las predicciones; etc. El uso del factor de impacto puede provocar que tanto los investigadores como las revistas tiendan a publicar aquellos trabajos que por su naturaleza son propensos a recibir citas en un período corto de tiempo, lo cual puede conducir a largo plazo a un sesgo importante hacia este tipo de actividad científica, que en absoluto tiene por qué ser de mayor ni de menor calidad que aquella que necesita de un período más largo para ser aceptada y citada por parte de los investigadores.

Otra conclusión destacable, es la importancia que tienen los métodos de simulación en la resolución, desde el punto de vista bayesiano, de modelos complicados de tratar teóricamente. Dada la vertiginosa velocidad de incremento de las posibilidades de los equipos informáticos, los procedimientos de simulación implementados proporcionan resultados en tiempos razonables prácticamente para cualquier modelo que pretendamos utilizar. Así, la metodología bayesiana ofrece unas posibilidades enormes pues prácticamente basta con elegir la distribución a priori y escribir la verosimilitud del modelo para, implementando correctamente los procesos de simulación, obtener una muestra del tamaño que queramos de la distribución

conjunta a posteriori de los parámetros, a partir de la cual podemos hacer estimaciones de cualquier característica de interés. Puesto que la simulación de distribuciones multidimensionales es complicada, mientras que en el caso unidimensional con los métodos de aceptación – rechazo es relativamente fácil simular cualquier distribución, el algoritmo de Gibbs constituye un eslabón fundamental pues permite precisamente obtener muestras de la distribución conjunta a partir de muestras de las distribuciones condicionadas unidimensionales.

Una cuestión no abordada en este trabajo y que es una de las líneas de investigación que pretendemos abordar en el futuro es el de la resolución de los modelos de duración pero con variables explicativas, es decir, usando modelos de regresión. En este caso, la particularidad más importante es que se suele hacer necesario abandonar las hipótesis de normalidad en las perturbaciones, además de la dificultad adicional que supone siempre la presencia de censura en los datos observados. Señalemos que ya hemos hecho algún trabajo en esta vía dentro del grupo de investigación dirigido por el Dr. Basulto, J. al cual pertenezco resolviendo concretamente el modelo de regresión exponencial, trabajo que fue presentado en las “II Jornadas de Trabajo sobre Metodologías Cualitativas Aplicadas a los Sistemas Socioeconómicos”, celebradas en Sevilla en Abril de 2000.

En conexión con lo anterior, otra cuestión que pretendemos abordar es la de la aplicación de los métodos bayesianos descritos en este trabajo junto con los modelos de regresión en el campo del desempleo, problema encuadrado en el actual proyecto del grupo de investigación al cual pertenezco.





## **APÉNDICE A: DATOS Y GRÁFICAS DE ANTIGÜEDAD DE LAS CITAS**

### **• APÉNDICE A.1**

A continuación ofrecemos los datos de antigüedad de las citas (tanto con censura como sin ella) obtenidos para todas las revistas analizadas así como los correspondientes histogramas (ver páginas 211 y 216).

#### **♦ Tres temáticas distintas.**

##### **1. Datos censurados.**

	<b>1</b>	<b>2</b>	<b>3</b>
<b>0,5</b>	11	21	351
<b>1,5</b>	64	118	889
<b>2,5</b>	95	191	937
<b>3,5</b>	108	241	917
<b>4,5</b>	103	194	817
<b>5,5</b>	88	233	604
<b>6,5</b>	69	195	563
<b>7,5</b>	76	187	443
<b>8,5</b>	62	170	344
<b>9,5</b>	67	185	261
<b>&gt;10</b>	861	2850	996

Fuente: Bases de datos “Social Science Journal Citation Reports” y “Science Journal Citation Reports” del ISI (datos del período 1996-1999 agregados).

**2. Datos sin censura.**

	<b>1</b>	<b>2</b>	<b>3</b>
<b>0,5</b>	8	11	262
<b>1,5</b>	54	68	720
<b>2,5</b>	92	99	691
<b>3,5</b>	90	141	689
<b>4,5</b>	102	127	602
<b>5,5</b>	81	141	458
<b>6,5</b>	71	125	398
<b>7,5</b>	75	96	303
<b>8,5</b>	56	105	231
<b>9,5</b>	60	108	187
<b>10,5</b>	39	94	149
<b>11,5</b>	52	96	115
<b>12,5</b>	55	108	84
<b>13,5</b>	49	93	73
<b>14,5</b>	47	110	52
<b>15,5</b>	45	90	37
<b>16,5</b>	39	86	25
<b>17,5</b>	41	65	10
<b>18,5</b>	38	55	7
<b>19,5</b>	31	65	8
<b>20,5</b>	22	64	3
<b>21,5</b>	27	69	2
<b>22,5</b>	18	79	4
<b>23,5</b>	16	69	1
<b>24,5</b>	28	56	0
<b>25,5</b>	19	55	1
<b>26,5</b>	20	53	0
<b>27,5</b>	20	48	1
<b>28,5</b>	20	36	
<b>29,5</b>	25	38	
<b>30,5</b>	21	42	
<b>31,5</b>	29	26	
<b>32,5</b>	12	31	
<b>33,5</b>	10	22	
<b>34,5</b>	7	10	
<b>35,5</b>	8	8	
<b>36,5</b>	6	1	
<b>37,5</b>	3		
<b>38,5</b>	3		

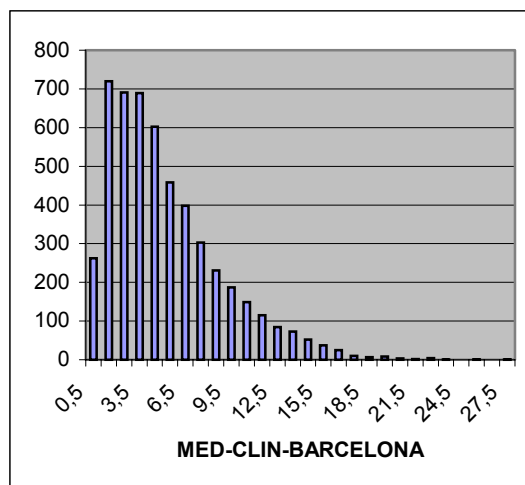
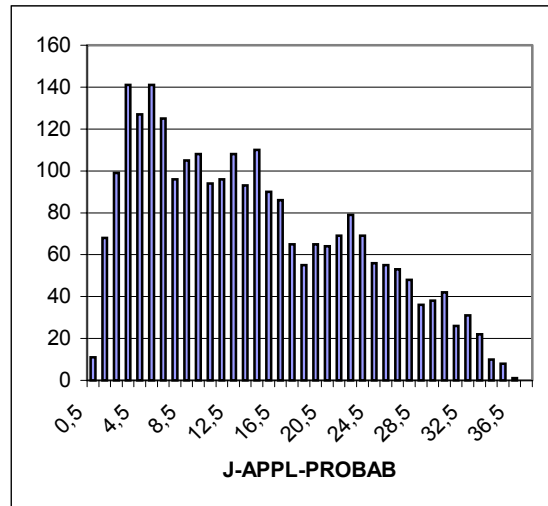
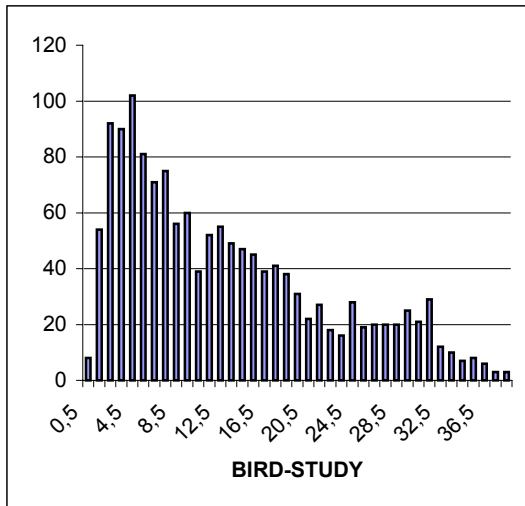
Fuente: Elaboración propia a partir de las bases de datos "Social Science Citation Index" y "Science Citation Index" del ISI (datos del período 1996-1999 agregados).

**Apéndice A: Datos y gráficas de antigüedad de las citas.**

Tanto en una como en otra tabla, en la primera columna están los valores de la variable antigüedad y en la primera fila una etiqueta de cada revista de acuerdo a la tabla que sigue.

Nº	Nombre de la revista	Abreviatura
1	BIRD STUDY	BIRD-STUDY
2	JOURNAL OF APPLIED PROBABILITY	J-APPL-PROBAB
3	MEDICINA CLÍNICA	MED-CLIN-BARCELONA

A continuación, ofrecemos los histogramas correspondientes a cada revista para los datos no censurados.



♦ Doce revistas del ámbito de Economía Aplicada.

1. Datos censurados.

	1	2	3	4	5	6	7	8	9	10	11	12
<b>0,5</b>	7	12	71	4	6	21	7	65	31	9	42	18
<b>1,5</b>	87	65	292	28	91	118	31	236	43	50	230	111
<b>2,5</b>	220	158	579	36	127	191	84	495	74	126	389	173
<b>3,5</b>	227	188	817	43	178	241	90	668	83	153	446	257
<b>4,5</b>	208	189	909	44	207	194	115	747	77	234	445	293
<b>5,5</b>	255	165	1180	22	179	233	45	687	85	210	471	259
<b>6,5</b>	246	154	1217	16	192	195	32	825	102	340	465	313
<b>7,5</b>	225	127	1305	25	161	187	23	680	100	306	460	279
<b>8,5</b>	193	115	1274	12	127	170	33	472	92	191	480	256
<b>9,5</b>	149	102	1572	10	113	185	51	523	72	146	442	278
<b>&gt;10</b>	547	240	22022	78	260	2850	78	4272	987	627	5585	6653

Fuente: Bases de datos “Social Science Journal Citation Reports” y “Science Journal Citation Reports” del ISI (datos del periodo 1996-1999 agregados).

2. Datos sin censura.

	1	2	3	4	5	6	7	8	9	10	11	12
<b>0,5</b>	11	8	46	2	3	11	1	42	25	4	25	16
<b>1,5</b>	119	61	278	20	82	68	21	217	29	48	189	72
<b>2,5</b>	287	125	569	32	138	99	65	461	66	117	377	100
<b>3,5</b>	313	167	774	31	171	141	70	613	78	143	408	140
<b>4,5</b>	264	147	907	37	205	127	46	713	74	242	430	147
<b>5,5</b>	261	143	1175	16	170	141	23	645	77	221	456	157
<b>6,5</b>	253	138	1230	14	200	125	16	773	99	339	456	181
<b>7,5</b>	230	111	1287	20	145	96	23	676	100	304	432	134
<b>8,5</b>	171	102	1240	8	103	105	24	448	94	196	463	141
<b>9,5</b>	120	91	1546	6	89	108	27	483	67	143	426	155
<b>10,5</b>	100	71	1512	8	42	94	19	445	67	130	439	181
<b>11,5</b>	63	55	1383	12	12	96	2	462	74	104	392	154
<b>12,5</b>	55	20	1360	5	5	108	3	428	67	90	355	138
<b>13,5</b>	54	20	999	1	3	93	0	430	66	77	319	128
<b>14,5</b>	43	7	1462	6	1	110	0	327	68	22	327	105
<b>15,5</b>	21	0	1492	4	1	90	1	313	72	31	301	100
<b>16,5</b>	18	0	1522	2		86	2	308	60	20	274	101
<b>17,5</b>	7	0	1827			65		286	41	27	260	168
<b>18,5</b>	8	0	1460			55		180	22	44	258	226
<b>19,5</b>	7	0	1338			65		203	26	16	192	236

Apéndice A: Datos y gráficas de antigüedad de las citas.

