

MIXTURAS DE DISTRIBUCIONES:  
MODELIZACIÓN DE EXPERIENCIAS  
CON ASIMETRÍA EN LOS DATOS

MARÍA NIEVES ATIENZA MARTÍNEZ

UNIVERSIDAD DE SEVILLA  
Instituto de Dpto. de Estadística e I.O.  
de la Fa. de Matemáticas  
de Sevilla, inscrita desde el día 13.2.03  
hasta el día 5.3.03  
Código 6 de matrícula de 2003  
EL DIRECTOR



Sevilla  
En el día del mes de 2003  
*Rosa María*

UNIVERSIDAD DE SEVILLA

MIXTURAS DE DISTRIBUCIONES:  
MODELIZACIÓN DE EXPERIENCIAS  
CON ASIMETRÍA EN LOS DATOS

MEMORIA DIRIGIDA POR  
PROF. DR. D. JOAQUÍN GARCÍA DE LAS HERAS  
PROF. DR. D. JUAN MUÑOZ PICHARDO

MEMORIA PRESENTADA POR MARÍA  
NIEVES ATIENZA MARTÍNEZ PARA  
OPTAR AL GRADO DE DOCTOR EN  
MATEMÁTICAS

VºBº DIRECTOR




FDO. JOAQUÍN GARCÍA DE LAS HERAS

FDO.:



MARÍA NIEVES ATIENZA MARTÍNEZ

VºBº DIRECTOR



FDO. JUAN MUÑOZ PICHARDO

SEVILLA, ABRIL DE 2003

043  
395

*A Rafa y a mis padres*

# Índice

Introducción	1
<b>1 Conceptos generales</b>	<b>5</b>
1.1 Introducción	5
1.2 Distribuciones Gamma, Lognormal y Weibull	6
1.2.1 Distribución Lognormal	6
1.2.2 Distribución Gamma	7
1.2.3 Distribución Weibull	8
1.3 Distribuciones mixtura finitas	11
1.4 Estimación de los parámetros	14
1.5 Formulación del algoritmo EM	20
1.6 Algoritmo EM aplicado a las mixturas finitas	24
<b>2 Mixturas finitas: Identificabilidad</b>	<b>29</b>
2.1 Introducción	29
2.2 Formulación del problema	30
2.3 Identificabilidad	31
2.4 Resultados previos	32
2.5 Condición suficiente de identificabilidad	36
2.6 Identificabilidad de algunas familias	39
2.6.1 Identificabilidad de mixturas de distribuciones de las familias Log-gamma y Log-gamma inversa	39
2.6.2 Identificabilidad de mixturas de distribuciones Gamma, Lognormal y Weibull	42
<b>3 Mixturas finitas: Teoremas de Consistencia</b>	<b>45</b>
3.1 Introducción	45
3.2 Resultados de integrabilidad	47
3.3 Consistencia para familias exponenciales	53
3.4 Consistencia para uniones de familias exponenciales	61
3.5 Consistencia para familias tipo $\mathcal{W}$	65
3.6 Consistencia para uniones de familias tipo $\mathcal{W}$	70
3.7 Consistencia para Gamma, Lognormal, Weibull	73

3.7.1	Condiciones de consistencia $C1$ , $C3$ y $C4$ . . . . .	73
3.7.2	Condición de consistencia $C2$ . . . . .	76
4	Estimación de parámetros. Algoritmo EM . . . . .	85
4.1	Introducción . . . . .	85
4.2	Aplicación del algoritmo EM . . . . .	86
4.3	Elección de los valores iniciales . . . . .	90
4.3.1	Valores iniciales para los pesos . . . . .	91
4.3.2	Valores iniciales de los parámetros de las funciones componentes . . . . .	92
4.4	Criterio de parada . . . . .	93
4.5	Aceleración del algoritmo EM . . . . .	94
4.5.1	Algoritmo del gradiente conjugado generalizado . . . . .	94
4.5.2	Aplicación del algoritmo del gradiente conjugado a la aceleración del algoritmo EM . . . . .	95
4.5.3	Aplicación del algoritmo del gradiente conjugado al modelo de mixturas mixtas $\mathcal{P}$ . . . . .	96
4.6	Simulación . . . . .	99
4.7	Aplicación a la estancia hospitalaria . . . . .	100
	Anexo A. Resultados de la simulación: Gráficos . . . . .	109
	Bibliografía . . . . .	127

# Introducción

Una de las estrategias de la Estadística para describir y explicar una realidad compleja es la representación de la misma mediante modelos probabilísticos. En esta línea se enmarca la presente memoria, teniendo como principal objetivo la modelización de fenómenos aleatorios a través de mixturas finitas de distribuciones. En concreto, aquellas experiencias aleatorias con asimetría en los datos.

Inicialmente, esta memoria surge tratando de abordar un problema real al que se tiene que enfrentar la gestión de hospitales y servicios de salud para programar y financiar adecuadamente los mismos: el estudio del tiempo de hospitalización de los pacientes, variable conocida como ‘estancia hospitalaria’.

La estancia hospitalaria es un índice del consumo de recursos ampliamente utilizado en la gestión hospitalaria. En particular, se utiliza en la dirección, planificación y control de calidad del servicio.

Usualmente los datos observados de esta variable presentan asimetría positiva, por ello, tradicionalmente se han utilizado modelos Lognormal, Gamma y Weibull. En esta perspectiva, Marazzi *y otros* [37], presentan criterios para decidir sobre cuál de los tres modelos anteriores representan mejor la variable citada.

En esta memoria, se pretende ampliar el estudio a través de un modelo más general basado en mixturas finitas de distribuciones, en particular, de las distribuciones antes citadas. El modelo consiste en una mixtura de tres componentes pertenecientes a cada una de las tres familias. Este modelo de ‘mixturas mixtas’ proporciona una flexibilidad extra necesaria cuando una sólo de las familias no conduce a una explicación satisfactoria de la realidad.

Conviene resaltar que, además de la gestión hospitalaria, existen otras muchas situaciones reales descritas por variables que se debaten entre la modelización por alguna de las tres distribuciones: Lognormal, Gamma o Weibull.

Para ilustrar esta aplicación, a continuación se recogen algunas de ellas:

**Microbiología** : Estudio de las posibles consecuencias de la preparación de comidas en ciertos patógenos alimenticios (Whiting [58]).

Neurología : Estudio de la actividad de las neuronas (McKeegan [38]).

Veterinaria : Estudio de la duración de viremias en ganado vacuno (Singer [49]).

Mecánica : Estudios de resistencia de materiales (Ahn [2]).

Meteorología : Concentración de vapor de agua en las nubes (Costa [13]) y en el estudio de pluviometría (Stephenson [50]).

Geología : Estudio de los tamaños de las partículas de sedimentos fluviales (Kondolf [30]).

Geofísica : Estudio de las frecuencias de terremotos de cierta magnitud (Parvez [44]).

Epidemiología : Estimación del periodo de incubación del SIDA (Cooley [12]).

La memoria se ha estructurado de la siguiente forma. En el primer capítulo se recogen definiciones, propiedades y conceptos necesarios para el desarrollo de la misma. Se presentan las distribuciones objeto de estudio con algunas de sus características. También, se introduce la modelización mediante mixturas de distribuciones. Se revisa el método de estimación de máxima verosimilitud, poniendo de manifiesto sus propiedades y recogiendo condiciones suficientes para que las mismas se verifiquen. Se finaliza este capítulo con el estudio de un procedimiento iterativo, ampliamente utilizado en el área de las mixturas, para el cálculo de los estimadores de máxima verosimilitud de los diferentes parámetros: el algoritmo EM.

En el segundo capítulo se estudian condiciones suficientes que permitan comprobar la identificabilidad de experimentos modelizados a través de mixturas finitas de distribuciones. Dado que los resultados existentes sobre identificabilidad de mixturas no son aplicables al modelo bajo estudio en esta memoria, se propone un nuevo resultado que relaja las condiciones de resultados propuestos en la literatura sobre el tema. El capítulo finaliza con la comprobación de la identificabilidad de distintas familias de mixturas, mediante este nuevo resultado.

Para comprobar que en el conjunto de mixturas mixtas de las familias Lognormal, Gamma y Weibull, las soluciones de las ecuaciones de verosimilitud cumplen las condiciones de consistencia (recogidas en el Capítulo 1), en el tercer capítulo se proponen algunos resultados que facilitan la verificación de dichas condiciones para las mixturas de uniones de ciertas familias paramétricas de funciones de densidad como son las familias exponenciales y otro tipo de familias más generales que incluyen a las distribuciones Weibull.

Por último, el Capítulo 4 aborda el problema de la estimación de máxima verosimilitud de los parámetros del modelo presentado, a través de la aplicación del algoritmo EM. Asimismo, se estudian los elementos necesarios para la aplicación

de dicho algoritmo y un método de aceleración del mismo. A continuación, se presenta un conjunto de simulaciones del modelo con el objetivo de ilustrar la validez de las técnicas propuestas en esta memoria. Este capítulo se concluye con la aplicación a conjuntos de datos reales de la variable estancia hospitalaria.



# Capítulo 1

## Conceptos generales

### 1.1 Introducción

En este capítulo se recogen definiciones y propiedades de algunos conceptos necesarios para el desarrollo de esta memoria. En primer lugar, se describen algunas características de las distribuciones Lognormal, Gamma y Weibull, por el papel fundamental que juegan en este estudio.

La sección tercera está dedicada al estudio de mixturas finitas de distribuciones. La modelización de fenómenos aleatorios mediante mixturas distribucionales está recibiendo cada vez más atención, tanto desde el punto de vista teórico como práctico, gracias a la flexibilidad que éstas proporcionan. Existen muchos campos donde los modelos mixtura se han aplicado, astronomía, biología, genética, medicina, economía, psiquiatría, ingeniería y marketing, entre otros. El secreto de su éxito está en su flexibilidad; un modelo mixtura es capaz de modelar distribuciones bastante complejas a través de la elección apropiada de sus componentes y puede describir situaciones donde una sola familia paramétrica no proporciona un modelo satisfactorio. Este es precisamente el caso de la variable estancia hospitalaria, ya que, debido a sus valores extremos y a su asimetría, no se ha encontrado hasta hoy ninguna familia de distribuciones que la describa adecuadamente. Tradicionalmente, para la descripción de esta variable se han utilizado las familias de distribuciones Lognormal, Gamma y Weibull como se recoge, por ejemplo, en Marazzi y otros [37]. Éstas serán las familias de las que se extraerán las componentes del modelo mixtura que se utiliza en esta memoria.

En la sección cuarta se revisa el método de máxima verosimilitud para la estimación de parámetros en una mixtura distribucional, poniendo de manifiesto sus buenas propiedades y recogiendo condiciones suficientes para que se cumplan.

Se finaliza este capítulo con el estudio de un procedimiento iterativo, ampliamente utilizado en el área de las mixturas, para el cálculo de los estimadores de

los diferentes parámetros: el algoritmo EM.

## 1.2 Distribuciones Gamma, Lognormal y Weibull

Las distribuciones Lognormal, Gamma y Weibull son de gran utilidad en estudios y experiencias prácticas en las que se deben aplicar técnicas de análisis de supervivencia o que tienen como objetivo la modelización de tiempos de respuesta. Así, por ejemplo, datos de supervivencia en investigaciones de salud pública, duración de epidemias de virus, respuestas a dosis farmacológicas, incubación de virus o datos de estancias hospitalarias suelen ser modelados con alguna de las distribuciones anteriormente citadas.

A continuación, se presentan sus definiciones junto con algunas características de las mismas, entre ellas la función generatriz de momentos de su logaritmo neperiano, necesaria en capítulos posteriores. Hay que hacer constar que si bien las dos primeras pertenecen a la familia exponencial paramétrica, la tercera no.