20,5	5	0	1020			64		130	33	19	176	216
21,5	1	0	488			69		131	42	13	169	156
22,5	1	0	293			79		177	84	4	116	157
23,5	2	0	302			69		106	63	4	123	154
24,5	3	1	272			56		81	45	3	100	157
25,5	2	0	294			55		66	25	2	118	179
26,5	2	0	323			53		11		3	81	177
27,5	6	0	264			48		0		1	94	120
28,5	4	0	206			36		0		3	83	155
29,5	1	0	175			38		0		1	70	110
30,5		4	160			42		0		1	75	125
31,5		0	101			26		1		1	102	80
32,5		7	140			31				1	80	80
33,5			166			22				1	78	73
34,5			115			10				0	75	103
35,5			158			8				2	34	99
36,5			166			1					27	108
37,5			197								28	99
38,5			204								29	71
39,5			163								46	58
40,5			120								65	24
41,5			113								73	
42,5			92								81	
43,5			107								49	
44,5			123								43	
45,5			158								43	
46,5			175								14	
47,5			110								15	
48,5			88								27	
49,5			76								21	
50,5			30								4	
51,5			25								7	
52,5			28								17	
53,5			34								2	
54,5			22								5	
55,5			20								5	
56,5			3								4	
57,5			9								9	
58,5			14								6	
59,5			31								7	
60,5			26								8	
61,5			35								8	
62,5			25								15	
63,5			26								5	

**Apéndice A: Datos y gráficas de antigüedad de las citas.**

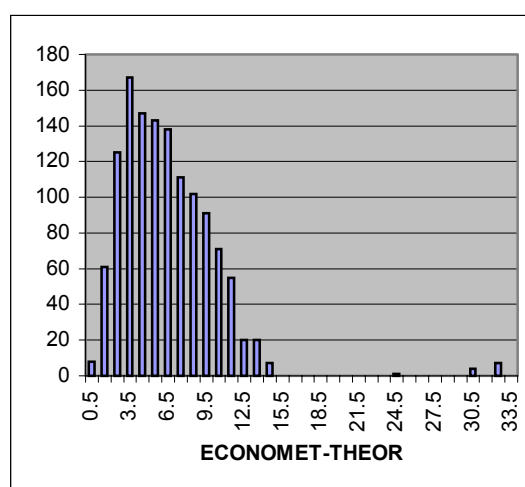
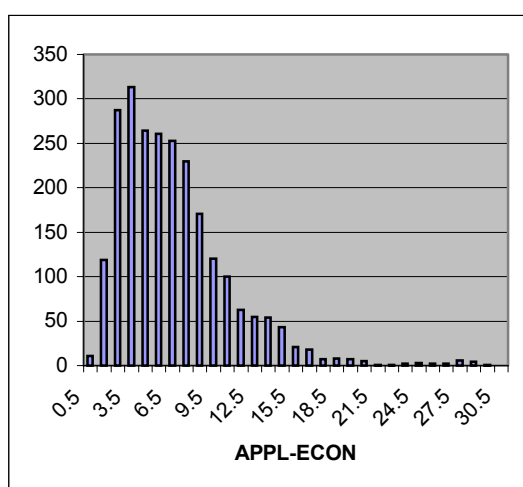
<b>64,5</b>			25							3
<b>65,5</b>			22							3
<b>66,5</b>			16							0
<b>67,5</b>										1

Fuente: Elaboración propia a partir de las bases de datos “Social Science Citation Index” y “Science Citation Index” del ISI (datos del período 1996-1999 agregados).

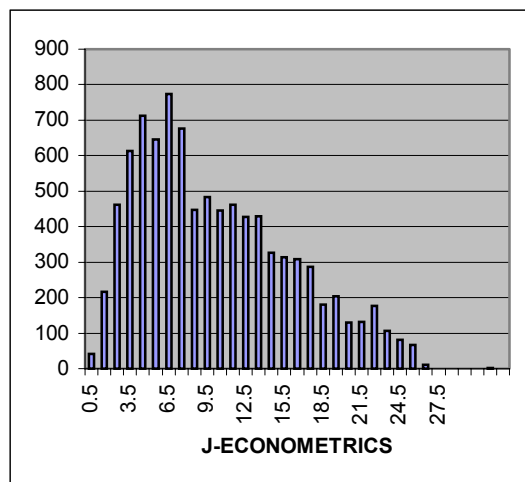
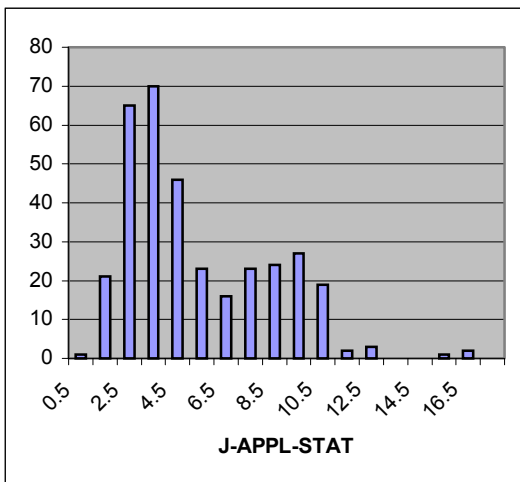
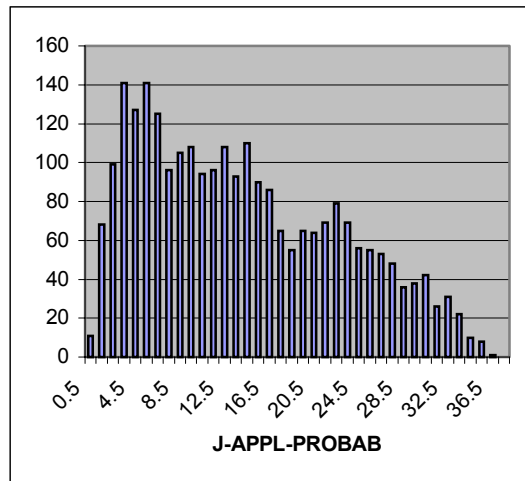
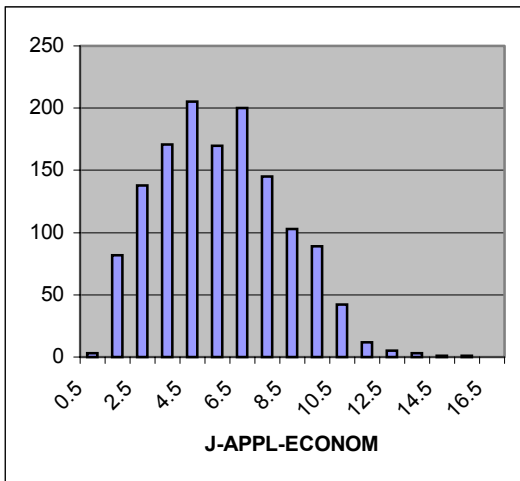
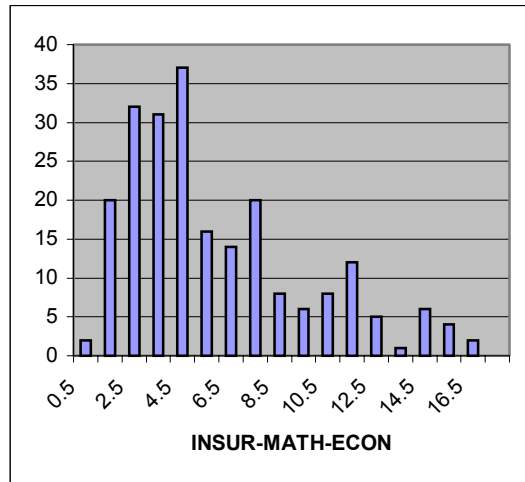
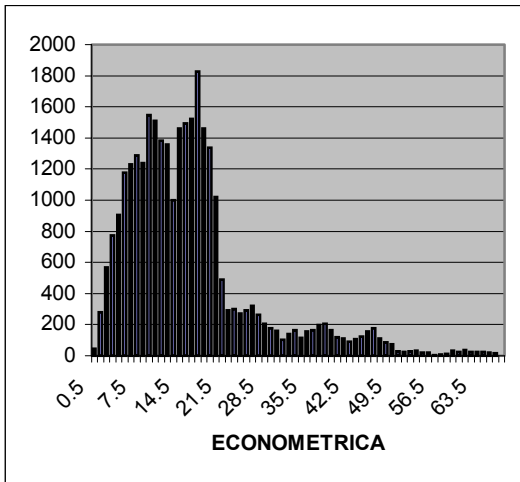
Tanto en una como en otra tabla, en la primera columna están los valores de la variable antigüedad y en la primera fila una etiqueta de cada revista de acuerdo a la tabla que sigue.

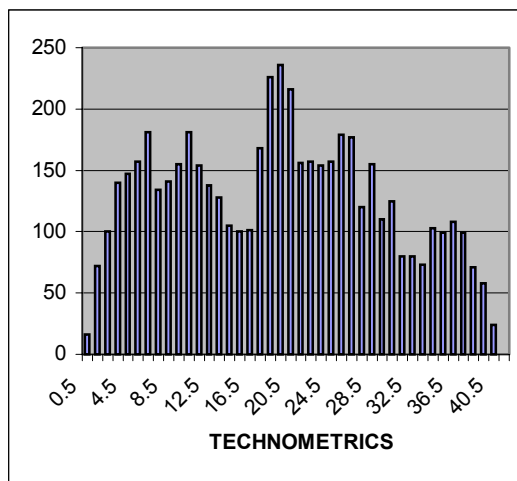
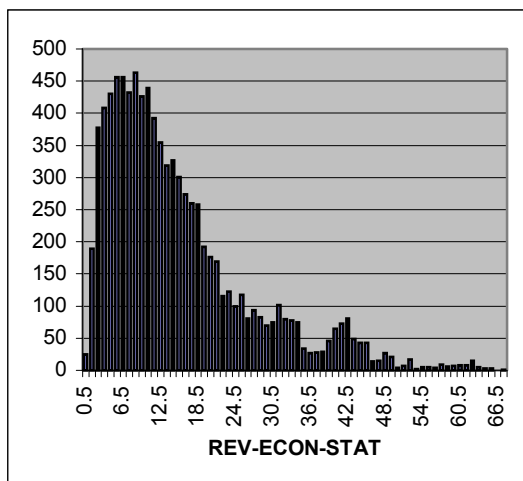
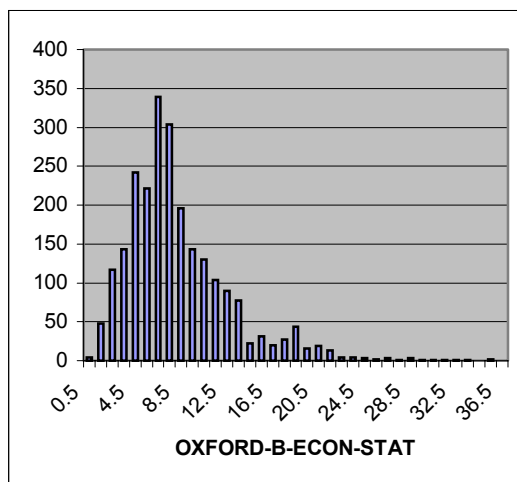
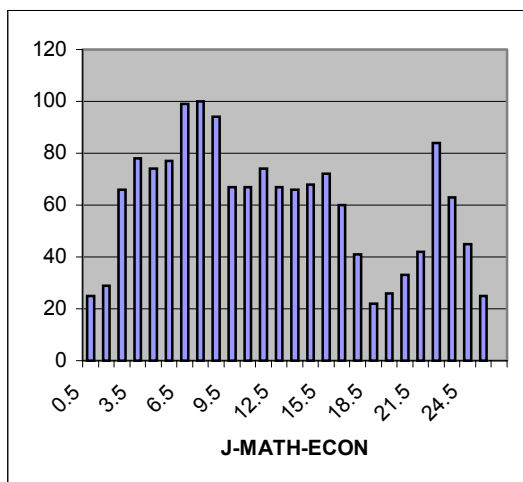
Nº	Nombre de la revista	Abreviatura
1	APPLIED ECONOMICS	APPL-ECON
2	ECONOMETRIC THEORY	ECONOMET-THEOR
3	ECONOMETRICA	ECONOMETRICA
4	INSURANCE MATHEMATICS&ECONOMICS	INSUR-MATH-ECON
5	JOURNAL OF APPLIED ECONOMETRICS	J-APPL-ECONOM
6	JOURNAL OF APPLIED PROBABILITY	J-APPL-PROBAB
7	JOURNAL OF APPLIED STATISTICS	J-APPL-STAT
8	JOURNAL OF ECONOMETRICS	J-ECONOMETRICS
9	JOURNAL OF MATHEMATICAL ECONOMICS	J-MATH-ECON
10	OXFORD BULLETIN OF ECONOMICS AND STATISTICS	OXFOR-B-ECON-STAT
11	REVIEW OF ECONOMICS AND STATISTICS	REV-ECON-STAT
12	TECHNOMETRICS	TECHNOMETRICS

A continuación, ofrecemos los histogramas correspondientes a cada revista para los datos no censurados.



Apéndice A: Datos y gráficas de antigüedad de las citas.





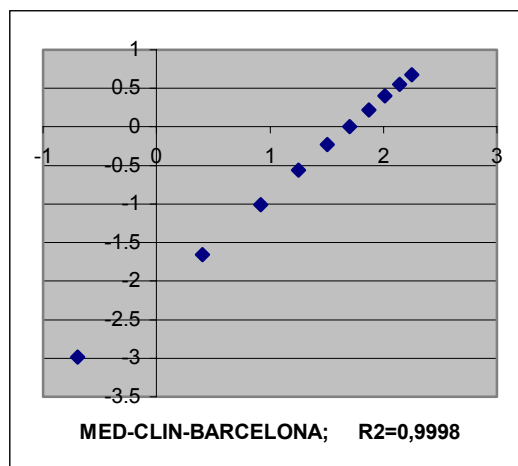
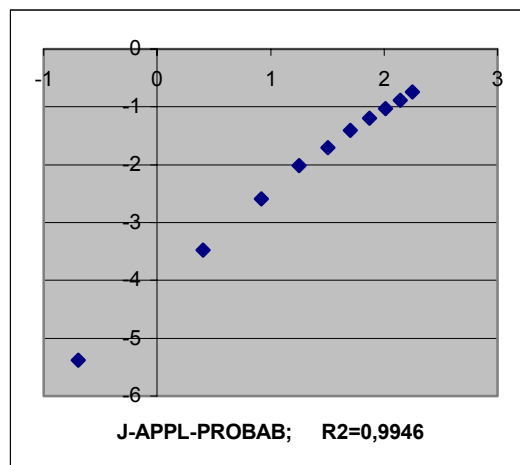
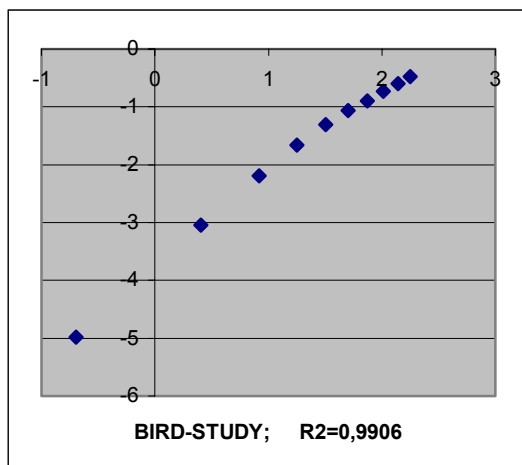
• APÉNDICE A.2

A continuación ofrecemos los gráficos que muestran si el modelo Weibull es adecuado para cada revista (ver página 212). Junto a cada gráfico, medimos el grado de ajuste de los puntos a una recta a través del coeficiente  $R^2$ .

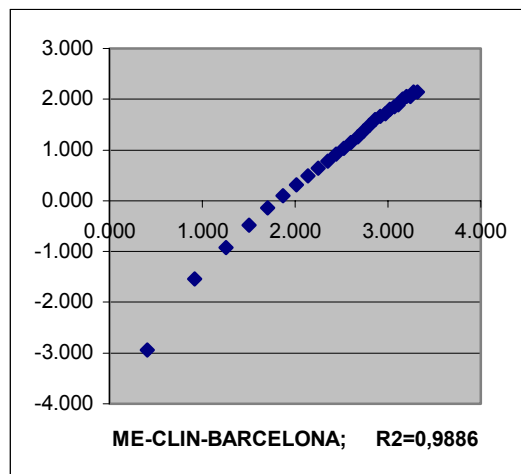
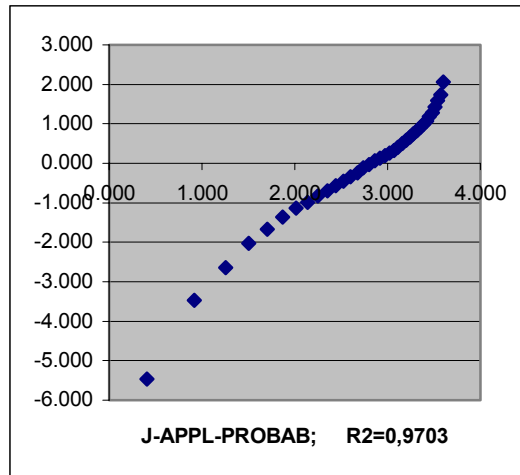
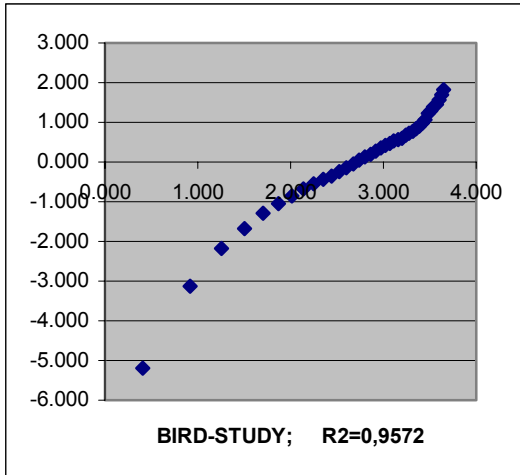


◆ Tres temáticas distintas.

1. Datos censurados.

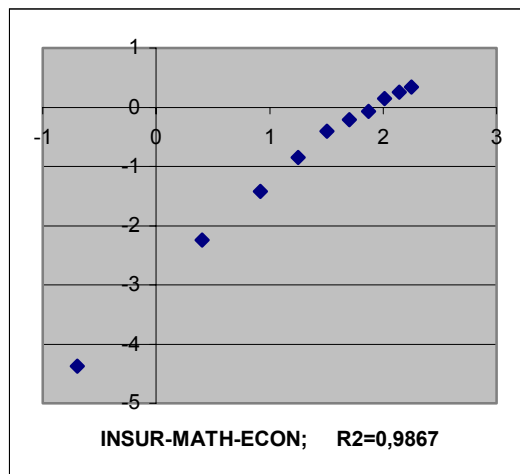
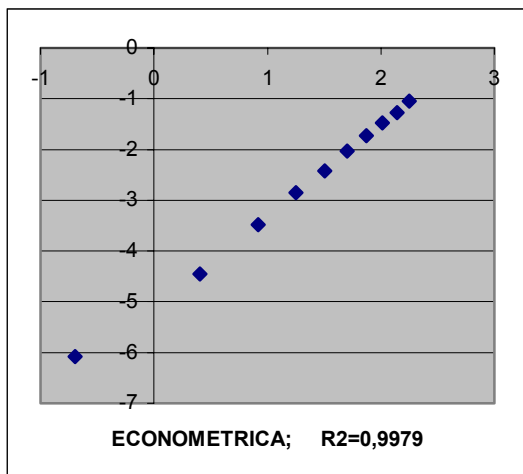
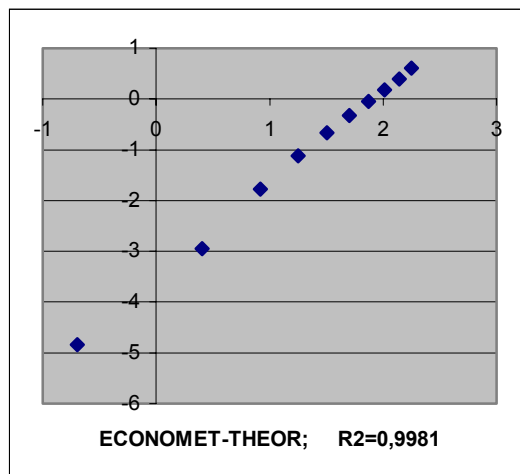
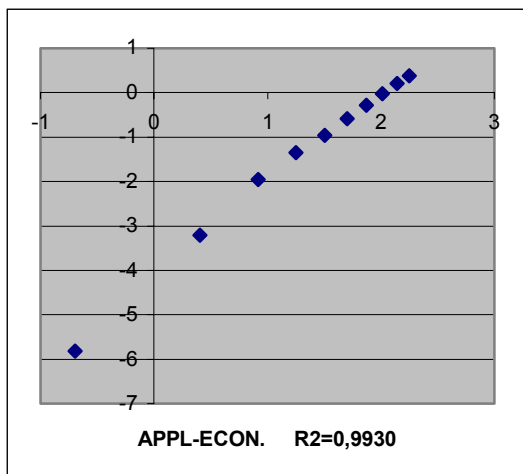