### 1.2.1 Distribución Lognormal

Históricamente, estas distribuciones aparecen por vez primera cuando Galton [16] en 1879 estudió medias geométricas de variables aleatorias cuyos logaritmos se distribuían siguiendo una ley normal. Desde entonces esta distribución ha demostrado ser útil para modelizar diversas situaciones tales como la concentración de contaminación en el aire, la distribución de ingresos y riquezas, el periodo de incubación de ciertas enfermedades o de recuperación de intervenciones quirúrgicas, el tiempo de supervivencia de pacientes con ciertos tipos de cáncer... y, en muchas ocasiones, en situaciones donde se necesita un modelo con asimetría positiva; particularmente, como modelo de la variable estancia hospitalaria. En Johnson and Kotz [25] y Bain [5] se recoge una extensa relación de trabajos aplicados en los que aparece dicha distribución.

**Definición 1.2.1.** *Una variable aleatoria,  $X$ , se distribuye según una ley Lognormal de parámetros  $\mu$  y  $\sigma$ , lo que se denota por  $X \sim LN(\mu, \sigma)$ , si su función de densidad viene dada por:*

$$f(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma x}} \exp \left\{ -(\log x - \mu)^2 / 2\sigma^2 \right\}, \quad x > 0$$

para  $\mu \in \mathbb{R}$ ,  $\sigma > 0$ . Generalmente,  $\sigma$  es conocido como parámetro de forma y  $e^\mu$  como parámetro de escala.

A continuación, se recogen algunas características básicas de esta distribución.

1.- El momento de orden  $r$  viene dado por:

$$\mathbb{E}[X^r] = \exp\left\{\mu r + \frac{r\sigma^2}{2}\right\}, \quad r > 0.$$

2.- La función generatriz de momentos del  $\log X$ , el cual sigue una distribución normal de parámetros  $\mu$  y  $\sigma$ , viene dada por la expresión:

$$\Phi_{\mu\sigma}(t) = \exp\left\{t\mu + \frac{t^2\sigma^2}{2}\right\}, \quad t \in (-\infty, +\infty).$$

3.- Los estimadores de máxima verosimilitud de los parámetros  $\mu$  y  $\sigma$  son:

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n \log x_i \quad \hat{\sigma} = \sqrt{\frac{1}{n} \sum_{i=1}^n (\log x_i - \hat{\mu})^2}.$$

### 1.2.2 Distribución Gamma

La familia de distribuciones Gamma fue introducida en el siglo XIX, estando estrechamente ligada a los estudios sobre las variables aleatorias normales. Por ejemplo, en Laplace [33] y Bienaymé [8] aparece la distribución Chi-cuadrado perteneciente a la familia de distribuciones Gamma. Estas distribuciones se han usado para la aproximación de formas cuadráticas en las técnicas de modelos lineales y análisis de la varianza. Además, permiten representaciones útiles de muchas situaciones físicas en trabajos aplicados en áreas tales como la meteorología, economía, teoría de inventarios... También se han utilizado para ajustar a distribuciones exponenciales negativas (que son un caso particular de ellas) la variable tiempo de vida. El hecho de que la suma de exponenciales negativas independientes siga una distribución gamma, hace que ésta aparezca en la teoría de contadores aleatorios y en otros temas asociados con procesos aleatorios en el tiempo, por ejemplo, procesos que tratan de modelizar las precipitaciones meteorológicas.

**Definición 1.2.2.** Una variable aleatoria,  $X$ , se distribuye según una ley Gamma de parámetros  $a$  y  $b$ , lo que se denota por  $X \sim G(a, b)$ , si su función de densidad es de la forma:

$$f(x; a, b) = \frac{b^{-a}}{\Gamma(a)} x^{a-1} e^{-x/b}, \quad x > 0$$

para  $a, b > 0$  y donde  $\Gamma(a) = \int_0^\infty t^{a-1} e^{-t} dt$  es la función gamma. Generalmente  $a$  es conocido como parámetro de forma y  $b$  como parámetro de escala.

A continuación se recogen algunas propiedades de interés de esta distribución.

1.- La distribución Gamma se reduce a la distribución Exponencial negativa si  $a = 1$  y a la distribución Chi-cuadrado con  $\nu$  grados de libertad cuando  $b = 2$  y  $a = \nu/2$ .

2.- Momento de orden  $r$

$$\mathbb{E}[X^r] = \frac{\Gamma(a+r)b^r}{\Gamma(a)}, \quad r > -a.$$

3.- La función generatriz de momentos del  $\log X$ , viene dada por la expresión

$$\Phi_{ab}(t) = \frac{b^{-a}\Gamma(t+a)}{\Gamma(a)}, \quad t \in (-a, +\infty).$$

4.- Los estimadores de máxima verosimilitud de la distribución Gamma son las soluciones del sistema de ecuaciones:

$$\begin{cases} \hat{b} = \frac{1}{\hat{a}n} \sum_{i=1}^n x_i \\ \log \frac{1}{n} \sum_{i=1}^n x_i - \log \prod_{i=1}^n x_i^{\frac{1}{n}} = \log \hat{a} - \text{Poly}[\hat{a}] \end{cases}$$

donde  $\text{Poly}[a] = \frac{\Gamma'(a)}{\Gamma(a)}$  (con  $\Gamma'$  la derivada de  $\Gamma$ ).

Dada la forma de la segunda ecuación, no ha sido posible obtener una fórmula explícita para calcular  $\hat{a}$ , por lo que diferentes autores han dado metodos aproximativos. Uno de estos procedimientos es el dado por Greenwood and Durand [17], que consiste en aproximar  $\hat{a}$  mediante la siguiente función  $s(M)$ :

$$s(M) = \begin{cases} \frac{0.5000876 + 0.1648852M - 0.0544276M^2}{M} & 0 \leq M \leq 0.5772 \\ \frac{8.898919 + 9.059950M + 0.9775373M^2}{M(17.79728 + 11.968477M + M^2)} & 0.5772 \leq M \leq 17 \\ \frac{1}{M} & M \geq 17 \end{cases}$$

siendo  $M = \log(\bar{x}/\tilde{x})$ , donde  $\bar{x}$  y  $\tilde{x}$  son la media aritmética y geométrica, respectivamente.

### 1.2.3 Distribución Weibull

En el año 1939 el físico Waloddi Weibull utilizó esta distribución para modelar el comportamiento de la fuerza de resistencia de materiales. Posteriormente, volvió

a usarla en una amplia gama de aplicaciones, entre las que se incluyen estudios de las estaturas de hombres adultos nacidos en las islas británicas. En Berrettoni [7], se recogen diversas aplicaciones de la distribución Weibull usando en la mayoría de los casos métodos gráficos.

La distribución Weibull incluye también a la distribución Exponencial negativa como un caso particular y, algunas veces, es considerada como una generalización de ella que proporciona una flexibilidad adicional necesaria para hacer que un modelo se ajuste de forma más exacta.

Si bien, el modelo esponencial tuvo un gran desarrollo y aplicabilidad en los años 50 y 60 del siglo XX, el modelo Weibull empezó a ser un serio competidor de éste, especialmente en problemas en los que la variable respuesta es el tiempo de fallo. Kao [26] y Rider [47] estudiaron mixturas de la distribución Weibull. Debido a su versatilidad en el ajuste a distribuciones de tiempo de vida en una extensa variedad de mecanismos complejos, la distribución Weibull juega un papel muy importante en el campo de la fiabilidad y del análisis de supervivencia.

**Definición 1.2.3.** Una variable aleatoria,  $X$ , se distribuye según una ley Weibull de parámetros  $c$  y  $d$ , lo que se denota por  $X \sim W(c, d)$  si su función de densidad es de la forma:

$$f(x; c, d) = \frac{c}{d^c} x^{c-1} e^{-x^c/d^c}, \quad x > 0,$$

para  $c, d > 0$ . Generalmente  $c$  es conocido como parámetro de forma y  $d$  como parámetro de escala.

A continuación se presentan algunas características de interés de esta distribución.

1.- En el caso particular  $c = 1$ , se tiene la distribución Exponencial negativa.

2.- Momento de orden  $r$

$$\mathbb{E}[X^r] = d^r \Gamma\left(1 + \frac{r}{c}\right), \quad r > 0.$$

3.- La función generatriz de momentos del log  $X$ , viene dada por la expresión:

$$\Phi_{cd}(t) = d^{-t} \Gamma\left(\frac{t}{c} + 1\right), \quad t \in (-c, +\infty).$$

Un inconveniente que presenta esta distribución es la dificultad en la estimación de sus parámetros, ya que los cálculos que envuelve no son siempre fáciles. Son muchos los autores que han escrito sobre este problema, así Leone y otros [35], Cohen [11] y Harter y Moore [19] obtuvieron los estimadores de máxima verosimilitud, a través de métodos iterativos; Sarhan y Greenberg [48] proponen unos estimadores insesgados y algunos estimadores simplificados, no obstante estos estimadores requieren cálculos muy complejos. Gumbel [18] dio estimadores más simples que posteriormente fueron modificados por Miller y Freud [43].

Dado que en esta memoria, los estimadores de los parámetros de la distribución Weibull se utilizarán para la obtención de unos valores iniciales en la aplicación del algoritmo EM, lo que significa que no serán las estimaciones definitivas de los parámetros, se ha optado por el método de estimación propuesto por Bain y Antle [5], ya que proporciona fórmulas explícitas y, por lo tanto, no necesita la aplicación de ningún método iterativo.

Un interesante y exhaustivo trabajo comparativo de los diferentes métodos propuestos es el realizado por Al-Baidhani y Sinclair [3], en el que están incluidos los propuestos por Bain y Antle [5].

A continuación, se recoge el método de estimación general que proponen estos últimos.

### Método de estimación de Bain y Antle

Sea  $x_1, \dots, x_n$  una m.a.s. extraída de una población con función de densidad  $f(x; \theta)$ .

Sea  $u(x; \theta)$  una función elegida arbitrariamente tal que verifica:

1. La distribución de  $u$  es independiente de  $\theta$ .
2. El orden de los  $u_i = u(x_i; \theta)$  se puede determinar a partir de los valores ordenados de  $x_i$ .

Sea  $x_{(1)} < \dots < x_{(n)}$  la muestra ordenada y sean  $u_{(1)} < u_{(2)} < \dots < u_{(n)}$  los valores  $u(x_i; \theta)$  ordenados. En estas circunstancias, un estimador para  $\theta$  es el valor  $\hat{\theta}$  que maximiza el *acuerdo*, según algún criterio, entre  $u_{(i)}$  y  $\mathbb{E}[u_{(i)}]$ .

Al igual que la elección de la función  $u(x; \theta)$ , la elección del criterio también es arbitraria y puede resultar difícil determinar el adecuado.

En el caso particular de la distribución Weibull proponen la función:

$$u(x; c, d) = \left(\frac{x}{c}\right)^d,$$

cuya función de densidad es de la forma:

$$f(u) = \exp[-u] \quad u \geq 0,$$

que no depende ni de  $c$  ni de  $d$ . Y como criterio de *acuerdo* entre  $u_{(i)}$  y  $\mathbb{E}[u_{(i)}]$  proponen considerar el cuadrado de la desviación entre  $\log(u_{(i)})$  y  $\log \mathbb{E}[u_{(i)}]$ . Así, el estimador será el valor que minimice:

$$\sum_{i=1}^n \left[ \log(x_{(i)}/c)^d - \log \mathbb{E}[u_{(i)}] \right]^2.$$

Utilizan la transformación logarítmica ya que el criterio usual de mínimos cuadrados no proporciona una forma explícita para  $d$ .

De acuerdo con el criterio anterior y denotando por  $\varepsilon_r$

$$\varepsilon_r = \mathbb{E}[u_{(r)}] = \sum_{i=1}^r \frac{1}{n-i+1},$$

los estimadores de los parámetros de la distribución Weibull vienen dados por:

$$\hat{c} = \frac{\prod_{i=1}^n x_{(i)}^{\frac{1}{n}}}{\prod_{i=1}^n \varepsilon_i^{\frac{1}{nd}}} \quad \text{y} \quad \hat{d} = \frac{\sum_{i=1}^n \log \varepsilon_i \log x_{(i)} - \frac{1}{n} \sum_{i=1}^n \log \varepsilon_i \sum_{i=1}^n \log x_{(i)}}{\sum_{i=1}^n \log^2 x_{(i)} - \frac{1}{n} \left( \sum_{i=1}^n \log x_{(i)} \right)^2}.$$

### 1.3 Distribuciones mixtura finitas

Como ya se ha comentado, cada una de las familias de distribuciones de la sección anterior han demostrado una gran utilidad en la modelización de datos de tiempo de vida o respuesta. A continuación, se plantea el estudio de mixturas de dichas familias, las cuales podrán permitir una mejor modelización de estos experimentos. En particular, en esta memoria se plantea la modelización de fenómenos aleatorios con asimetría positiva en los datos a través de una distribución mixtura formada por la combinación lineal convexa de tres funciones pertenecientes a cada una de las tres familias anteriores.