2. Datos sin censura.



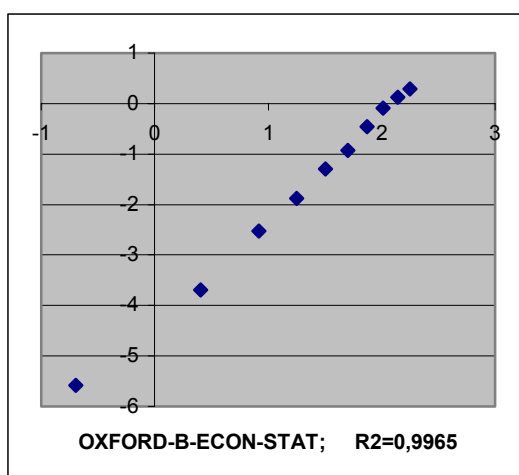
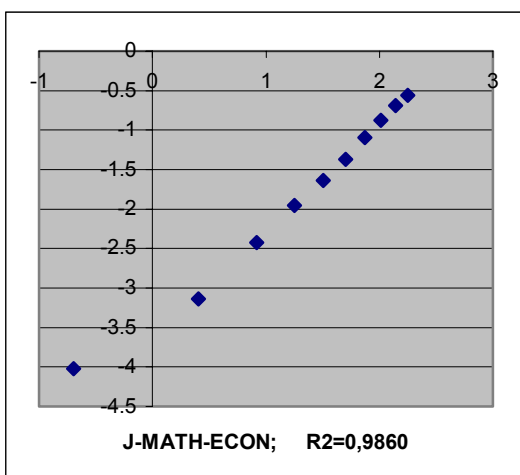
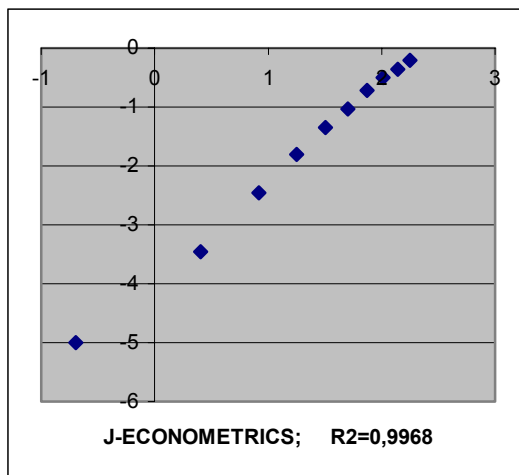
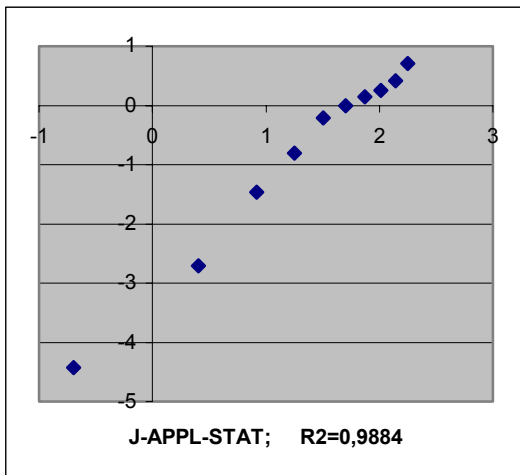
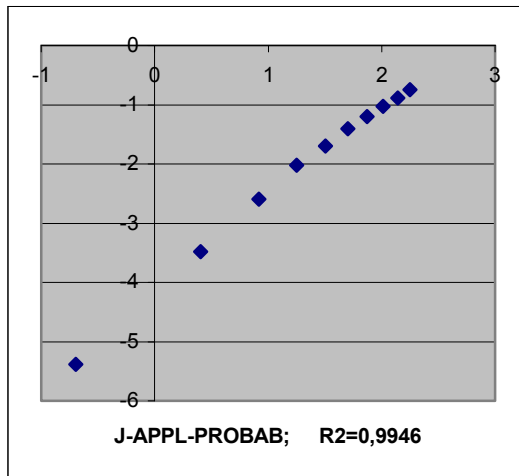
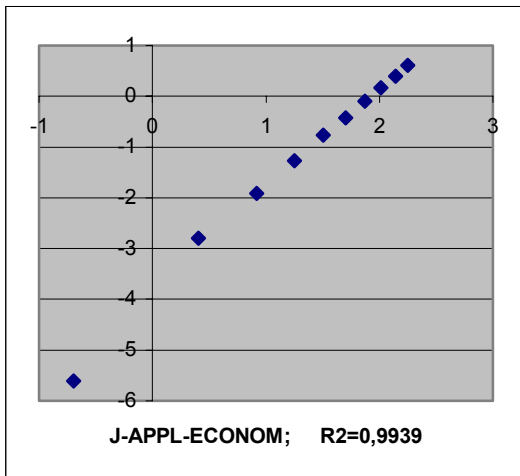
◆ Doce revistas del ámbito de Economía Aplicada.

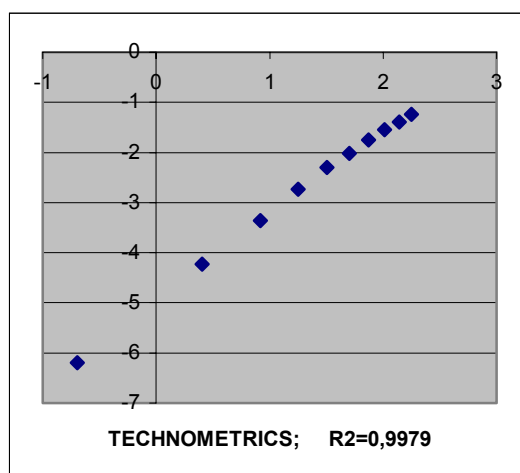
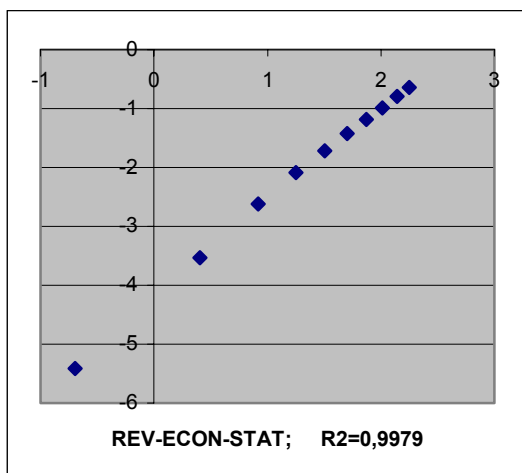
1. Datos censurados.



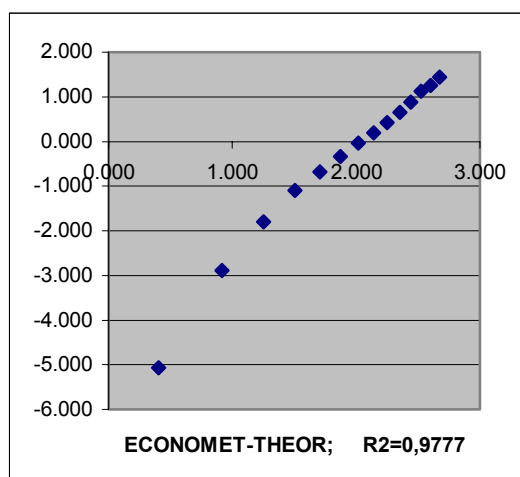
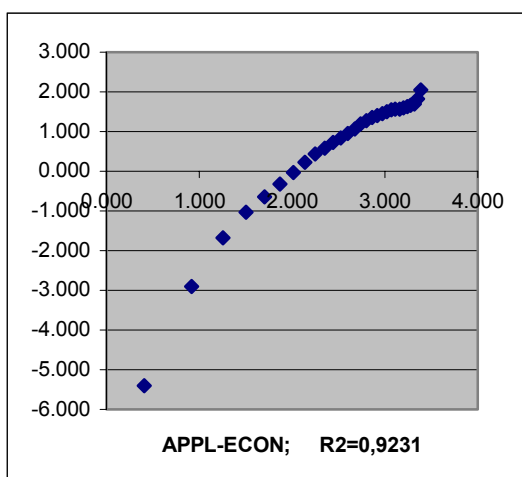
Apéndice A: Datos y gráficas de antigüedad de las citas.

---



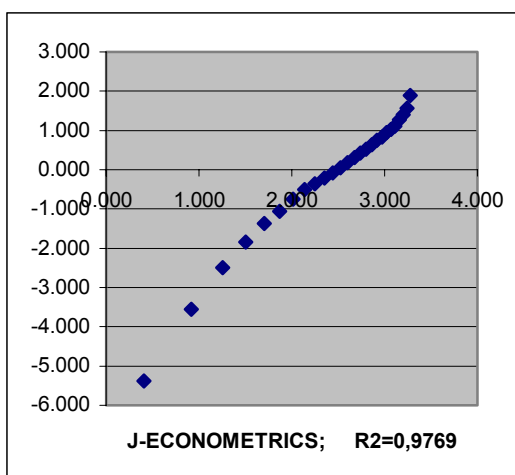
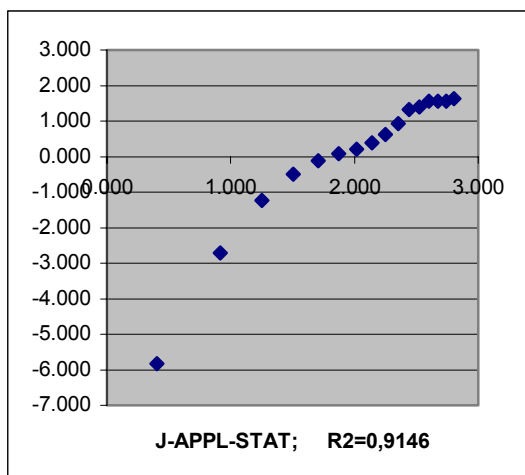
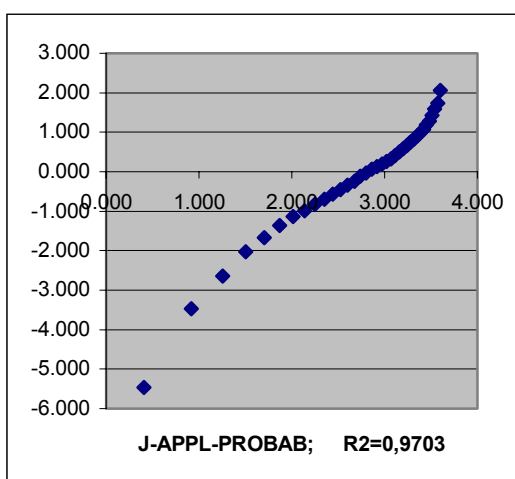
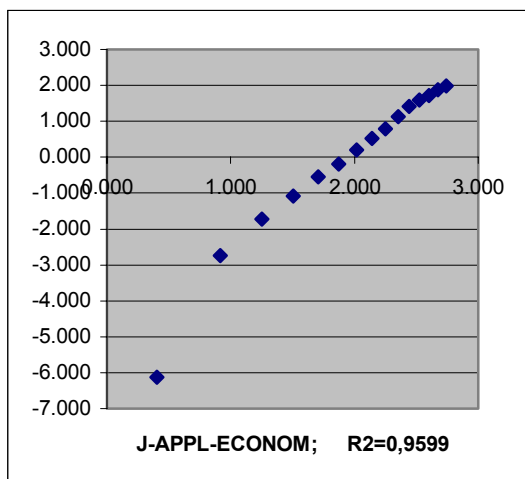
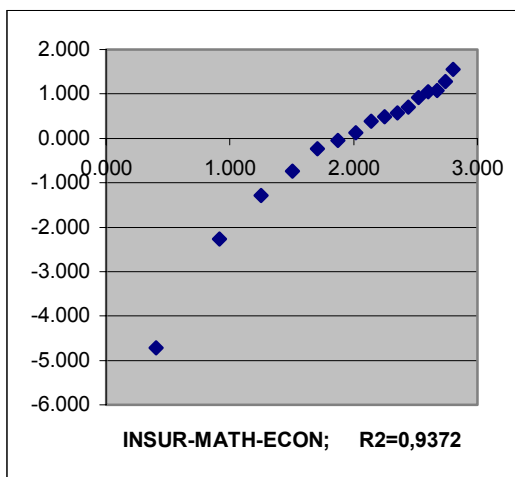
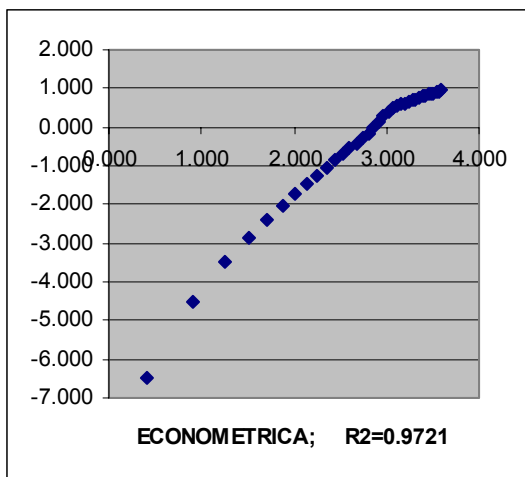


## 2. Datos sin censura.

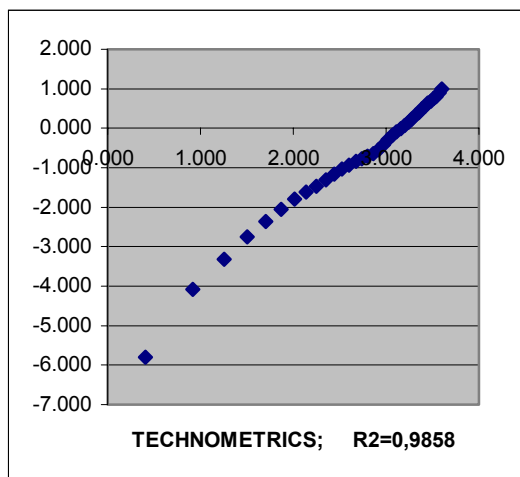
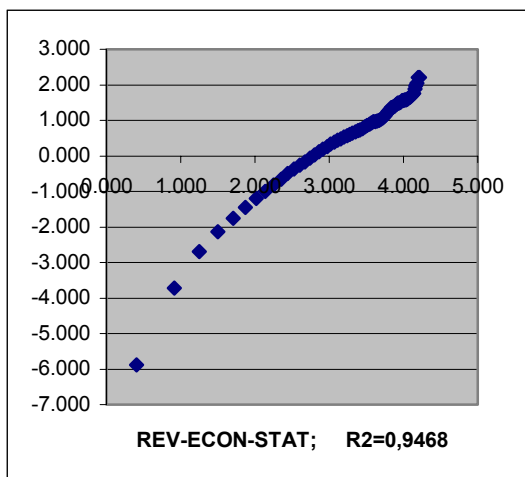
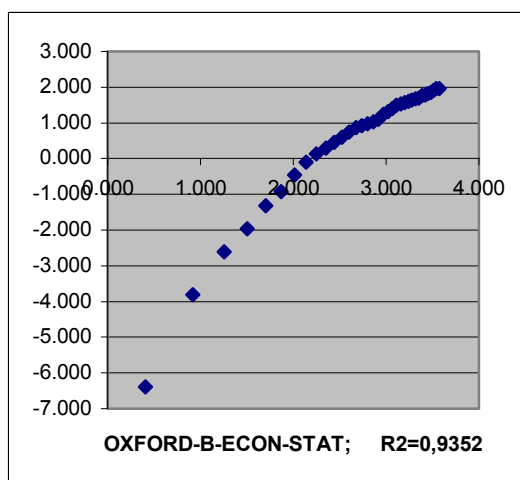
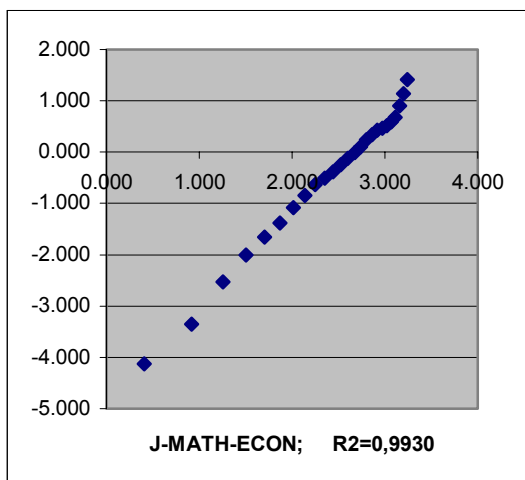


Apéndice A: Datos y gráficas de antigüedad de las citas.

---



Apéndice A: Datos y gráficas de antigüedad de las citas.







## **APÉNDICE B: LISTADOS DE LOS PROGRAMAS CON MATEMÁTICA 3.0**

### **• APÉNDICE B.1**

A continuación ofrecemos el listado del programa con el cual hemos estudiado el comportamiento frecuencialista de los intervalos bayesianos obtenidos para el modelo exponencial con parámetro umbral y censura Tipo I (ver página 186).

```
m=1000;
t0=5;
l0=1/5;
a=0.1;
Needs["Statistics`ContinuousDistributions`"]
Do[
  cl=0;
  ct=0;
  L=Table[t0-Log[0.2]*l0,{i,1,n}];
  Do[
    s=Table[t0+Random[GammaDistribution[1,l0]},{i,1,n}];
    f[x_,y_]=If[x≤y,x,y];
    t=Table[f[s[[i]],L[[i]]],{i,1,n}];
    d=Table[If[t[[i]]<L[[i]],1,0],{i,1,n}];
    r=Sum[d[[i]],{i,1,n}];
    T=Sum[t[[i]]*d[[i]]+L[[i]]*(1-d[[i]]),{i,1,n}];
    G[x_]=T-n*x;
    tmin=Min[t];
    K=((r-2)!/G[tmin]^(r-1)-(r-2)!/G[0]^(r-1))/n;
```

```
pt[x_]=Gamma[r]G[x]^(-r)/K;
Pt[x_]=Integrate[pt[y],{y,0,x}];
pl[x_]=(Exp[-x*G[tmin]]-Exp[-x*G[0]])*x^(r-2)/(n*K);
Pl[x_]=((r-2)!*CDF[GammaDistribution[r-1,1/G[tmin]],x]/
G[tmin]^(r-1)-(r-2)!*CDF[GammaDistribution[r-
1,1/G[0]],x]/G[0]^(r-1))/(K*n);
ext=G[0]*(1-((r-2)!/(a*n*K*G[0]^(r-1)+(r-2)!))^(1/(r-
1)))/n;
exl1=Re[x/.FindRoot[Pl[x]==a/2,{x,1/10},MaxIterations-
>50]];
exl2=Re[x/.FindRoot[Pl[x]==1-
a/2,{x,1/10},MaxIterations->50]];
If[ext<=t0,ct++,];
If[exl1<=1/10&&1/10<=exl2,cl++,];

,{m}];

Print["n=",n,";α=",a,";m=",m,";θ=",t0,";λ=",1/10,";NC(θ)="
,N[ct/m],";NC(λ)=",N[cl/m]];

,{n,10,50,10}]
```

## • APÉNDICE B.2

A continuación ofrecemos el listado del programa con el cual hemos hecho la comparación de los resultados que se obtienen para el modelo exponencial con parámetro umbral y censura Tipo I usando las distribuciones a priori  $\pi(\theta,\lambda)$  y  $\pi^*(\theta,\lambda)$  (ver páginas 193 y 194).

```
(**** Parte Común ****)
(**** Definiciones ****)
t0=5;
l0=1;
n=30;
a=0.05;
m1=100;
m2=100;
pa=15;
laf=Table[0,{j,1,m1}];
tef=Table[0,{j,1,m1}];
Needs["Statistics`DataManipulation`"]
Needs["Statistics`ContinuousDistributions`"]
L=Table[t0-Log[0.2]*10,{i,1,n}];
s=Table[t0+Random[GammaDistribution[1,l0]},{i,1,n}];
c[x_,y_]=If[x<=y,x,y];
t=Table[c[s[[i]],L[[i]]],{i,1,n}];
d=Table[If[t[[i]]<L[[i]],1,0],{i,1,n}];
T=Sum[t[[i]]*d[[i]]+L[[i]]*(1-d[[i]]),{i,1,n}];
r=Sum[d[[i]],{i,1,n}];
G[x_]=T-n*x;
tmin=Min[t];

(**** Aquí empieza el bloque del Algoritmo
de Gibbs en el primer caso ****)
f[x_,y_]=x^(r-1)*Exp[-x*(T-n*y)]*Sqrt[Sum[1-Exp[-
x*(L[[i]]-y)],{i,1,n}]];
te=Table[0,{i,1,m2+1+pa*m1}];
la=Table[0,{i,1,m2+1+pa*m1}];
te[[1]]=tmin/2+tmin/2*Random[];
la[[1]]=l0;
```

---

```
j=0;
Do[
  u=1;
  coc=0;
  While[u>coc,
    li=Random[GammaDistribution[r,1/(T-n*te[[i-1]])]];
    u=Random[];
    coc=Sqrt[(1/n)*Sum[1-Exp[-li*(L[[k]]-te[[i-
1]])]],{k,1,n}]];
    la[[i]]=li;
  ];
  u=1;
  coc=0;
  While[u>coc,
    tei=Log[1+Random]*(Exp[n*la[[i]]*tmin]-
1)]/(n*la[[i]]);
    u=Random[];
    coc=Sqrt[(1/n)*Sum[1-Exp[-la[[i]]*(L[[k]]-
te[[i]])]],{k,1,n}]];
    te[[i]]=tei;
  ];
  If[i>m2+1&&Mod[(i-m2-
1),pa]==0,++j;laf[[j]]=la[[i]];tef[[j]]=te[[i]];
  Print[j]];
,{i,2,m2+1+pa*m1}];

Print["Intervalo para  $\lambda$  (Gibbs)"];
Print[Quantile[laf,0.025]," ",Quantile[laf,0.975]];
Print["Media de  $\lambda$  (Gibbs): ",Mean[laf]];
Print["Intervalo para  $\theta$  (Gibbs)"];
Print[Quantile[tef,0.05]," ",tmin];
```

```
Print["Media de  $\theta$  (Gibbs): ", Mean[tef]];

(**** Aquí empieza el bloque de fórmulas
      exactas ****)

K=Integrate[x^(r-1)*Exp[-x*G[y]],{x,0,+∞},{y,0,tmin}];
k=((r-2)!/G[tmin]^(r-1)-(r-2)!/G[0]^(r-1))/n;
pt[x_]=Gamma[r]G[x]^(-r)/K;
Pt[x_]=Integrate[pt[y],{y,0,x}];
pl[x_]=(Exp[-x*G[tmin]]-Exp[-x*G[0]])*x^(r-2)/(n*K);
Pl[x_]=((r-2)!*CDF[GammaDistribution[r-
1,1/G[tmin]],x]/G[tmin]^(r-1)-(r-2)!*
CDF[GammaDistribution[r-1,1/G[0]],x]/G[0]^(r-1))/(K*n);
elt=G[0]*(1-((r-2)!/(a*n*K*G[0]^(r-1)+(r-2)!))^(1/(r-
1)))/n;
e11=FindRoot[Pl[x]==a/2,{x,1/10}];
e21=FindRoot[Pl[x]==1-a/2,{x,1/10}];
ml=Integrate[x*pl[x],{x,0,+∞}];
mt=Integrate[x*pt[x],{x,0,tmin}];
Print["Intervalo para  $\lambda$  (Fórmulas)"];
Print[e11,"  ",e21];
Print["Media de  $\lambda$  (fórmulas): ",ml];
Print["Intervalo para  $\theta$  (Fórmulas)"];
Print[elt,"  ",tmin];
Print["Media de  $\theta$  (fórmulas): ",mt];

(**** Ahora los gráficos ****)
(**** Gráficos del primer bloque ****)

Needs["Graphics`Graphics`"]
lmi=Min[laf];
```