En esta sección, se recogen algunas nociones básicas referentes a mixturas finitas, comenzando con la definición de densidad mixtura.

**Definición 1.3.1.** Sea  $\{f(\cdot; \theta) : \theta \in \Theta\}$  una familia de densidades de probabilidad parametrizada por  $\Theta$ . A toda función de la forma

$$q(x; \Psi) = \int_{\Theta} f(x; \theta) dG(\theta)$$

donde  $G$  es una distribución de probabilidad sobre  $\Theta$ , se le denomina densidad mixtura, y a  $G$  distribución mixing.

Esta memoria se centra en el caso particular en que  $G$  es discreta y asigna probabilidad positiva a sólo un conjunto finito de puntos, en cuyo caso se puede sustituir la integral por una suma y, así, obtener la forma general de una mixtura finita. Una definición formal sería:

**Definición 1.3.2.** Sea  $\mathcal{F}$  una familia de densidades de probabilidad. A toda función de la forma

$$q = \pi_1 f_1 + \cdots + \pi_k f_k$$

donde

$$\{\pi_i\}_{i=1}^k \in \mathcal{C} = \left\{ \{\pi_i\}_{i=1}^k / h \in \mathbb{N}; \pi_i > 0, \forall i = 1, \dots, k; \sum_{i=1}^k \pi_i = h \right\}$$

y  $f_1, \dots, f_k \in \mathcal{F}$ , se le denomina densidad mixtura finita de elementos de  $\mathcal{F}$ .

Los parámetros  $\pi_1, \dots, \pi_k$  se denominan pesos de la mixtura y  $f_1, \dots, f_k$  densidades componentes de la mixtura.

Es decir, una mixtura finita se puede ver como una combinación convexa de funciones. A continuación, se recogen dos formas diferentes de interpretar un problema de modelos mixtura finitos, tal y como vienen recogidos en Titterington [52].

*Aplicación directa.* El problema de modelos de mixturas finitas se interpreta como un problema de aplicación directa cuando se supone que existen  $k$  categorías subyacentes o fuentes diferentes a las que podría pertenecer cada una de las observaciones realizadas de la variable aleatoria  $X$ , sin que se llegue a conocer la fuente de la que proviene. En este caso,  $f_j$  denota la función de densidad de  $X$  condicionada a que la observación fuese de la categoría  $j$ , y  $\pi_j$  denotará la probabilidad de que la observación obtenida fuese de dicha fuente.

*Aplicación indirecta.* Se puede interpretar como un problema de aplicación indirecta aquellas situaciones en las que se recurre al modelo de mixtura finita simplemente como una herramienta matemática para proporcionar un medio indirecto de obtener una forma de análisis flexible y más tratable, es decir, donde las categorías subyacentes no tienen necesariamente una interpretación física directa. En principio, el estudio que se realiza en esta memoria se encuadra en este segundo enfoque, sin descartar la posibilidad de que, una vez conseguido el modelo, las componentes resultantes puedan identificarse con algunas subpoblaciones determinadas.

Dependiendo de la forma en que se han obtenido los datos, se pueden originar varios tipos de muestras. En algunas aplicaciones de las mixturas finitas, como se puede ver en Titterington [52], aparecen muestras aleatorias resultantes de formas de muestreo que mezclan observaciones de las que se conoce la categoría subyacente a la que pertenecen (estas se llaman etiquetadas o categorizadas) con otras de las que no (no etiquetadas). Hosmer [21] hace una clasificación de estas formas de muestreo.

Siguiendo la notación de Hosmer, el caso que centra el interés de esta memoria se corresponde con el modelo que él denota por M0, este modelo es el más tratado en la literatura clásica de mixturas. En este caso, la muestra está compuesta en su totalidad por  $n$  observaciones no etiquetadas en la mixtura. Además, en esta situación,  $f_1, \dots, f_k$  tienen formas paramétricas específicas y la mixtura finita tiene



una representación más explícita que viene dada por:

$$q(\mathbf{x}; \Psi) = \sum_{j=1}^k \pi_j f_j(\mathbf{x}; \theta_j),$$

donde  $\theta_j$  denota los parámetros correspondientes a la función densidad componente  $f_j$  y  $\Psi = (\pi_1, \dots, \pi_{k-1}, \theta_1, \dots, \theta_k)$  denota la colección completa de parámetros que intervienen en el modelo mixtura. Obviamente, uno de los problemas principales que deben ser abordados es la estimación de dichos parámetros.

Previo a la estimación de los parámetros, se debe abordar el problema de la identificabilidad de la distribución mixtura. La identificabilidad se refiere a la cuestión teórica de si es posible estimar los parámetros de un modelo de manera única. En general, una familia paramétrica de funciones de distribución es identificable si distintos valores de los parámetros determinan distintos miembros de la familia.

Para dar una definición formal de identificabilidad, en el caso de mixturas finitas, se considera  $\mathcal{T} = \{F(\cdot, \theta), \theta \in \Theta \subseteq \mathbb{R}^d, \mathbf{x} \in \mathbb{R}^p\}$ , una clase de funciones de distribución  $p$ -dimensionales. La clase de mixturas finitas de  $\mathcal{T}$  es la clase de funciones de distribución,  $\mathcal{H}$ , definida por:

$$\begin{aligned} \mathcal{H} &= \left\{ H / \exists \{\pi_i\}_{i=1}^k \in \mathcal{C}, \exists F_1, \dots, F_k \in \mathcal{T} : H(\cdot) = \sum_{i=1}^k \pi_i F_i(\cdot) \right\} \\ &= \left\{ H / \exists \{\pi_i\}_{i=1}^k \in \mathcal{C}, \exists \theta_1, \dots, \theta_k \in \Theta : H(\cdot) = \sum_{i=1}^k \pi_i F(\cdot; \theta_i) \right\}, \end{aligned}$$

es decir,  $\mathcal{H}$  es el casco convexo de  $\mathcal{T}$ .

En esta situación y asumiendo que cualquier  $H \in \mathcal{H}$  viene representada de la forma:

$$H = \sum_{j=1}^k \pi_j F_j \tag{1.1}$$

donde  $F_1, F_2, \dots, F_k$  son miembros distintos de  $\mathcal{T}$ , se dirá que  $\mathcal{H}$  es identificable si y sólo si todos los miembros de  $\mathcal{H}$  tienen una única representación de la forma recogida en (1.1). Formalmente:

**Definición 1.3.3.** *Sea  $\mathcal{T}$  una familia de distribuciones  $p$ -dimensionales y  $\mathcal{H}$  la clase de mixturas finitas de  $\mathcal{T}$ . Se dice que  $\mathcal{H}$  es identificable si para cualesquiera  $H, H' \in \mathcal{H}$  tales que*

$$H = \sum_{j=1}^k \pi_j F_j \quad \text{y} \quad H' = \sum_{j=1}^{k'} \pi'_j F'_j$$

se verifica que  $H = H'$  si  $k = k'$  y se pueden ordenar las sumas de forma que  $\pi_j = \pi'_j, F_j = F'_j, j = 1, \dots, k$ .

## 1.4 Estimación de los parámetros

La utilización de las mezclas finitas, como modelización de fenómenos reales bajo estudio, conduce a la necesidad de seleccionar el elemento de la clase completa de las mezclas que mejor refleje la realidad bajo algún criterio de optimalidad. Es necesario, por tanto, estimar los parámetros que determinan unívocamente, en caso de identificabilidad, la mezcla adecuada partiendo del conocimiento de una muestra aleatoria. Así pues, supuesto que  $\Psi^* \in \Omega$  es el valor del parámetro desconocido, se plantea el problema de cómo estimarlo.

Si bien existe una amplia variedad de métodos de estimación aplicados al problema de mezclas finitas (método de los momentos, de mínima chi-cuadrado, de mínimos cuadrados, métodos gráficos...), tal y como aparece recogido en Titterton(1992), el de máxima verosimilitud es el más usado por presentar propiedades estadísticas deseables, entre ellas su teoría asintótica. En concreto, se trata de obtener para cada realización muestral  $x = (x_1, \dots, x_n)$  el valor  $\Psi \in \Omega$  que maximiza la función de verosimilitud, que en el caso bajo estudio, viene dada por la expresión:

$$L(\Psi) = \prod_{i=1}^n \sum_{j=1}^k \pi_j f_j(x_i; \theta_j)$$

o, equivalentemente, que maximiza su logaritmo:

$$\mathcal{L}(\Psi) = \sum_{j=1}^n \log \sum_{j=1}^k \pi_j f_j(x_i; \theta_j).$$

Hay que hacer constar que este método conlleva dos problemáticas asociadas. La primera radica en que esta función no tiene porqué estar acotada superiormente en  $\Omega$ , tal es el caso del ejemplo recogido por Kiefer y Wolfowitz [28] con la mezcla de dos variables normales.

La segunda problemática que puede presentar es que, a menudo, en un modelo de mezclas la verosimilitud alcanza su máximo para distintos valores de  $\Psi$ . En este sentido, esta segunda dificultad es motivo de interés o no dependiendo de si los parámetros de cada densidad componente son de interés o, por el contrario, se desea sólo una aproximación de la densidad mezcla. Conviene hacer notar la existencia de un problema relacionado con el segundo aspecto citado: la verosimilitud puede tener máximos locales.

Con respecto a la primera problemática, si la función de verosimilitud es diferenciable (como ocurre en las mezclas objeto de la presente memoria), la maximización de la misma se aborda a través de la resolución de las llamadas ecuaciones normales, resultantes de igualar a cero las derivadas de la función de verosimilitud:

$$\frac{\partial L(\Psi)}{\partial \Psi} = 0.$$

A continuación, se recogen dos teoremas relacionados con la bondad de los estimadores obtenidos por este método. El primero de los teoremas establece las condiciones para que en un entorno del valor del parámetro haya una única solución de las ecuaciones normales fuertemente consistente y para que, además, esta solución maximice al menos localmente la función log-verosimilitud y se distribuya asintóticamente según una ley normal. Este teorema es un compendio de resultados que generalizan el trabajo inicial de Cramér [14] referente a la existencia, consistencia y normalidad asintótica del estimador de máxima verosimilitud de un sólo parámetro escalar. Las condiciones, en las cuáles se basan los siguientes resultados, fueron esencialmente dadas por Chanda [9] como generalizaciones multidimensionales de las de Cramér [14]. Con ellas, Chanda [9] establece que existe una única solución de las ecuaciones de verosimilitud consistente y da su conducta asintótica normal. Usando estas mismas condiciones, Peters and Walker [45] demostraron que hay una única solución fuertemente consistente y que, al menos localmente, maximiza la función log-verosimilitud.

Previo a recoger los teoremas, se presenta la notación y las condiciones mencionadas, tal y como se recogen en Peters and Walker [45], que serán referidas en este y otros capítulos como *Condiciones de consistencia* C1-C2-C3-C4.

Sea  $\mathbf{x} = (x_1, \dots, x_n)$ , una muestra del modelo mixtura dado por:

$$q(\mathbf{x}, \Psi) = \sum_{j=1}^k \pi_j f_j(x_i, \theta_j),$$

donde  $\{\pi_j\} \in \mathcal{C}$  y cada componente de la mixtura  $f_j$  es diferenciable respecto de sus parámetros  $\theta_j$ . En adelante se denotará por

$$\mathcal{C}_k = \left\{ (\pi_1, \dots, \pi_{k-1}) : \sum_{j=1}^{k-1} \pi_j < 1, \pi_j > 0 \quad \forall j = 1, \dots, k-1 \right\}$$

y por  $\Omega$  se considera el espacio paramétrico  $\mathcal{C}_k \times \Theta_1 \times \dots \times \Theta_k$ , dado por:

$$\Omega = \left\{ (\pi_1, \dots, \pi_{k-1}, \theta_1, \dots, \theta_k) : \{\pi_j\}_{j=1}^{k-1} \in \mathcal{C}_k, \quad \theta_j \in \Theta_j \text{ para } j = 1, \dots, k \right\}.$$

Por comodidad, los elementos de  $\Omega$ , para los dos teoremas siguientes, serán denotados por  $\Psi = (\psi_1, \dots, \psi_\nu)$  con  $\nu$  la dimensión de  $\Omega$ .