## Apéndice B: Listados de los programas con Mathematica 3.0.

---

```
lma=Max[laf];
tmi=Min[tef];
tma=Max[tef];
ninter=10;
lol=(lma-lmi)/ninter;
rel=BinCounts[laf,{lmi,lma,lol}]/(m1*lol);
rel[[1]]+=1/(m1*lol);
exl=Table[lmi+lol*(k-1),{k,1,ninter+1}];
lot=(tma-tmi)/ninter;
ext=Table[tmi+lot*(k-1),{k,1,ninter+1}];
ret=BinCounts[tef,{tmi,tma,lot}]/(m1*lot);
ret[[1]]+=1/(m1*lot);
g1l=Graphics[
Table[Line[{{exl[[i]],0},{exl[[i]],rel[[i]]},{exl[[i+1]],
rel[[i]]},{
exl[[i+1]],0}}],{i,1,ninter}]];
g1t=Graphics[
Table[Line[{{ext[[i]],0},{ext[[i]],ret[[i]]},{ext[[i+1]],
ret[[i]]},{
ext[[i+1]],0}}],{i,1,ninter}]];

(**** Gráficos del segundo bloque ****)

wwt=0.1*(tma-tmi);
g2t=Plot[pt[x],{x,tmi-wwt,tmin}];
wwl=0.1*(lma-lmi);
g2l=Plot[pl[x],{x,lmi-wwl,lma+wwl}];
Show[{g1l,g2l},PlotRange->{{lmi-wwl,lma+wwl},All},
Frame->{True,True,False,False},
PlotLabel->FontForm["Parámetro  $\lambda$ ",{"Arrus-Bold",10}],
FrameTicks->{N[{lmi,(lmi+lma)/2,lma},3],Automatic}]
```

---

```
Show[{g1t,g2t},PlotRange->{{tmi-wwt,tmin},All},Frame-  
>{True,True,False,False},  
PlotLabel->FontForm["Parámetro  $\theta$ ","Arrus-Bold",10]],  
FrameTicks->{N[{tmi,(tmi+tma)/2,tma},3],Automatic}}
```

### • APÉNDICE B.3

A continuación ofrecemos el listado del programa con el cual hemos obtenido las estimaciones de los parámetros para un modelo Weibull sin censura. (ver página 198). Para ejecutarlo, hay que cambiar previamente el vector 'n' introduciendo los datos correspondientes para cada revista.

```
Needs["Statistics`DataManipulation`"]  
Needs["Statistics`ContinuousDistributions`"]  
n={46,278,569,774,907,1175,1230,1287,1240,1546,1512,1383,  
1360,999,1462,1492,1522,1827,1460,1338,1020,488,293,302,2  
72,294,323,264,206,175,160,101,140,166,115,158,166,197,20  
4,163,120,113,92,107,123,158,175,110,88,76,30,25,28,34,22  
,20,3,9,14,31,26,35,25,26,25,22,16};  
k=Length[n];  
nt=Sum[n[[i]],{i,1,k}];  
m1=1000;  
m2=500;  
pa=15;  
uf=Table[0,{1,1,m1}];  
bf=Table[0,{1,1,m1}];  
xx=Table[Log[i-0.5],{i,1,k}];  
xme=Sum[xx[[i]]*n[[i]],{i,1,k}]/nt  
x=Table[xx[[i]]-xme,{i,1,k}];  
ur=1.89778;
```

## Apéndice B: Listados de los programas con Mathematica 3.0.

---

```
br=0.618168809;
bmv=b/.
FindRoot[Sum[n[[i]]*x[[i]]*Exp[x[[i]]/b],{i,1,k}]/
Sum[n[[i]]*Exp[x[[i]]/b],{i,1,k}]==b,{b,br}]/N;
umv=bmv*(-
Log[nt]+Log[Sum[n[[i]]*Exp[x[[i]]/bmv],{i,1,k}]])/N;
fu[u_,b_]=Exp[-nt*u/b]*Exp[-Sum[n[[i]]*Exp[(x[[i]]-
u)/b],{i,1,k}]];
fb[u_,b_]=b^-(nt+1)*Exp[-nt*u/b]*Exp[-
Sum[n[[i]]*Exp[(x[[i]]-u)/b],{i,1,k}]];
dlb[u_,b_]=(-b*(nt+1)+nt*u+Sum[n[[i]]*(x[[i]]-
u)*Exp[(x[[i]]-u)/b],{i,1,k}])/
b^2;
d2lb[u_,b_]=D[dlb[u,b],b];
d2lu[u_,b_]=-Sum[n[[i]]*Exp[(x[[i]]-u)/b],{i,1,k}]/b^2;
us=Table[0,{1,1,m2+1+pa*m1}];
bs=Table[0,{1,1,m2+1+pa*m1}];
us[[1]]=umv;
bs[[1]]=bmv;

(***) Bloque del muestreo de Gibss (***)
l=0;
Do[

(***) Comienza la generación de b (***)

bmax=b/.FindRoot[dlb[us[[j-1]],b]==0,{b,bs[[1]]}];
print["bmax = ",bmax];
vma=fb[us[[j-1]],bmax];
var=-1/d2lb[us[[j-1]],bmax];
r=6*Sqrt[var];
```

---



```
u1=bmv;
u2=1000;
z=1000;
While[fb[us[[j-1]],u1]<u2,
  z=Random[];
  u1=2*r*Random[]-r+bmax;
  u2=vma*Random[];
];
bs[[j]]=u1;

  (** Comienza la generación de u **)

  umax=bs[[j]]*Log[Sum[n[[i]]*Exp[x[[i]]/bs[[j]]],{i,1,k}
/nt]//N;
  vma=fu[umax,bs[[j]]];
  var=-1/d2lu[umax,bs[[j]]];
  r=6*Sqrt[var];
  u1=0;
  u2=1000;
  z=1000;
  While[fu[u1,bs[[j]]]<u2,
    z=Random[];
    u1=2*r*Random[]-r+umax;
    u2=vma*Random[];
  ];
  us[[j]]=u1;
  If[j>m2+1&&Mod[(j-m2-
1),pa]==0,++1;uf[[1]]=us[[j]];bf[[1]]=bs[[j]];Print[1]];
,{j,2,m2+1+pa*m1}];

Print["Valores de u:",uf];
```

---

```
Print["Valores de b:",bf];
Print["Máximo-Verosímiles: "];
Print["u: ",umv];
Print["b: ",bmv];
Print["Estimaciones bayesianas (a través de la media) e
intervalos (95%):"];
Print["u: ",Sum[uf[[1]],{1,1,m1}]/m1,";(",Quantile[uf,0.025]," ",
Quantile[uf,0.975],")"];
Print["b: ",Sum[bf[[1]],{1,1,m1}]/m1," ;
(",Quantile[bf,0.025]," ",
Quantile[bf,0.975],")"]
```

#### • APÉNDICE B.4

A continuación ofrecemos el listado del programa con el cual hemos obtenido las estimaciones de los parámetros para un modelo Weibull con censura tipo I. (ver página 202). Para ejecutarlo, hay que cambiar previamente el vector 'n' introduciendo los datos correspondientes para cada revista.

```
Needs["Statistics`DataManipulation`"]
Needs["Statistics`ContinuousDistributions`"]
k=11;
n={71,292,579,817,909,1180,1217,1305,1274,1572,22022};
nt=Sum[n[[i]],{i,1,k}];
nd=nt-n[[k]];
m1=1000;
m2=500;
pa=15;
uf=Table[0,{1,1,m1}];
bf=Table[0,{1,1,m1}];
```

```

xx=Table[Log[i-0.5],{i,1,k}];
xme=Sum[xx[[i]]*n[[i]},{i,1,k-1}]/nd
x=Table[xx[[i]]-xme,{i,1,k}];
eta=Log[10.5]/N;
ur=1.702159;
br=0.797246;
bmv=b/.
FindRoot[Sum[n[[i]]*x[[i]]*Exp[x[[i]]/b],{i,1,k}]/
Sum[n[[i]]*Exp[x[[i]]/b],{i,1,k}]==b,{b,br}]/N;
umv=bmv*(-
Log[nd]+Log[Sum[n[[i]]*Exp[x[[i]]/bmv],{i,1,k}]])/N;
fu[u_,b_]=Exp[-nd*u/b]*Exp[-Sum[n[[i]]*Exp[(x[[i]]-
u)/b],{i,1,k}]];
fb[u_,b_]=b^-(nd+1)*Exp[-nd*u/b]*Exp[-
Sum[n[[i]]*Exp[(x[[i]]-u)/b],{i,1,k}]];
s[u_,b_]=Sqrt[1-Exp[-Exp[(eta-u)/b]]];
dlb[u_,b_]=(-b*(nd+1)+nd*u+Sum[n[[i]]*(x[[i]]-
u)*Exp[(x[[i]]-u)/b],{i,1,k}])/
b^2;
d2lb[u_,b_]=D[dlb[u,b],b];
d2lu[u_,b_]=-Sum[n[[i]]*Exp[(x[[i]]-u)/b],{i,1,k}]/b^2;
us=Table[0,{1,1,m2+1+pa*m1}];
bs=Table[0,{1,1,m2+1+pa*m1}];
us[[1]]=umv;
bs[[1]]=bmv;

(***) Bloque del muestreo de Gibss (***)

l=0;
Do[

```

## Apéndice B: Listados de los programas con Mathematica 3.0.

---

```
(** Comienza la generación de b **)

bmax=b/.FindRoot[dlb[us[[j-1]],b]==0,{b,bs[[1]]}];
vma=fb[us[[j-1]],bmax];
var=-1/d2lb[us[[j-1]],bmax];
r=6*Sqrt[var];
u1=bmv;
u2=1000;
z=1000;
While[fb[us[[j-1]],u1]<u2&& z>s[us[[j-1]],u1],
  z=Random[];
  u1=2*r*Random[]-r+bmax;
  u2=vma*Random[];
];
bs[[j]]=u1;

(** Comienza la generación de u **)

umax=bs[[j]]*Log[Sum[n[[i]]*Exp[x[[i]]/bs[[j]]],{i,1,k}]/
nd]/N;
vma=fu[umax,bs[[j]]];
var=-1/d2lu[umax,bs[[j]]];
r=6*Sqrt[var];
u1=0;
u2=1000;
z=1000;
While[fu[u1,bs[[j]]]<u2&& z>s[u1,bs[[j]]],
  z=Random[];
  u1=2*r*Random[]-r+umax;
  u2=vma*Random[];
];
```

---

```
us[[j]]=u1;

If[j>m2+1&&Mod[(j-m2-
1),pa]==0,++1;uf[[1]]=us[[j]];bf[[1]]=bs[[j]];Print[l]];

,{j,2,m2+1+pa*m1}}];

Print["Valores de u:",uf];
Print["Valores de b:",bf];
Print["Máximo-Verosímiles: "];
Print["u: ",umv];
Print["b: ",bmv];
Print["Estimaciones bayesianas (a través de la media) e
intervalos (95%):"];
Print["u: ",Sum[uf[[1]},{1,1,m1}]/m1," ;
(",Quantile[uf,0.025]," , ",
Quantile[uf,0.975],")"];
Print["b: ",Sum[bf[[1]},{1,1,m1}]/m1," ;
(",Quantile[bf,0.025]," , ",
Quantile[bf,0.975],")"]
```



• **BIBLIOGRAFÍA.**

1. **BIBLIOGRAFÍA GLOBAL.**

- ◆ Akahira, M. and Takeuchi, K. (1991), “A Definition of Information Amount Applicable to Non-Regular Cases”, *Journal of Computing and Information*, **2**, 71-92.
- ◆ Akahira, M. and Takeuchi, K. (1995), *Non-Regular Statistical Estimation*, Springer-Verlag, New-York.
- ◆ Azzalini, A. (1996), *Statistical Inference Based on the Likelihood*, Chapman and Hall, London.
- ◆ Bardoff-Nielsen, O.E. and Cox, D.R. (1994), *Inference and Asymptotic*. Chapman and Hall, London.
- ◆ Barry, C.A., Castillo, E. and Sarabia, J.M. (1992), *Conditionally Specified Distributions*, Lecture Notes in Statistics, Springer- Verlag, Berlin-Heidelberg.
- ◆ Basulto, J. (1997), “Funciones a Priori Imparciales Unidimensionales”, *Estadística Española*, **39**, nº **142**, 99-128.
- ◆ Bayes, T.R. (1763), “An essay towards solving a problem in the doctrine of chances”, *Philosophical Transactions of the Royal Society*, **53**, 370-418. Reprinted in *Biometrika*, **45** (1958), 243-315.
- ◆ Berger, J. and Bernardo, J.M. (1992), “On the development of the reference prior method”, in *Bayesian Statistics 4: Proceeding of the Fourth Valencia International*

## **Bibliografía.**

---

*Meeting*, eds. J.M. Bernardo, J.M., Berger, J.O., Dawid, A.P. and Smith, A.F., 35-60, Clarendon Press, Oxford.