Para  $\Psi \in \Omega$  y un  $r > 0$  suficientemente pequeño, se denota por  $N_r(\Psi)$  la bola cerrada de centro  $\Psi$  y radio  $r$  en  $\Omega$  y por  $q_r(\mathbf{x}; \Psi)$  y  $q_r^*(\mathbf{x}; \Psi)$  las siguientes funciones:

$$q_r(\mathbf{x}; \Psi) = \sup_{\Psi' \in N_r(\Psi)} q(\mathbf{x}; \Psi') \quad \text{y} \quad q_r^*(\mathbf{x}; \Psi) = \max \{1, q_r(\mathbf{x}; \Psi)\}.$$

Debido a que los resultados de los teoremas son estrictamente locales, no hay pérdida de generalidad en restringir  $\Omega$  a un entorno de  $\Psi^*$ ,  $\hat{\Omega}$ , si fuese necesario.

## Condiciones de consistencia

C1. Para todo  $\Psi \in \Omega$ , c.p.t.  $x \in \mathbb{R}^p$  y para  $i, j, s = 1, \dots, \nu$ , las derivadas parciales,  $\partial q / \partial \psi_j$ ,  $\partial^2 q / \partial \psi_j \partial \psi_i$  y  $\partial^3 q / \partial \psi_j \partial \psi_i \partial \psi_s$ , existen y verifican las desigualdades:

$$\left| \frac{\partial q(x; \Psi)}{\partial \psi_j} \right| \leq h_j(x), \quad \left| \frac{\partial^2 q(x; \Psi)}{\partial \psi_j \partial \psi_i} \right| \leq h_{ij}(x), \quad \left| \frac{\partial^3 \log q(x; \Psi)}{\partial \psi_j \partial \psi_i \partial \psi_s} \right| \leq h_{ijs}(x)$$

donde  $h_j$  y  $h_{ij}$  son integrables y  $h_{ijs}$  satisfacen:

$$\int h_{ijs}(x) q(x; \Psi^*) dx < \infty.$$

C2. La matriz información de Fisher,  $I(\Psi)$ , dada por:

$$I(\Psi) = \int [\nabla_{\Psi} \log q(x; \Psi)] [\nabla_{\Psi} \log q(x; \Psi)]^T q(x; \Psi) dx,$$

existe y es definida positiva en  $\Psi^*$ , donde  $\nabla_{\Psi}$  denota el gradiente de primeras derivadas parciales de la función con respecto a las componentes de  $\Psi$ .

C3. Para cada  $\Psi \in \Omega$  y para  $r > 0$  suficientemente pequeño,

$$\int \log q_r^*(x; \Psi) q(x; \Psi^*) dx < +\infty.$$

C4.

$$\int |\log q(x; \Psi^*)| q(x; \Psi^*) dx < +\infty.$$

**Teorema 1.4.1.** Si se verifican las condiciones de consistencia C1 y C2, entonces:

a) para cualquier entorno suficientemente pequeño de  $\Psi^*$  en  $\Omega$  y para  $n$  suficientemente grande existe, con probabilidad 1, una única solución  $\Psi^n$  de las ecuaciones normales, la cual maximiza localmente su función de verosimilitud.

b)  $\sqrt{n}(\Psi^n - \Psi^*)$  se distribuye asintóticamente según una ley normal con vector media cero y matriz de covarianzas  $I(\Psi^*)^{-1}$ .

Con este teorema se justifica por tanto que, independientemente de que la función de verosimilitud esté acotada o no, existe una única solución de las ecuaciones normales en un entorno de  $\Psi^*$  que es, al menos, un máximo local de dicha función (que se denotará por  $\hat{\Psi}$  y que será tratada como estimador de máxima verosimilitud aun no siendo máximo global) y tiene, bajo ciertas condiciones de regularidad, propiedades de consistencia, eficiencia y normalidad asintótica, lo que la convierte en un buen estimador de  $\Psi^*$ .

No obstante, existen dos cuestiones no resueltas por el teorema anterior acerca de  $\Psi^n$ , la única solución fuertemente consistente de las ecuaciones de verosimilitud. La primera cuestión es si  $\Psi^n$  es realmente un estimador de máxima verosimilitud, es decir, un punto donde la función log-verosimilitud alcanza el mayor máximo local. La segunda cuestión es si, aun siéndolo, existen otros estimadores de máxima verosimilitud que llevan a otras densidades límites distintas a  $q(x; \Psi^*)$ . El teorema siguiente está relacionado con estas cuestiones.

**Teorema 1.4.2.** *Sea  $\Omega' \subset \Omega$  cualquier subconjunto compacto que contiene a  $\Psi^*$  en su interior y sea el conjunto*

$$E = \{ \Psi \in \Omega' : q(x; \Psi) = q(x; \Psi^*) \text{ c.p.t. } x \in \mathbb{R}^p \}.$$

*Si se verifican las condiciones de consistencia C3 y C4, entonces, con probabilidad 1, para cualquier  $D \subset \Omega' \setminus E$ ,*

$$\lim_{n \rightarrow \infty} \sup_{\Psi \in D} \frac{\prod_{i=1}^n q(x_i; \Psi)}{\prod_{i=1}^n q(x_i; \Psi^*)} = 0.$$

Si la familia de densidades mixtura  $q(x; \Psi)$ , con  $\Psi \in \Omega$ , es identificable, el teorema anterior garantiza que para cualquier subconjunto compacto  $\Omega' \subseteq \Omega$  que contenga a  $\Psi^*$  en su interior,  $\Psi^n$  es un estimador de máxima verosimilitud, con probabilidad 1, siempre que el tamaño muestral,  $n$ , sea suficientemente grande. Además, cualquier otro estimador de máxima verosimilitud en  $\Omega'$  se obtiene de  $\Psi^n$  sin más que intercambiar las componentes de la mixtura (Peters and Walker [45]).

Dada la dificultad que desde el punto de vista práctico conlleva la comprobación de las condiciones de consistencia C3 y C4 referidas a la mixtura  $q(x, \Psi)$ , a continuación se proponen dos nuevas condiciones C3' y C4', referidas a las funciones componentes de la mixtura, que facilitan la citada comprobación.

**C3'.** *Para cada  $\theta = (\theta_1, \dots, \theta_k) \in \Theta = \Theta^1 \times \dots \times \Theta^k$  y para  $r > 0$  suficientemente pequeño,*

$$\int (\log f_{ir}^*(x; \theta_i)) f_j(x; \theta_j^*) dx < +\infty \quad \forall i, j = 1, \dots, k,$$

*donde  $f_{ir}^*$  se define de manera análoga a  $q_r$ .*

**C4'.** *Para  $i, j = 1, \dots, k$*

$$\int |\log f_i(x, \theta_i^*)| f_j(x, \theta_j^*) dx < +\infty.$$

Las siguientes proposiciones demuestran que estas dos condiciones son suficientes para que se verifiquen C3 y C4.

**Lema 1.4.1.** *Dados  $a_j \geq 0 \quad \forall j = 1, \dots, k$  y  $\pi_j \in [0, 1]$  tales que  $\sum_{j=1}^k \pi_j = 1$ , se verifica*

$$\left| \log \left( \sum_{j=1}^k \pi_j a_j \right) \right| \leq \sum_{j=1}^k |\log a_j|.$$

*Demostración.* Sean  $\alpha = \min\{a_j : j = 1, \dots, k\}$  y  $\beta = \max\{a_j : j = 1, \dots, k\}$ , se verifica

$$\alpha \leq \sum_{j=1}^k \pi_j a_j \leq \beta$$

y por ser el logaritmo creciente,

$$\log \alpha \leq \log \left( \sum_{j=1}^k \pi_j a_j \right) \leq \log \beta$$

por tanto

$$\left| \log \left( \sum_{j=1}^k \pi_j a_j \right) \right| \leq \max\{|\log \alpha|, |\log \beta|\} \leq \sum_{j=1}^k |\log a_j|.$$

□

**Proposición 1.4.1.** *La condición C3' es suficiente para que se verifique la condición C3.*

*Demostración.* Por la definición de  $q_r^*(x; \Psi)$ ,

$$\begin{aligned} \log q_r^*(x; \Psi) &= \sup_{\Psi' \in N_r(\Psi)} \log \max \left\{ \sum_{i=1}^k \pi'_i, \sum_{i=1}^k \pi'_i f_i(x, \theta'_i) \right\} \leq \\ &\leq \sup_{\Psi' \in N_r(\Psi)} \log \sum_{i=1}^k \pi'_i \max\{1, f_i(x, \theta'_i)\} \end{aligned}$$

que por la aplicación del Lema 1.4.1 es

$$\begin{aligned} &\leq \sup_{(\theta'_1, \dots, \theta'_k) \in N_r(\theta_1, \dots, \theta_k)} \sum_{i=1}^k \log \max\{1, f_i(x, \theta'_i)\} \leq \\ &\leq \sum_{i=1}^k \log \sup_{(\theta'_1, \dots, \theta'_k) \in N_r(\theta_1, \dots, \theta_k)} \max\{1, f_i(x, \theta'_i)\} = \end{aligned}$$

$$= \sum_{i=1}^k \log f_{ir}^*(x, \theta_i).$$

Por tanto:

$$\begin{aligned} \int (\log q_r^*(x; \Psi)) q(x; \Psi^*) dx &\leq \int \left( \sum_{i=1}^k \log f_{ir}^*(x, \theta_i) \right) \sum_{j=1}^k \pi_j^* f_{jr}^*(x, \theta_j^*) dx = \\ &= \sum_{i,j=1}^k \int \pi_j^* (\log f_{ir}^*(x, \theta_i)) f_j(x, \theta_j^*) dx, \end{aligned}$$

lo que demuestra, por la linealidad de la integral, que si  $\mathbf{C3}'$  se verifica entonces se verifica  $\mathbf{C3}$ .  $\square$

**Proposición 1.4.2.** *La condición  $\mathbf{C4}'$  es suficiente para que se verifique la condición  $\mathbf{C4}$ .*

*Demostración.* Aplicando el Lema 1.4.1, se obtiene la desigualdad

$$\begin{aligned} \int |\log q(x; \Psi^*)| q(x; \Psi^*) dx &= \int \left| \log \sum_{i=1}^k \pi_i^* f_i(x, \theta_i^*) \right| \sum_{j=1}^k \pi_j^* f_j(x, \theta_j^*) dx \\ &\leq \int |\log f_i(x, \theta_i^*)| \sum_{i,j=1}^k \pi_j^* f_j(x, \theta_j^*) dx. \end{aligned}$$

Y, por la linealidad de la integral y  $\mathbf{C4}'$ ,

$$\leq \sum_{i,j=1}^k \pi_j^* \int |\log f_i(x, \theta_i^*)| f_j(x, \theta_j^*) dx < +\infty.$$

$\square$

Desde el punto de vista teórico, los Teoremas 1.4.1 y 1.4.2 son adecuados a los problemas de densidad mixtura porque proporcionan la seguridad de la existencia de estimadores de máxima verosimilitud fuertemente consistentes, caracterizándolos como soluciones de las ecuaciones de verosimilitud y describiendo su conducta asintótica. En la práctica, sin embargo, es frecuente encontrar ciertas dificultades estadísticas, matemáticas e incluso numéricas asociadas con los estimadores de máxima verosimilitud: no siempre existe una solución explícita de las ecuaciones normales, por lo que son necesarios métodos numéricos tales como el

método de Newton-Raphson, el método de Fisher's scoring o alguna de sus variantes. Otra alternativa para muchos de estos problemas es considerarlos como problemas de datos incompletos de otro problema de datos completos, formulado de manera que permita calcular los estimadores del problema original mediante las relaciones entre las funciones de verosimilitud de los dos problemas de una forma más sencilla. Esta es la filosofía del algoritmo EM que a continuación se describe.