◆ Box, G.E. and Tiao, G.C. (1973), *Bayesian Inference in Statistical Analysis*, Addison-Wesley: Reading, Massachusetts.

◆ Brookes, B.C. (1974), “Obsolescence of special library periodical: sampling errors utility contours”, *Journal of the American Society for Information Science*, **21**, nº 5, 320-329.

◆ Brown, L.P. (1967), “The Conditional Level of Students’s t test”, *Ann. Math. Statist.*, **38**, 1068-1071.

◆ Buehler, R.J. (1959), “Some Validity Criteria for Statistical Inferences”, *Ann. Math. Statist.*, **30**, 845-863.

◆ Buehler, R.J. and Feddersen, A.P. (1963), “Note on a Conditional Property of Student’s t”, *Ann. Math. Statist.*, **34**, 1098-1100.

◆ Callon, M., Courtial, J.P. and Penan, H. (1995), *Cienciometría. La Medición de la Actividad Científica: de la Bibliometría a la Vigilancia Tecnológica*. Ediciones Trea, Gijón.

◆ Cox, D.R. and Reid, N. (1987), “Parameter Orthogonality and Approximate Conditional Inference (with Discussion)”, *J.R. Statist. Soc., Ser. B*, **49**, 1-39.

◆ Datta, G.S. and Gosh, J.K. (1995), “On Priors Providing Frequentist Validity for Bayesian Inference”, *Biometrika*, **82**, 37-45.



- ◆ Fisher, R.A. (1956), *Statistical Methods and Scientific Inference*. 3<sup>rd</sup>. Edición. Oliver and Boyd .Edinburgh.
  
- ◆ Gelfand, A.E. and Smith, A.F. (1990), “Sampling-Based Approaches to Calculating Marginal Densities”, *J.Amer. Statist. Assoc.*, **85**, 398-409.
  
- ◆ Geman, S. And Geman, D. (1984), “Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **6**, 721-741.
  
- ◆ George, E.I. and McCulloch, R. (1993), “On obtaining Prior Distributions”, *Journal of Statistical Planning and Inference*”, **37**, 169-179.
  
- ◆ Ghosal, S. and Samanta, T. (1997), “Asymptotic Expansions of Posterior Distributions in Non-Regular Cases”, *Ann. Inst. Statist. Math.*, **49**, n° **1**, 181-197.
  
- ◆ Ghosal, S., Ghosh, J.K. and Samanta, T. (1995), “On Convergence of Posterior Distributions”, *Ann. Statist.*, **23**, 2145-2152.
  
- ◆ Glänzel, W. And Shoepflin, U. (1994), “A Stochastic Model for the Ageing of Scientific Literature”, *Scientometrics*, **30**, n° **1**, 94-108.
  
- ◆ Glänzel, W. And Shoepflin, U. (1995), “A Bibliometric Study on Ageing and Reception Processes of Scientific Literature”, *Journal of Information Science*, **21**, 37-53
  
- ◆ Gosh, J.K., Ghosal, S. and Samanta, T. (1994), “Stability and Convergence of Posterior in Non-Regular Problems”, *Statistical Decision Theory and Related Topics V*, (eds. S.S. Gupta and J.O. Berger), 183-199, Springer, New-York.

## **Bibliografía.**

---

- ◆ Goutis, C. and Casella, G. (1995), “Frequentist Post-Data Inference”, *International Statistical Review*, **63**, n° 3, 325-344.
  
- ◆ Gupta, B.M. (1997), “Análisis of Distribution of the Age of Citations in Theoretical Population Genetics”, *Scientometrics*, **40**, n° 1, 139-162.
  
- ◆ Gupta, B.M. (1998), “Growth and Obsolescence of Literature in Theoretical Population Genetics”, *Scientometrics*, **42**, n° 3, 335-347.
  
- ◆ Hartigan, J.A. (1966), “Note on the Confidence-Prior of Welch and Peers”, *J.R. Statist. Soc., Ser. B*, **28**, 55-56.
  
- ◆ Huzurbazar, V.S. (1950), “Probability Distributions and Orthogonal Parameters”, *Proc. Camb. Phil. Soc.*, **46**, 281-284.
  
- ◆ Huzurbazar, V.S. (1956), “Sufficient Statistics and Orthogonal Parameters”, *Sankhya*, **17**, 217-220.
  
- ◆ Jeffreys, H. (1946), “An Invariant Form for the Prior Probability in Estimation Problems”, *Proc. Roy. Soc. (London), Ser. A*, **186**, 453-461.
  
- ◆ Jeffreys, H. (1961), *Theory of Probability*, 3<sup>rd</sup>. edition, Oxford University Press, London.
  
- ◆ Kappenman, R.F. (1975), “Conditional Confidence Intervals for Double Exponential Distributions Parameters”, *Technometrics*, **17**, 233-235.
  
- ◆ Kass, R.E. (1989), “The Geometry of Asymptotic Inference”, *Statistical Science*, **4**, n° 3, 188-234.

- ◆ Kass, R.E. (1990), “Data-Translated Likelihood and Jeffrey’s rule”, *Biometrika*, **77**, 107-114.
  
- ◆ Kass, R.E. and Wasserman, L. (1995,a), “A Reference Bayesian Test for Nested Hypotheses and its Relationship to the Schwarz criterion”, *J. Amer. Statist. Assoc.*, **90**, 928-934.
  
- ◆ Kass, R.E. and Wasserman, L. (1995,b), “The Selection of Priors Distributions by Formal Rules”, *Technical Report #583, Department of Statistics, Carnegie Mellon University*.
  
- ◆ Kosmas, K.F. (1990), “Shortest Confidence Intervals for Families of Distributions Involving Truncation Parameters”, *The American Statistician*, **44**, 167-168.
  
- ◆ Kullback, S. (1968), *Information Theory and Statistics*”, Dover Publications, Inc., New York.
  
- ◆ Lawless, J.F. (1982), *Statistical Models and Methods for Lifetime Data*, John Wiley & Sons, New York.
  
- ◆ Lehman, E.L. (1991), *Theory of Point Estimation*, Wadsworth & Brooks/Cole Advanced Books & Software. Pacific Grove, California.
  
- ◆ Lehmann, E.L. (1999), *Elements of Large-Sample Theory*, Springer-Verlag, New-York.
  
- ◆ Lindley, D.V. (1958), “Fiducial Distributions and Baye’s Theorem”, *J.R. Statist. Soc., Ser. B*, **20**, 102-107.

## **Bibliografía.**

---

- ◆ Matriccioni, E. (1994), “Shannon’s Entropy as a Measure of the “Life” of the Literature of a Discipline”, *Scientometrics*, **30**, n° 1, 129-145.
  
- ◆ Matusita, K. (1955), “Decision Rules Based on the Distance for Problems of Fit, two Samples and Estimation”, *Ann. Math. Statist.*, **26**, 631-640.
  
- ◆ Moed, H.F., Van Leeuwen, T.N. and Reedijk, J. (1999), “Towards Appropriate Indicators of Journal Impact”, *Scientometrics*, **46**, n°3, 575-589.
  
- ◆ Nelson, W.B. (1972), “Theory and Applications of Hazard Plotting for Censored Failure Data”, *Technometrics*, **14**, 945-965.
  
- ◆ Nicolau, A. (1993), “Bayesian Intervals with Good Frequentist Behaviour in the Presence of Nuisance Parameters”, *J.R. Statist. Soc., Ser. B*, **55**, 377-390.
  
- ◆ Peña, D. (1986), *Estadística. Modelos y métodos. Volumen I: Fundamentos*, Alianza Universidad Textos, Madrid.
  
- ◆ Peers, H.W. (1965), “On Confidence Points and Bayesian Probability Points in the Case of Several Parameters”, *J.R. Statist. Soc., Ser. B*, **27**, 9-16.
  
- ◆ Pitman, E.J. (1979), *Some Basic Theory for Statistical Inference*, Chapman and Hall, London.
  
- ◆ Price, D. (1986), *Little Science, Big Science... and Beyond*, Columbia University Press, New York.
  
- ◆ Reid, N. (1995), “The Roles of Conditioning in Inference”, *Statistical Science*, **10**, n° 2, 138-199.

- ◆ Robinson, G.K. (1975), “Some Counterexamples to the Theory of Confidence Intervals”, *Biometrika*, **62**, 155-161.
  
- ◆ Ruiz, R. y Jiménez, E. (1996), “Envejecimiento de la Literatura Científica en Documentación. Influencia del Origen Nacional de las Revistas. Estudio de una Muestra”, *Revista Española de Documentación Científica*, **19**, n° 1, 39-49.
  
- ◆ Smith, A.F. and Gelfand, A.E. (1992), “Bayesian Statistics Without Tears: A Sampling-Resampling Perspective”, *The American Statistician*, **46**, n°2, 84-88.
  
- ◆ Stein, C. (1956), “Inadmissibility of the Usual Estimator for the Mean of a Multivariate Normal Distribution”, in *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, **1**, 197-206, University of California Press, Berkeley.
  
- ◆ Stein, C. (1959), “An Example of Wide Discrepancy Between Fiducial and Confidence Intervals”, *Ann. Math. Statist.*, **30**, 877-880.
  
- ◆ Stigker, S.M. (1986), *The History of Statistics: The Measurement of Uncertainty Before 1900*, The Belknap Press of Harvard University Press, Cambridge-Massachusetts.
  
- ◆ Stigler, S.M. (1994), “Citation Patterns in the Journals of Statistics and Probability”, *Statistical Science*, **9**, n° 1, 94-108.
  
- ◆ Tibshirani, R. (1989), “Noninformative Priors for One Parameter of Many”, *Biometrika*, **76**, 604-608.
  
- ◆ Welch, B.L. and Peers, H.W. (1963), “On Formulae for Confidence Points Based on Integral of Weighted Likelihoods”, *J.R. Statist. Soc., Ser. B*, **25**, 318-329.

- ◆ Welsh, A.H. (1996), *Aspects of Statistical Inference*, John Wiley & Sons, New-York.
- ◆ Ye, K. (1993), “Reference Priors When the Stopping Rule Depends on the Parameter of Interest”, *Journal of the American Statistical Association*, **88**, nº 421, 360-363.
- ◆ Zabell, S.L. (1992), “R.A. Fisher and the Fiducial Argument”, *Statistical Science*, **7**, nº3, 369-387.

## **2. BIBLIOGRAFÍA POR CAPÍTULOS.**

### **CAPÍTULO 1.**

- ◆ Akahira, M. and Takeuchi, K. (1991), “A Definition of Information Amount Applicable to Non-Regular Cases”, *Journal of Computing and Information*, **2**, 71-92.
- ◆ Azzalini, A. (1996), *Statistical Inference Based on the Likelihood*, Chapman and Hall, London.
- ◆ Brown, L.P. (1967), “The Conditional Level of Students’s t test”, *Ann. Math. Statist.*, **38**, 1068-1071.
- ◆ Buehler, R.J. (1959), “Some Validity Criteria for Statistical Inferences”, *Ann. Math. Statist.*, **30**, 845-863.
- ◆ Buehler, R.J. and Feddersen, A.P. (1963), “Note on a Conditional Property of Student’s t”, *Ann. Math. Statist.*, **34**, 1098-1100.