## 1.5 Formulación del algoritmo EM

El algoritmo EM es un procedimiento general que se utiliza para el cálculo de estimadores de máxima verosimilitud en una amplia variedad de situaciones. Estas situaciones pueden describirse como *problemas de datos incompletos* de los cuáles McLachlan [39] presenta una gran diversidad de aplicaciones. En estos problemas, la estimación de máxima verosimilitud se dificulta por la ausencia de algunos datos, como son los casos de experimentos con datos missing, distribuciones truncadas, observaciones censuradas o agrupadas, etc...

La idea intuitiva del algoritmo EM consiste en asociar el problema original de datos incompletos con otro de datos completos que sea más tratable y aprovechar su facilidad de cálculo. Así pues, el problema original será reformulado en términos de un conjunto de datos completos, estableciendo una relación entre las funciones de verosimilitud de los dos problemas (usualmente se trabaja con log-verosimilitud).

Además de su aplicación en experimentos con datos incompletos, el algoritmo EM también es de gran utilidad en una variedad de situaciones donde la incompletitud de los datos no aparece de forma natural, sino que es creada de forma ficticia para la adecuada aplicación del mismo. Un ejemplo de esta estrategia es su aplicación a las mixturas distribucionales, ya que tratar problemas de datos completos como si fuesen incompletos proporciona una mejora, desde el punto de vista computacional, cuando han sido reformulados adecuadamente.

El algoritmo EM fue propuesto inicialmente por Demster y otros [15] y consta de dos pasos en cada iteración: uno de cálculo de esperanza (E) y otro de maximización (M).

El paso E consiste en "fabricar" los datos *faltantes*, para el problema de datos completos, usando para ello los datos observados del problema de datos incomple-



tos y los valores de los parámetros de la iteración anterior (la primera iteración se realiza a partir de unos valores iniciales).

Posteriormente, en el paso M de maximización, se actualizan los valores de los parámetros aprovechando la moderada complejidad de la estimación de máxima verosimilitud del problema de datos completos.

Un vector de datos observados  $\mathbf{x}$  puede interpretarse como incompleto y considerarse como una función observable de los llamados datos completos. Si bien la noción de “datos incompletos” incluye el sentido convencional de datos perdidos, también se aplica en situaciones donde los datos completos representan los que estarían disponibles en algún experimento hipotético. En este último caso, los datos completos pueden contener algunas variables que nunca son observables. En este contexto, se denota por  $\mathbf{y}$  el vector que contiene los datos aumentados o los llamados datos completos y por  $\mathbf{z}$  el vector que contiene los datos adicionales referidos como no observables, *faltantes* o perdidos.

Sea  $\mathbf{x} = (x_1, \dots, x_n)$  una muestra aleatoria simple obtenida a partir del vector aleatorio  $X$ , con f.d.d. conjunta  $g(\mathbf{x}; \Psi)$ , y sea  $\Psi = (\psi_1 \dots \psi_d)^T$  un vector de parámetros desconocidos del espacio paramétrico  $\Omega$ .

Se denotará por  $g_c(\mathbf{y}; \Psi)$  la f.d.d. conjunta de la muestra que se obtendría a partir del vector aleatorio  $Y$ , correspondiente al vector de datos completos  $\mathbf{y} = (\mathbf{x}, \mathbf{z})$ . Entonces el logaritmo de la función de verosimilitud de los datos completos para  $\Psi$ , si  $\mathbf{y}$  fuese completamente observable, vendría dado por

$$\log L_c(\Psi) = \log g_c(\mathbf{y}; \Psi).$$

El algoritmo EM aproxima el problema de resolver la ecuación de los datos incompletos

$$\partial \log L(\Psi) / \partial \Psi = 0$$

indirectamente a través de un proceso iterativo en términos del logaritmo de la función de verosimilitud de los datos completos,  $\log L_c(\Psi)$ .

Formalmente, se tienen dos espacios muestrales,  $\mathcal{Y}$  y  $\mathcal{X}$ , y una aplicación no inyectiva  $f : \mathcal{Y} \rightarrow \mathcal{X}$  que hace corresponder a cada vector de datos completos  $\mathbf{y} = (\mathbf{x}, \mathbf{z}) \in \mathcal{Y}$  el vector de datos incompletos  $f(\mathbf{y}) = \mathbf{x}$  en  $\mathcal{X}$ .

Denotando por  $k(\mathbf{z}|\mathbf{x}, \Psi)$  a la función de densidad del vector aleatorio  $Z$  condicionada a que el vector aleatorio  $X$  ha tomado el valor  $\mathbf{x}$ , para todo  $\mathbf{y} \in \mathcal{Y}$ , se verifica:

$$g_c(\mathbf{y}; \Psi) = k(\mathbf{z}|\mathbf{x}, \Psi)g(\mathbf{x}; \Psi)$$

y por tanto, en términos de las verosimilitudes:

$$L_c(\Psi) = k(z|x, \Psi)L(\Psi).$$

Tomando logaritmos, se establece la siguiente relación entre los dos problemas:

$$\log L_c(\Psi) = \log k(z|x, \Psi) + \log L(\Psi).$$

Así pues, si en lugar de observar el vector de datos completos  $Y$  se observa únicamente  $X$ , calculando en ambos lados la esperanza condicionada respecto de la componente  $Z$  de la variable  $Y$  dado  $X = x$  y para un valor  $\Psi'$  del vector paramétrico se tiene:

$$\mathbb{E}_{\Psi'}(\log L_c(\Psi)|x) = \mathbb{E}_{\Psi'}(\log k(Z|x, \Psi)) + \log L(\Psi)$$

(donde el subíndice  $\Psi'$  en el operador  $\mathbb{E}$  significa que en la función respecto a la que se calcula la esperanza el parámetro  $\Psi$  toma el valor  $\Psi'$ ).

Denotando por:

$$Q(\Psi; \Psi') = \mathbb{E}_{\Psi'}(\log L_c(\Psi)|x) \quad \text{y} \quad H(\Psi; \Psi') = \mathbb{E}_{\Psi'}(\log k(Z|x, \Psi)|x),$$

a partir de la igualdad anterior se consiguen distintas descomposiciones de la función de verosimilitud de los datos incompletos, una para cada valor  $\Psi'$  del parámetro:

$$\log L(\Psi) = Q(\Psi; \Psi') - H(\Psi; \Psi').$$

El algoritmo EM, partiendo de una descomposición inicial, con  $\Psi' = \Psi_0$ , se va aproximando a la solución maximizando la función  $Q$  respecto al primer parámetro y utilizando la solución para descomponer  $Q$  en la etapa siguiente. Como se verá más adelante, este proceso conduce a un aumento de la función de verosimilitud en cada etapa.

Desde un punto de vista más formal las etapas del algoritmo EM, tal y como viene recogido en la obra de Dempster, Laird and Rubin (1977), son:

Sea  $\Psi^{(0)}$  un valor inicial para  $\Psi$ . En la primera iteración, en el paso E se calcula:

$$Q(\Psi; \Psi^{(0)}) = \mathbb{E}_{\Psi^{(0)}}\{\log L_c(\Psi)|x\}. \quad (E_0)$$

En el paso M se maximiza  $Q(\Psi; \Psi^{(0)})$  con respecto a  $\Psi$  sobre el espacio paramétrico  $\Omega$ , por lo que se debe encontrar  $\Psi^{(1)}$  tal que

$$Q(\Psi^{(1)}; \Psi^{(0)}) \geq Q(\Psi; \Psi^{(0)}) \quad \forall \Psi \in \Omega. \quad (M_0)$$

Los pasos E y M se realizan de nuevo, pero reemplazando  $\Psi^{(0)}$  por  $\Psi^{(1)}$ .

En la iteración  $k$ , los pasos E y M se definen de la siguiente forma:

**Paso E.** Se calcula  $Q(\Psi; \Psi^{(k)})$ , donde

$$Q(\Psi; \Psi^{(k)}) = \mathbb{E}_{\Psi^{(k)}} \{ \log L_c(\Psi) | \mathbf{x} \}. \quad (E_k)$$

**Paso M.** Se elige  $\Psi^{(k+1)}$ , un valor de  $\Psi \in \Omega$ , tal que

$$Q(\Psi^{(k+1)}; \Psi^{(k)}) \geq Q(\Psi; \Psi^{(k)}) \quad \forall \Psi \in \Omega. \quad (M_k)$$

El sentido del algoritmo EM radica en que la función de verosimilitud es monótona creciente en cualquier iteración generada por el algoritmo EM. Para demostrar este resultado se usa la desigualdad de Jensen que establece  $\mathbb{E}[g(X_1)|X_2] \geq g(\mathbb{E}[X_1|X_2])$  para toda función convexa real y cualesquiera variables aleatorias  $X_1$  e  $X_2$  tal que  $\mathbb{E}[X_1]$  es finita (ver Ash[4]). Por la definición de  $H(\Psi; \Psi')$  y por la convexidad del logaritmo se deduce que  $H(\Psi^{(k+1)}; \Psi^{(k)}) \leq H(\Psi^{(k)}; \Psi^{(k)})$ . Además, el procedimiento de elección de  $\Psi^{(k+1)}$  garantiza que  $Q(\Psi^{(k+1)}; \Psi^{(k)}) \geq Q(\Psi^{(k)}; \Psi^{(k)})$ , y de estas dos desigualdades se deduce que  $L(\Psi^{(k+1)}) \geq L(\Psi^{(k)})$ .

Los pasos E y M se repiten alternativamente hasta que la diferencia

$$L(\Psi^{(k+1)}) - L(\Psi^{(k)})$$

sea suficientemente pequeña, supuesta la convergencia de la sucesión de valores de verosimilitud  $\{L(\Psi^{(k)})\}_k$ . La convergencia depende de que la sucesión de valores de verosimilitud esté acotada superiormente y deberá ser estudiada para cada caso concreto.

Por todo lo anterior, no es necesario especificar la aplicación de  $\mathcal{Y}$  en  $\mathcal{X}$ , ni la correspondiente representación de la función de densidad del conjunto de datos incompletos  $g$  en términos de la función de densidad de los datos completos  $g_c$ . Sólo se necesita especificar el vector de datos completos  $\mathbf{y}$  y la densidad condicionada de  $Z$  dado el vector de datos observados  $\mathbf{x}$  para, así, poder llevar a cabo el paso E.

## 1.6 Algoritmo EM aplicado a las mezclas finitas

Como se ha comentado anteriormente, una de las aplicaciones del algoritmo EM, es la estimación de parámetros para el caso de las mezclas finitas. A continuación, se recoge una breve introducción sobre la misma.

A partir de una muestra  $\mathbf{x} = (x_1, \dots, x_n)$  de una distribución mezcla finita, se tendrá que la función de verosimilitud viene dada por:

$$L(\Psi) = \prod_{i=1}^n \left[ \sum_{j=1}^k \pi_j f(x_i; \theta_j) \right].$$

La maximización de  $L(\Psi)$  con respecto a  $\Psi$  para los datos  $\mathbf{x}$  es equivalente a la maximización de  $\mathcal{L}(\Psi) = \log L(\Psi)$ . La aplicación del algoritmo EM a las mezclas se basa en considerar  $\mathbf{x}$  como un conjunto de datos incompletos en  $\mathcal{Y}(\mathbf{x})$ , el conjunto de los datos completamente categorizados,  $\mathbf{y}$ , que se representa en general como

$$\mathcal{Y}(\mathbf{y}) = \{\mathbf{y}_i : i = 1, \dots, n\} = \{(x_i, \mathbf{z}_i) : i = 1, \dots, n\},$$

donde cada  $\mathbf{z}_i = (z_{ij})$ , con  $j = 1, \dots, k$ , es un vector indicador de longitud  $k$ , con un 1 en la posición correspondiente a la categoría apropiada (a la que pertenece  $x_i$ ) y 0 en el resto.

La verosimilitud correspondiente a  $(\mathbf{y}_1, \dots, \mathbf{y}_n)$  viene dada por:

$$L_c(\Psi) = g(\mathbf{y}_1, \dots, \mathbf{y}_n; \Psi) = \prod_{i=1}^n \prod_{j=1}^k \pi_j^{z_{ij}} f_j(x_i; \theta_j)^{z_{ij}} \quad (1.2)$$

con logaritmo:

$$\mathcal{L}_c(\Psi) = \sum_{i=1}^n \mathbf{z}_i^T \begin{pmatrix} \log \pi_1 \\ \log \pi_2 \\ \dots \\ \log \pi_k \end{pmatrix} + \sum_{i=1}^n \mathbf{z}_i^T \begin{pmatrix} \log f_1(x_i; \theta_1) \\ \log f_2(x_i; \theta_2) \\ \dots \\ \log f_k(x_i; \theta_k) \end{pmatrix}.$$

La función de verosimilitud de los datos incompletos obtenidos de la mezcla,  $L(\Psi)$ , se corresponde con la densidad marginal de  $x_1, \dots, x_n$  obtenida por la suma de (1.2) sobre todos los posibles valores de  $\mathbf{z}_1, \dots, \mathbf{z}_n$ . En efecto, teniendo

en cuenta que  $g(\mathbf{y}_1, \dots, \mathbf{y}_n; \Psi)$  es el producto de  $n$  funciones que dependen cada una de sólo un  $\mathbf{z}_i$ , la densidad marginal de  $g(\mathbf{y}_1, \dots, \mathbf{y}_n; \Psi)$  respecto de  $\mathbf{z}_1, \dots, \mathbf{z}_n$  es igual al producto de las densidades marginales de cada una de estas funciones con respecto a cada  $\mathbf{z}_i$ . Puesto que  $\mathbf{z}_1, \dots, \mathbf{z}_n$  son en este caso variables aleatorias discretas con  $k$  valores posibles ( $\mathbf{e}_1 = (1 \ 0 \dots 0)$ ,  $\mathbf{e}_2 = (0 \ 1 \dots 0)$ ,  $\dots$ ,  $\mathbf{e}_k = (0 \dots 0 \ 1)$ ), se tiene que la densidad marginal de  $g(\mathbf{y}_1, \dots, \mathbf{y}_n; \Psi)$  sobre todos los posibles valores de  $\mathbf{z}_1, \dots, \mathbf{z}_n$  es

$$\prod_{i=1}^n \left[ \sum_{h=1}^k \left[ \prod_{j=1}^k \pi_j^{z_{hj}} f_j(\mathbf{x}_i; \theta_j)^{z_{hj}} \right] \right]. \quad (1.3)$$

Además las  $z_{hj}$  son las componentes de los vectores unitarios  $\mathbf{e}_h$  que se anulan salvo para el caso  $h = j$ , por lo que la expresión (1.3) queda:

$$\prod_{i=1}^n \left[ \sum_{j=1}^k \pi_j f_j(\mathbf{x}_i; \theta_j) \right].$$

No es posible el cálculo explícito de los estimadores de máxima verosimilitud de esta función, sin embargo se pueden aproximar mediante el algoritmo EM usando la función de verosimilitud del conjunto de datos completos que se ha definido.

En esta situación el algoritmo EM genera desde un valor inicial de  $\Psi$ ,  $\Psi_0$ , una secuencia de estimadores  $\{\Psi^{(m)}\}$ . Cada iteración está constituida por los dos pasos siguientes:

**Paso E:** Evalúa  $\mathbb{E}[\log g(\mathbf{y}; \Psi) | \mathbf{x}, \Psi^{(m)}] = Q(\Psi, \Psi^{(m)})$ .

Dado que  $\mathbf{z}_1, \dots, \mathbf{z}_n$  son desconocidas, se tiene que  $Q(\Psi, \Psi^{(m)})$  viene dada por la expresión:

$$\sum_{i=1}^n \mathbf{w}_i(\Psi^{(m)})^T \begin{pmatrix} \log \pi_1 \\ \log \pi_2 \\ \dots \\ \log \pi_k \end{pmatrix} + \sum_{i=1}^n \mathbf{w}_i(\Psi^{(m)})^T \begin{pmatrix} \log f_1(\mathbf{x}_i; \theta_1) \\ \log f_2(\mathbf{x}_i; \theta_2) \\ \dots \\ \log f_k(\mathbf{x}_i; \theta_k) \end{pmatrix}, \quad (1.4)$$

con  $\mathbf{w}_i(\Psi^{(m)}) = \mathbb{E}(\mathbf{z}_i | \mathbf{x}_i, \Psi^{(m)}) = h(\mathbf{e}_1 | \mathbf{x}_i) \mathbf{e}_1 + h(\mathbf{e}_2 | \mathbf{x}_i) \mathbf{e}_2 + \dots + h(\mathbf{e}_k | \mathbf{x}_i) \mathbf{e}_k =$   
 $= (h(\mathbf{e}_1 | \mathbf{x}_i) \ h(\mathbf{e}_2 | \mathbf{x}_i) \ \dots \ h(\mathbf{e}_k | \mathbf{x}_i))^T.$

Aplicando el teorema de Bayes, se obtiene que cada componente de  $\mathbf{w}_i(\Psi^{(m)})$  puede ser expresada por:

$$w_{ij}(\Psi^{(m)}) = \frac{\pi_j^{(m)} f_j(\mathbf{x}_i; \theta_j^{(m)})}{q(\mathbf{x}_i; \Psi^{(m)})},$$

siendo  $q(\mathbf{x}_i; \Psi^{(m)})$  la función de densidad mixtura en  $\mathbf{x}_i$ , es decir la marginal de la función de densidad de los datos completos en  $\mathbf{x}_i$ .

Paso M: Encuentra  $\Psi = \Psi^{(m+1)}$  para maximizar  $Q(\Psi, \Psi^{(m)})$ .

Como se observa en la expresión (1.4), el primer sumatorio sólo depende de los pesos  $\pi_1, \dots, \pi_k$  mientras que en el segundo aparecen el resto de los parámetros  $\theta_1, \dots, \theta_k$ , por tanto estas partes se maximizan por separado. Como la función depende linealmente de  $\log \pi_1, \dots, \log \pi_k$ , el primer problema de maximización tiene una única solución que se resuelve por los multiplicadores de Lagrange y es independiente de la forma funcional de las densidades componentes  $f_i(\mathbf{x}; \theta_i)$ :

$$\pi_j^{(m+1)} = \frac{1}{n} \sum_{i=1}^n w_{ij}(\Psi^{(m)}) \quad j = 1, \dots, k.$$

De modo que para resolver el paso M, basta maximizar el segundo sumando de  $Q(\Psi, \Psi^{(m)})$ , respecto de los vectores paramétricos  $\theta_1, \dots, \theta_k$ . Si, además, cada uno de estos vectores paramétricos aparece únicamente en una de las funciones componentes de la mixtura, el segundo problema de maximización se divide a su vez en  $k$  problemas de maximización componentes, interviniendo en cada uno un único parámetro o vector paramétrico. Todos estos problemas de maximización pueden verse como problemas de estimación de máxima verosimilitud "ponderados" en los que aparecen sumas de logaritmos ponderadas por las probabilidades *a posteriori* de que las observaciones muestrales pertenezcan a la apropiada población componente una vez dado el estimador de máxima verosimilitud aproximado de  $\Psi^*$ . Así pues, la dificultad del problema dependerá ahora de la dificultad que presente el cálculo de los estimadores de las funciones de densidad que intervengan en la mixtura.

De la comparación del algoritmo EM con otros métodos numéricos ya mencionados, como el de Newton-Raphson y el de Fisher's scoring, conviene reflejar, como recoge McLachlan [39], los siguientes aspectos:

a) EM es más fácil de aplicar generalmente y es numéricamente estable pues con cada iteración EM incrementa la verosimilitud.

b) Los métodos el de Newton-Raphson y de Fisher's scoring son más difíciles de aplicar en la práctica porque necesitan invertir una matriz información en cada etapa, además de no garantizar la monotonía de la verosimilitud.

c) El método de Newton-Raphson tiene convergencia de segundo orden, mientras que el algoritmo EM es generalmente lento. No obstante, si la separación

entre las componentes de la mixtura no es muy grande, Newton-Raphson también puede ser muy lento.

d) Si se satisfacen las condiciones de regularidad exigidas en la teoría asintótica de la familia, los métodos de Newton-Raphson y Scoring proporcionan en cada iteración una aproximación a la matriz de covarianzas, sin embargo, en el algoritmo EM no se dispone de ella directamente, sino que hay que calcularla por otro procedimiento.

e) La convergencia al estimador de máxima verosimilitud no se tiene garantizada con ninguno de los métodos, pero, bajo condiciones bastante generales, el algoritmo EM tiene convergencia global, es decir que converge generalmente al máximo global salvo elecciones inadecuadas de los valores iniciales o existencia de singularidades. Sin embargo, otros métodos como Newton-Raphson tienden a dirigirse hacia puntos de silla y mínimos locales tan a menudo como a máximos locales.

f) Existen extensiones del algoritmo EM como el GEM (algoritmo EM generalizado) o ECM (esperanza-condicional maximización) que se implementan iterativamente de forma bastante simple y que conservan la convergencia monótona estable del algoritmo EM, para aquellos casos en los que la resolución del paso M no es directa.

## Capítulo 2

# Mixturas finitas de distribuciones: Identificabilidad

### 2.1 Introducción

El objetivo de este capítulo es la búsqueda de condiciones suficientes que permitan comprobar la identificabilidad de algunos experimentos modelizados a través de mixturas finitas. Se comienza con la formulación del problema de modelización de variables con asimetría positiva en los datos a través de una mixtura mixta de funciones de las familias Lognormal, Gamma y Weibull. Una vez planteado el modelo, el siguiente objetivo que se persigue es el estudio de la identificabilidad del mismo.

En la sección cuarta, se recogen, detalladamente, los resultados más importantes sobre la identificabilidad de mixturas. Dado que estas condiciones no son aplicables a algunos modelos de interés, en la siguiente sección, se propone una nueva condición suficiente para demostrar la identificabilidad que relaja las condiciones de los resultados anteriormente recogidos.

Se finaliza este capítulo, con aplicaciones de los resultados obtenidos a algunas mixturas de interés.



## 2.2 Formulación del problema

Tal y como se recoge en Titterington [52], la mayoría de los resultados acerca de mixturas utilizados en la literatura tratan sobre aquéllas cuyas componentes provienen de una misma familia paramétrica. No obstante, existen experiencias que pueden ser modelizadas a través de mixturas de diferentes familias, éste es el caso planteado en esta memoria. Así, se estudian las mixturas resultantes de la combinación lineal convexa de una distribución Lognormal, una Gamma y una Weibull. Es decir, se está tratando con mixturas muy concretas obtenidas de las combinaciones de tres funciones de tres familias respectivas distintas. Una forma alternativa a este planteamiento, hubiera sido considerar estas familias como un caso particular de una única familia paramétrica que englobase a las tres anteriores, pero, bajo este enfoque, se aumentaría el número de parámetros y, del mismo modo la complejidad en la resolución del problema. Si se considera la familia de funciones resultante de la unión de estas tres familias, la mixtura del problema bajo estudio pertenecerá a un subconjunto del conjunto de mixturas que genera dicha unión. Seguidamente se pasa a la formulación del problema.

Sea  $X$  una variable aleatoria con función de densidad mixtura  $g(\cdot; \Psi)$  de la forma:

$$g(x; \Psi) = \sum_{j=1}^3 \pi_j f_j(x; \theta_j),$$

donde  $\{\pi_i\}_{i=1}^3 \in \mathcal{C}$  y  $f_1, f_2$  y  $f_3$  son las funciones de densidad Lognormal, Gamma y Weibull, respectivamente, dadas por:

$$\begin{aligned} f_1(x; \mu, \sigma) &= \frac{1}{\sqrt{2\pi\sigma x}} \exp\left\{-\frac{(\log x - \mu)^2}{2\sigma^2}\right\}, \\ f_2(x; a, b) &= \frac{b^{-a} x^{a-1} \exp\{-x/b\}}{\Gamma(a)} \quad \text{y} \\ f_3(x; c, d) &= \frac{c}{d^c} x^{c-1} \exp\{-x^c/d^c\}. \end{aligned}$$

Dada una muestra  $(x_1, \dots, x_n)$ , la función de verosimilitud asociada viene dada por:

$$L(\Psi) = \prod_{i=1}^n \left[ \sum_{j=1}^3 \pi_j f_j(x_i; \theta_j) \right].$$

De esta forma,  $L$  depende de los ocho parámetros componentes del vector  $\Psi = (\pi_1, \pi_2, \mu, \sigma, a, b, c, d)$ .