- ◆ Goutis, C. and Casella, G. (1995), “Frequentist Post-Data Inference”, *International Statistical Review*, **63**, n° 3, 325-344.
  
- ◆ Jeffreys, H. (1946), “An Invariant Form for the Prior Probability in Estimation Problems”, *Proc. Roy. Soc. (London), Ser. A*, **186**, 453-461.
  
- ◆ Kullback, S. (1968), *Information Theory and Statistics*”, Dover Publications, Inc., New York.
  
- ◆ Lehman, E.L. (1991), *Theory of Point Estimation*, Wadsworth & Brooks/Cole Advanced Books & Software. Pacific Grove, California.
  
- ◆ Matusita, K. (1955), “Decision Rules Based on the Distance for Problems of Fit, two Samples and Estimation”, *Ann. Math. Statist.*, **26**, 631-640.
  
- ◆ Peña, D. (1986), *Estadística. Modelos y métodos. Volumen I: Fundamentos*, Alianza Universidad Textos, Madrid.
  
- ◆ Pitman, E.J. (1979), *Some Basic Theory for Statistical Inference*, Chapman and Hall, London.
  
- ◆ Reid, N. (1995), “The Roles of Conditioning in Inference”, *Statistical Science*, **10**, n° 2, 138-199.
  
- ◆ Robinson, G.K. (1975), “Some Counterexamples to the Theory of Confidence Intervals”, *Biometrika*, **62**, 155-161.

**CAPÍTULO 2.**

- ◆ Barry, C.A., Castillo, E. and Sarabia, J.M. (1992), *Conditionally Specified Distributions*, Lecture Notes in Statistics, Springer-Verlag, Berlin-Heidelberg.
- ◆ Box, G.E. and Tiao, G.C. (1973), *Bayesian Inference in Statistical Analysis*, Addison-Wesley: Reading, Massachusetts.
- ◆ Cox, D.R. and Reid, N. (1987), “Parameter Orthogonality and Approximate Conditional Inference (with Discussion)”, *J.R. Statist. Soc., Ser. B*, **49**, 1-39.
- ◆ Datta, G.S. and Gosh, J.K. (1995), “On Priors Providing Frequentist Validity for Bayesian Inference”, *Biometrika*, **82**, 37-45.
- ◆ Hartigan, J.A. (1966), “Note on the Confidence-Prior of Welch and Peers”, *J.R. Statist. Soc., Ser. B*, **28**, 55-56.
- ◆ Huzurbazar, V.S. (1950), “Probability Distributions and Orthogonal Parameters”, *Proc. Camb. Phil. Soc.*, **46**, 281-284.
- ◆ Huzurbazar, V.S. (1956), “Sufficient Statistics and Orthogonal Parameters”, *Sankhya*, **17**, 217-220.
- ◆ Jeffreys, H. (1946), “An Invariant Form for the Prior Probability in Estimation Problems”, *Proc. Roy. Soc. (London), Ser. A*, **186**, 453-461.
- ◆ Jeffreys, H. (1961), *Theory of Probability*, 3<sup>rd</sup>. edition, Oxford University Press, London.



- ◆ Kass, R.E. (1990), “Data-Translated Likelihood and Jeffrey’s rule”, *Biometrika*, **77**, 107-114.
  
- ◆ Kass, R.E. and Wasserman, L. (1995,a), “A Reference Bayesian Test for Nested Hypotheses and its Relationship to the Schwarz criterion”, *J. Amer. Statist. Assoc.*, **90**, 928-934.
  
- ◆ Lindley, D.V. (1958), “Fiducial Distributions and Baye’s Theorem”, *J.R. Statist. Soc., Ser. B*, **20**, 102-107.
  
- ◆ Nicolau, A. (1993), “Bayesian Intervals with Good Frequentist Behaviour in the Presence of Nuisance Parameters”, *J.R. Statist. Soc., Ser. B*, **55**, 377-390.
  
- ◆ Peers, H.W. (1965), “On Confidence Points and Bayesian Probability Points in the Case of Several Parameters”, *J.R. Statist. Soc., Ser. B*, **27**, 9-16.
  
- ◆ Stein, C. (1956), “Inadmissibility of the Usual Estimator for the Mean of a Multivariate Normal Distribution”, in *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, **1**, 197-206, University of California Press, Berkeley.
  
- ◆ Stein, C. (1959), “An Example of Wide Discrepancy Between Fiducial and Confidence Intervals”, *Ann. Math. Statist.*, **30**, 877-880.
  
- ◆ Welch, B.L. and Peers, H.W. (1963), “On Formulae for Confidence Points Based on Integral of Weithed Likelihoods”, *J.R. Statist. Soc., Ser. B*, **25**, 318-329.
  
- ◆ Ye, K. (1993), “Reference Priors When the Stopping Rule Depends on the Parameter of Interest”, *Journal of the American Statistical Association*, **88**, nº **421**, 360-363.

**CAPÍTULO 3.**

- ◆ Akahira, M. and Takeuchi, K. (1991), “A Definition of Information Amount Applicable to Non-Regular Cases”, *Journal of Computing and Information*, **2**, 71-92.
- ◆ Bardoff-Nielsen, O.E. and Cox, D.R. (1994), *Inference and Asymptotic*. Chapman and Hall, London.
- ◆ Basulto, J. (1997), “Funciones a Priori Imparciales Unidimensionales”, *Estadística Española*, **39**, nº **142**, 99-128.
- ◆ Cox, D.R. and Reid, N. (1987), “Parameter Orthogonality and Approximate Conditional Inference (with Discussion)”, *J.R. Statist. Soc., Ser. B*, **49**, 1-39.
- ◆ Fisher, R.A. (1956), *Statistical Methods and Scientific Inference*. 3<sup>rd</sup>. Edición. Oliver and Boyd .Edinburgh.
- ◆ George, E.I. and McCulloch, R. (1993), “On obtaining Prior Distributions”, *Journal of Statistical Planning and Inference*”, **37**, 169-179.
- ◆ Ghosal, S. and Samanta, T. (1997), “Asymptotic Expansions of Posterior Distributions in Non-Regular Cases”, *Ann. Inst. Statist. Math.*, **49**, nº **1**, 181-197.
- ◆ Ghosal, S., Ghosh, J.K. and Samanta, T. (1995), “On Convergence of Posterior Distributions”, *Ann. Statist.*, **23**, 2145-2152.
- ◆ Gosh, J.K., Ghosal, S. and Samanta, T. (1994), “Stability and Convergence of Posterior in Non-Regular Problems”, *Statistical Decision Theory and Related Topics V*, (eds. S.S. Gupta and J.O. Berger), 183-199, Springer, New-York.

- ◆ Huzurbazar, V.S. (1950), “Probability Distributions and Orthogonal Parameters”, *Proc. Camb. Phil. Soc.*, **46**, 281-284.
  
- ◆ Jeffreys, H. (1946), “An Invariant Form for the Prior Probability in Estimation Problems”, *Proc. Roy. Soc. (London), Ser. A*, **186**, 453-461.
  
- ◆ Jeffreys, H. (1961), *Theory of Probability*, 3<sup>rd</sup>. edition, Oxford University Press, London.
  
- ◆ Kappenman, R.F. (1975), “Conditional Confidence Intervals for Double Exponential Distributions Parameters”, *Technometrics*, **17**, 233-235.
  
- ◆ Kass, R.E. (1989), “The Geometry of Asymptotic Inference”, *Statistical Science*, **4**, n° 3, 188-234.
  
- ◆ Kosmas, K.F. (1990), “Shortest Confidence Intervals for Families of Distributions Involving Truncation Parameters”, *The American Statistician*, **44**, 167-168.
  
- ◆ Lehman, E.L. (1991), *Theory of Point Estimation*, Wadsworth & Brooks/Cole Advanced Books & Software. Pacific Grove, California.
  
- ◆ Lehmann, E.L. (1999), *Elements of Large-Sample Theory*, Springer-Verlag, New-York.
  
- ◆ Nicolau, A. (1993), “Bayesian Intervals with Good Frequentist Behaviour in the Presence of Nuisance Parameters”, *J.R. Statist. Soc., Ser. B*, **55**, 377-390.
  
- ◆ Tibshirani, R. (1989), “Noninformative Priors for One Parameter of Many”, *Biometrika*, **76**, 604-608.

## **Bibliografía.**

---

◆ Welsh, A.H. (1996), *Aspects of Statistical Inference*, John Wiley & Sons, New-York.

## **CAPÍTULO 4.**

◆ Gelfand, A.E. and Smith, A.F. (1990), “Sampling-Based Approaches to Calculating Marginal Densities”, *J.Amer. Statist. Assoc.*, **85**, 398-409.

◆ Geman, S. And Geman, D. (1984), “Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **6**, 721-741.

◆ Lawless, J.F. (1982), *Statistical Models and Methods for Lifetime Data*, John Wiley & Sons, New York.

◆ Smith, A.F. and Gelfand, A.E. (1992), “Bayesian Statistics Without Tears: A Sampling-Resampling Perspective”, *The American Statistician*, **46**, n°2, 84-88.

◆ Welsh, A.H. (1996), *Aspects of Statistical Inference*, John Wiley & Sons, New-York.

## **CAPÍTULO 5.**

◆ Brookes, B.C. (1974), “Obsolescence of special library periodical: sampling errors utility contours”, *Journal of the American Society for Information Science*, **21**, n° 5, 320-329.

- ◆ Callon, M., Courtial, J.P. and Penan, H. (1995), *Cienciometría. La Medición de la Actividad Científica: de la Bibliometría a la Vigilancia Tecnológica*. Ediciones Trea, Gijón.
- ◆ Glänzel, W. And Shoepflin, U. (1994), “A Stochastic Model for the Ageing of Scientific Literature”, *Scientometrics*, **30**, n° **1**, 94-108.
- ◆ Glänzel, W. And Shoepflin, U. (1995), “A Bibliometric Study on Ageing and Reception Processes of Scientific Literature”, *Journal of Information Science*, **21**, 37-53
- ◆ Gupta, B.M. (1997), “Análisis of Distribution of the Age of Citations in Theoretical Population Genetics”, *Scientometrics*, **40**, n° **1**, 139-162.
- ◆ Gupta, B.M. (1998), “Growth and Obsolescence of Literature in Theoretical Population Genetics”, *Scientometrics*, **42**, n° **3**, 335-347.
- ◆ Lawless, J.F. (1982), *Statistical Models and Methods for Lifetime Data*, John Wiley & Sons, New York.
- ◆ Matricciani, E. (1994), “Shannon’s Entropy as a Measure of the “Life” of the Literature of a Discipline”, *Scientometrics*, **30**, n° **1**, 129-145.
- ◆ Moed, H.F., Van Leeuwen, T.N. and Reedijk, J. (1999), “Towards Appropriate Indicators of Journal Impact”, *Scientometrics*, **46**, n°**3**, 575-589.
- ◆ Nelson, W.B. (1972), “Theory and Applications of Hazard Plotting for Censored Failure Data”, *Technometrics*, **14**, 945-965.

## **Bibliografía.**

---

- ◆ Price, D. (1986), *Little Science, Big Science... and Beyond*, Columbia University Press, New York.
  
- ◆ Ruiz, R. y Jiménez, E. (1996), “Envejecimiento de la Literatura Científica en Documentación. Influencia del Origen Nacional de las Revistas. Estudio de una Muestra”, *Revista Española de Documentación Científica*, **19**, n° 1, 39-49.
  
- ◆ Stigler, S.M. (1994), “Citation Patterns in the Journals of Statistics and Probability”, *Statistical Science*, **9**, n° 1, 94-108.