Tal y como se recoge en el siguiente resultado, igual que en el caso de una mezcla de dos distribuciones normales, (Kiefer y Wolfowitz [28]),  $L$  no está acotada superiormente.

**Proposición 2.2.1.**  $L(\Psi)$  no está acotada superiormente.

*Demostración.* Basta considerar un punto  $\Psi$  de la forma

$$(1/3, 1/3, \log x_1, \sigma, a, b, c, d)$$

siendo  $x_1$  un dato cualquiera de la muestra obtenida y  $a, b, c, d$  cualesquiera valores. Entonces la función de verosimilitud  $L(\Psi)$  tendrá la forma:

$$\frac{1}{3} \left[ \frac{1}{\sqrt{2\pi}\sigma x_1} e^{-\frac{(\log x_1 - \log x_1)^2}{2\sigma^2}} + f_2(x_1) + f_3(x_1) \right] \prod_{i=2}^n \left[ \sum_{j=1}^3 \frac{1}{3} f_j(x_i) \right].$$

El exponente en el primer sumando se anula por lo que, si se hace tender  $\sigma$  a 0, el primer sumando no está acotado superiormente y, como en cada uno de los demás factores, el segundo y tercer sumando son cantidades fijas no nulas e independientes de  $\sigma$ , la función de verosimilitud tenderá a infinito cuando  $\sigma$  tienda a 0.  $\square$

Este resultado indica que el máximo global no existe y, por lo tanto, como se citó en el capítulo anterior, el interés se debe centrar en la búsqueda de máximos locales.

## 2.3 Identificabilidad

Tal y como se ha recogido en la Sección 1.3, el problema de la identificabilidad de una familia de distribuciones es una cuestión teórica importante que permite conocer si es posible estimar de manera única los parámetros de un modelo. En general, una familia paramétrica de funciones de distribución es identificable si distintos valores de los parámetros determinan distintos miembros de la familia.

En la literatura aparece una gran cantidad de trabajos en los que se estudia el problema de la identificabilidad de mezclas finitas (e infinitas numerables). En este sentido, Teicher [51] proporcionó una condición suficiente para que la mezcla de una familia fuera identificable, dicho resultado lo aplicó a las distribuciones Normal y Gamma. Posteriormente, Chandra [10] propuso una modificación

del resultado de Teicher. Ahmad [1] lo utilizó para probar la identificabilidad de mixturas finitas de distribuciones Weibull, Lognormal, Chi-cuadrado y Pareto usando la función generatriz de momentos del logaritmo de la variable en estudio. Barndorff-Nielsen [6] estudiaron la identificabilidad de mixturas de familias exponenciales. Una modificación del argumento usado por Teicher le sirvió a Henna [20] para probar un resultado de identificabilidad basado en un orden biparamétrico, utilizándolo para comprobar la identificabilidad de mixturas de distribuciones log-gamma y log-gamma inversa.

Dado que en esta memoria se considera un caso especial, puesto que en la mixtura intervienen una sola función de cada una de tres familias diferentes de distribuciones, se considerará  $\mathcal{F}_L$  la familia de distribuciones Lognormal,  $\mathcal{F}_G$  la familia de distribuciones Gamma y  $\mathcal{F}_W$  la familia de distribuciones Weibull y a partir de ellas se define la siguiente familia:

**Definición 2.3.1.** Se define el conjunto de mixturas mixtas  $\mathcal{P}$  de las familias  $\mathcal{F}_L$ ,  $\mathcal{F}_G$  y  $\mathcal{F}_W$  como el conjunto formado por las mixturas de la forma

$$G(x, \Psi) = \pi_1 F_1(x, \theta_1) + \pi_2 F_2(x, \theta_2) + \pi_3 F_3(x, \theta_3),$$

donde  $F_1 \in \mathcal{F}_L$ ,  $F_2 \in \mathcal{F}_G$ ,  $F_3 \in \mathcal{F}_W$ , y  $F_2 \neq F_3$  y  $\{\pi_i\}_{i=1}^3 \in \mathcal{C}$ .

*Nota 2.3.1.* Nótese que la familia Gamma y Weibull tienen en común las distribuciones exponenciales, de ahí el requisito  $F_2 \neq F_3$ .

De acuerdo con la Definición 1.3.3,  $\mathcal{P}$  será identificable si para cualesquiera  $G(x, \Psi), G'(x, \Psi') \in \mathcal{P}$  tales que

$$G(x, \Psi) = G'(x, \Psi')$$

se verifica que

$$[\pi_j = \pi'_j, \quad F_j(x, \theta_j) = F'_j(x, \theta'_j) \quad j = 1, 2, 3].$$

Para demostrar la identificabilidad de la familia  $\mathcal{P}$  se probará un resultado más general: la identificabilidad del conjunto  $\mathcal{H}_{\mathcal{F}_U}$  de mixturas finitas de elementos de la familia  $\mathcal{F}_U = \mathcal{F}_L \cup \mathcal{F}_G \cup \mathcal{F}_W$ . Dado que  $\mathcal{P} \subseteq \mathcal{H}_{\mathcal{F}_U}$ , es obvio que si  $\mathcal{H}_{\mathcal{F}_U}$  es identificable entonces  $\mathcal{P}$  también lo es.

## 2.4 Resultados previos

Como se ha comentado anteriormente, el problema de la identificabilidad ha centrado la atención de muchos autores durante décadas. Yakowitz and Spragins

[54] demostraron que la clave para responder a la cuestión de la identificabilidad está en la independencia lineal de los miembros de la familia de funciones de distribución de la cual se forman las mixturas,  $\mathcal{F}$ .

**Teorema 2.4.1.** (*Yakowitz and Spragins [54]*) Una condición necesaria y suficiente para que  $\mathcal{H}$  sea identificable es que  $\mathcal{F}$  sea un conjunto linealmente independiente sobre el cuerpo de los números reales  $\mathbb{R}$  (i.e. no existe una combinación lineal no nula de miembros de  $\mathcal{F}$  tal que  $\sum_{j=1}^k \lambda_j F_j = 0$ ).

Sin embargo, esta condición es difícil de verificar en muchas situaciones. El primer resultado sobre identificabilidad aplicable a muchos modelos diferentes fue dado por Teicher [51]. Este resultado puede considerarse como el punto de partida del resto de trabajos sobre esta materia.

**Teorema 2.4.2.** (*Teicher [51]*) Sea  $\mathcal{F}$  una familia de funciones de distribución para la que existe una aplicación  $M$  que asocia a cada elemento  $F$  de la misma una transformada  $\phi_F(t)$  con dominio de definición  $S_F$ . Si  $M$  es lineal y biyectiva, y existe un orden total sobre  $\mathcal{F}$ , denotado por ' $\prec$ ', tal que:

- (a)  $F_1 \prec F_2 \Rightarrow S_{F_1} \subseteq S_{F_2}$   
 (b)  $\exists t(F_1) \in \bar{S}_{F_1}$  (la clausura de  $S_{F_1}$ ) tal que

$$F_1 \prec F_2 \Rightarrow \lim_{t \rightarrow t(F_1)} \frac{\phi_{F_2}(t)}{\phi_{F_1}(t)} = 0,$$

entonces la clase de todas las mixturas finitas de  $\mathcal{F}$  es identificable.

De este resultado Teicher [51] dedujo directamente la identificabilidad de las mixturas finitas generadas por la familia de distribuciones Gamma, así como de las generadas por normales univariantes, usando como transformada la función generatriz de momentos. Posteriormente, Ahmad [1] aplicó este teorema para demostrar la identificabilidad de las mixturas generadas por las familias de distribuciones Weibull y Lognormal, pero utilizando como transformada la función generatriz de momentos del logaritmo neperiano de la variable.

Sin embargo, este teorema, con las transformaciones usuales, no es aplicable a la clase bajo estudio  $\mathcal{P}$  ya que la condición (a) del teorema no se verifica. En algunas situaciones, dicha condición puede ser muy restrictiva debido a que ordena las funciones a partir de la contención de los dominios. Así pues, puede verse dificultada su aplicación en aquellos casos donde los conjuntos de definición  $S_F$  dependan de más de un parámetro.

Basándose en el argumento de Teicher [51] e imponiendo otras condiciones, Henna [20] propone un nuevo resultado que permite demostrar la identificabilidad

de ciertas mixturas. Estas condiciones dependen de la construcción de un orden sobre el espacio paramétrico, considerándolo como un espacio producto. A continuación, se presenta este resultado que va dirigido a familias de distribuciones de la forma:

$$\mathcal{F}_1 = \{F(x; \alpha, \beta) : \alpha \in \Omega_1 \subseteq \mathbb{R}^m, \beta \in \Omega_2 \subseteq \mathbb{R}^n\}$$

es decir, donde el espacio paramétrico puede ser representado por

$$\Omega = \Omega_1 \times \Omega_2 \subseteq \mathbb{R}^{m+n}.$$

Previamente, es necesario recoger algunos conceptos y consideraciones.

1.- La existencia de un orden total en  $\Omega_1, \prec^\alpha$ , y otro en  $\Omega_2, \prec^\beta$ , induce un orden total en  $\Omega$  dado por:

$$(\alpha_1, \beta_1) \prec (\alpha_2, \beta_2) \Leftrightarrow \begin{cases} \alpha_1 \prec^\alpha \alpha_2 \\ \text{ó} \\ \alpha_1 = \alpha_2, \quad \beta_1 \prec^\beta \beta_2. \end{cases}$$

2.- Un subconjunto  $A \subseteq \Omega$ , se dice que está bien ordenado respecto a la relación de orden  $\prec$  si todo subconjunto de  $A$  no vacío tiene mínimo, es decir

$$\forall A_1 \subset A \quad \exists (\alpha_1, \beta_1) \in A_1 : (\alpha_1, \beta_1) \prec (\alpha, \beta), \quad \forall (\alpha, \beta) \in A_1 \setminus \{(\alpha_1, \beta_1)\}.$$

Y se denotará  $(A, \prec)$  bien ordenado.

La familia de todos los subconjuntos de  $\Omega$  bien ordenados se denotará por  $\mathcal{B}_\Omega$ , es decir:

$$\mathcal{B}_\Omega = \{(A, \prec) : A \subset \Omega, (A, \prec) \text{ bien ordenado}\}$$

3.- A través de  $\mathcal{B}_\Omega$  se puede definir una amplia clase de distribuciones discretas

$$\mathcal{G}_{m,n}^* = \bigcup_{A \in \mathcal{B}_\Omega} \left\{ G(\alpha, \beta) : \sum_{(\alpha_r, \beta_r) \in A} P_G(\alpha_r, \beta_r) = 1, P_G(\alpha_r, \beta_r) > 0 \right\},$$

donde  $P_G$  es la probabilidad asociada a la función de distribución  $G$ . Esta clase incluye todas las distribuciones discretas con soporte finito.

En consecuencia, la clase de las mixturas de  $\mathcal{F}$ , con distribución mixing en  $\mathcal{G}_{m,n}^*$ ,  $\mathcal{G}_{m,n}^*$  - mixtura de  $\mathcal{F}$ , incluye a la clase de todas las mixturas finitas de  $\mathcal{F}$ .

4.-Para cualquier otra clase de distribuciones

$$\mathcal{G}_{m,n} = \{G(\alpha, \beta) : P_G(\Omega) = 1\}$$

tal que  $\mathcal{G}_{m,n}^* \subset \mathcal{G}_{m,n}$ , se considera la  $\mathcal{G}_{m,n}$  - *mixtura* de  $\mathcal{F}$ , que se denotará por  $\mathcal{H}$ , y una transformación  $\mathcal{M}$  definida sobre  $\mathcal{H}$  que transforma cualquier elemento  $H_G \in \mathcal{H}$  en una función real con dominio  $S_{H_G} \subset \mathbb{R}$ ,

$$\begin{aligned} \mathcal{M} : \mathcal{H} &\rightarrow \Phi \\ : H_G &\rightarrow \mathcal{M}(H_G) = \phi_{H_G} : S_{H_G} \subseteq \mathbb{R} \rightarrow \mathbb{R} \end{aligned}$$

El resultado de Henna [20] se basa en la existencia de una transformación  $\mathcal{M}$  definida como en el punto 4 que verifique las condiciones que se recogen a continuación:

(H1) Existe un subconjunto no vacío  $S$ , tal que  $S \subset S_{H_G}$ , para cualquier  $H_G \in \mathcal{H}$ , o sea la intersección de los dominios de definición de la clase de las transformadas  $\Phi$  es un subconjunto real no vacío.

(H2)  $\phi_{(\alpha,\beta)}(t) \neq 0$  para cualquier  $t \in S$  y para cualquier  $(\alpha, \beta) \in \mathbb{R}^{m+n}$ . Es decir, en dicha intersección las transformadas de los elementos de  $\mathcal{F}$  no se anulan.

(H3)  $\phi_{H_G}(t) = \int_{\mathbb{R}^{m+n}} \phi_{(\alpha,\beta)}(t) dG(\alpha, \beta)$  para cualquier  $t \in S$  y para cualquier  $H_G \in \mathcal{H}$ .

(H4) Para todo  $G \in \mathcal{G}_{m,n}$  y para cualquier  $\alpha^* \in \mathbb{R}^m$ , siendo  $\Theta_G$  el soporte de  $G$ , se verifica que  $A = \{\alpha : \alpha^* \prec^\alpha \alpha, (\alpha, \beta) \in \Theta_G\}$  es no vacío y  $(A, \prec^\alpha)$  bien ordenado.

(H5) Para cualquier  $(\alpha^*, \beta^*)$  de  $\mathbb{R}^{m,n}$ , existe un  $\bar{t} (\in \bar{S})$  independiente de  $(\alpha^*, \beta^*)$  tal que en un entorno de  $\bar{t}$  se satisface:

(i) Si  $\alpha^* \prec^\alpha \alpha$  entonces  $\phi_{(\alpha,\beta)}(t)/\phi_{(\alpha^*,\beta^*)}(t)$  está acotado y tiene signo constante en  $(\alpha, \beta, t)$  y  $\phi_{(\alpha,\beta)}(t)/\phi_{(\alpha^*,\beta^*)}(t) \rightarrow 0$  cuando  $t \rightarrow \bar{t}$ , donde  $(\alpha, \beta) \in \mathbb{R}^{m+n}$  y  $\alpha \preceq^\alpha \alpha$ .

(ii)  $\phi_{(\alpha^*,\beta)}(t)/\phi_{(\alpha^*,\beta^*)}(t)$  está acotada y tiene signo constante en  $(\beta, t)$  y  $\phi_{(\alpha^*,\beta)}(t)/\phi_{(\alpha^*,\beta^*)}(t) \rightarrow 0$  cuando  $t \rightarrow \bar{t}$  donde  $(\alpha^*, \beta) \in \mathbb{R}^{m+n}$  y  $\beta^* \prec^\beta \beta$ .

**Teorema 2.4.3.** (Henna [20]) Bajo las condiciones (H1)-(H5), sean  $G, G^* \in \mathcal{G}_{m,n}$  tales que:

$$H_G(x) = H_{G^*}(x) \quad y \quad G^* \in \mathcal{G}_{m,n}^*$$

entonces  $G = G^*$ .

Como consecuencia del teorema se deduce el siguiente corolario de identificabilidad para  $\mathcal{F}$ .

**Corolario 2.4.1.** *Si se verifican las condiciones H1-H5, la clase de mixturas finitas de  $\mathcal{F}$  es identificable.*

Henna [20] usa este resultado para probar la identificabilidad de mixturas finitas de distribuciones log-gamma y log-gamma inversa. Sin embargo, para las mixturas que se tratan en este capítulo es bastante restrictivo y no puede ser aplicado. El resultado, que se da en la próxima sección, permite relajar las condiciones del Teorema 2.4.2 de Teicher sin necesidad de complicar el problema con órdenes biparamétricos. Tanto es así, que dicho resultado puede ser usado para probar la identificabilidad de las mixturas formadas a partir de las funciones de la unión de las familias de distribuciones Gamma, Lognormal y Weibull, que es el objetivo fundamental de este capítulo, pero también para probar la identificabilidad de mixturas finitas de distribuciones log-gamma y log-gamma inversa de forma más sencilla que con el resultado de Henna.

## 2.5 Una nueva condición suficiente de identificabilidad

A continuación, se propone un nuevo resultado que evita la estructura del espacio paramétrico como espacio producto y sustituye la condición (a) en el Teorema 2.4.2 de Teicher por una condición mucho más general.

**Teorema 2.5.1.** *Sea  $\mathcal{F}$  una familia de distribuciones. Sea  $M$  una aplicación lineal que transforma cada  $F \in \mathcal{F}$  en una función real  $\phi_F$  definida en un subconjunto  $S_F \subseteq \mathbb{R}^d$ . Si existe un orden total  $\prec$  sobre  $\mathcal{F}$ , tal que  $\forall F \in \mathcal{F} \exists t(F) \in S_0(F)'$ , donde  $S_0(F)'$  es el conjunto de acumulación de  $S_0(F) = \{t \in S_F : \phi_F(t) \neq 0\}$ , verificando:*

$$(i) \forall F_1, F_2, \dots, F_m \in \mathcal{F} \quad / \quad F_1 \prec F_i \quad 2 \leq i \leq m \Rightarrow$$

$$t(F_1) \in [S_0(F_1) \cap [\cap_{i=2}^m S(F_i)]]'.$$

$$(ii) \forall F_1, F_2 \in \mathcal{F} \quad / \quad F_1 \prec F_2 \Rightarrow$$

$$\lim_{t \rightarrow t(F_1)} \frac{\phi_{F_2}(t)}{\phi_{F_1}(t)} = 0.$$

entonces la clase  $\mathcal{H}_{\mathcal{F}}$  de todas las mixturas finitas de  $\mathcal{F}$  es identificable.

*Demostración.* Supóngase que existen dos conjuntos finitos de elementos distintos de  $\mathcal{F}$ ,  $\mathcal{F}_1 = \{F_i, 1 \leq i \leq \hat{k}\}$  y  $\mathcal{F}_2 = \{\hat{F}_j, 1 \leq j \leq \hat{k}\}$  de modo que:

$$\sum_{i=1}^{\hat{k}} \pi_i F_i = \sum_{j=1}^{\hat{k}} \hat{\pi}_j \hat{F}_j \quad \{\pi_i\}_{i=1}^{\hat{k}}, \{\hat{\pi}_i\}_{i=1}^{\hat{k}} \in \mathcal{C}. \quad (2.1)$$

Sin pérdida de generalidad, se puede suponer que  $F_i \prec F_j, \hat{F}_i \prec \hat{F}_j$  para  $i < j$  y  $F_1 \preceq \hat{F}_1$ . De este modo,  $F_1 \prec \hat{F}_j \quad \forall j = 2, \dots, \hat{k}$ .

Sea:

$$A = S_0(F_1) \cap \left[ \bigcap_{i=2}^{\hat{k}} S(F_i) \right] \cap \left[ \bigcap_{i=1}^{\hat{k}} S(\hat{F}_i) \right].$$

Por la condición (i) existe  $t(F_1) \in A'$ , conjunto de puntos de acumulación de  $A$ . Así,  $A$  es un conjunto con infinitos puntos, ya que su conjunto de puntos de acumulación es no vacío. Para cada  $t \in A$ , por la linealidad de  $M$  y (2.1) se tiene que

$$\sum_{i=1}^{\hat{k}} \pi_i \phi_{F_i}(t) = \sum_{j=1}^{\hat{k}} \hat{\pi}_j \phi_{\hat{F}_j}(t).$$

Puesto que  $t \in S_0(F_1)$ , se pueden dividir por  $\phi_{F_1}(t)$  ambos miembros de la igualdad y, haciendo  $t \rightarrow t(F_1)$ , se tiene

$$\sum_{i=1}^{\hat{k}} \pi_i \lim_{t \rightarrow t(F_1)} \frac{\phi_{F_i}(t)}{\phi_{F_1}(t)} = \sum_{i=1}^{\hat{k}} \hat{\pi}_i \lim_{t \rightarrow t(F_1)} \frac{\phi_{\hat{F}_i}(t)}{\phi_{F_1}(t)}.$$

Teniendo en cuenta que  $F_1 \prec F_i$ , para todo  $i = 2, 3, \dots, \hat{k}$  y  $F_1 \prec \hat{F}_j$ , para todo  $j = 2, 3, \dots, \hat{k}$ , por la condición (ii), los respectivos sumandos se anulan, con lo que quedaría

$$\pi_1 = \hat{\pi}_1 \lim_{t \rightarrow t(F_1)} \frac{\phi_{\hat{F}_1}(t)}{\phi_{F_1}(t)}.$$

Si fuese  $F_1 \prec \hat{F}_1$  se tendría que  $\pi_1 = 0$ , lo cual es contradictorio dado que  $\{\pi_i\}_{i=1}^{\hat{k}} \in \mathcal{C}$ . Por tanto  $F_1 = \hat{F}_1$  y  $\pi_1 = \hat{\pi}_1$ . Cancelando ambos sumandos en (2.1) queda:

$$\sum_{i=2}^{\hat{k}} \pi_i F_i(x) = \sum_{j=2}^{\hat{k}} \hat{\pi}_j \hat{F}_j(x).$$



Repetiendo el argumento un número finito de veces, se llega a que  $\pi_i = \hat{\pi}_i$  y  $F_i = \hat{F}_i$  para todo  $i = 1, 2, \dots, \min\{k, \hat{k}\}$ . Además, se tiene que  $k = \hat{k}$ . En efecto, si  $\hat{k} > k$ , entonces se tendría  $\sum_{j=k+1}^{\hat{k}} \hat{\pi}_j \hat{F}_j(x) = 0$ , y por tanto  $\hat{\pi}_j = 0$ ,  $k+1 \leq j \leq \hat{k}$ , de nuevo una contradicción. Así pues,  $\mathcal{H}_{\mathcal{F}}$  es identificable.  $\square$

*Nota 2.5.1.* El argumento dado en la demostración del Teorema 2.5.1 sigue siendo válido considerando  $t(F_i) = \pm\infty$ .

*Nota 2.5.2.* Si la familia de distribuciones  $\mathcal{F}$  es una familia paramétrica, la existencia del orden total en el espacio paramétrico  $\Omega$  induce un orden total en  $\mathcal{F}$ , y viceversa. Esto permite encontrar el orden total requerido en el conjunto de las distribuciones a través de uno definido en el espacio paramétrico, generalmente más sencillo.

Por otra parte, en el enunciado del Teorema 2.5.1 se observa que el punto  $t(F)$  depende de cada distribución  $F \in \mathcal{F}$ . En el caso de que el punto  $t(F) = t_0$  sea independiente de  $F$  y teniendo en cuenta que la relación

$$F_1 \prec F_2 \text{ si y sólo si } \lim_{t \rightarrow t_0} \frac{\phi_{F_2}(t)}{\phi_{F_1}(t)} = 0$$

define un orden parcial sobre  $\mathcal{F}$  para  $t_0$  fijo (por ser antireflexiva, antisimétrica y transitiva) el siguiente corolario simplifica la comprobación de las hipótesis.

**Corolario 2.5.1.** *Sea  $\mathcal{F}$  una familia de distribuciones. Sea  $M$  una aplicación lineal que transforma cada  $F \in \mathcal{F}$  en una función real  $\phi_F$  definida en un subconjunto  $S_F \subseteq \mathbb{R}^d$  y sea  $S_0(F) = \{t \in S_F : \phi_F(t) \neq 0\}$ . Si existe un punto  $t_0$  verificando:*

(a) *Para toda colección finita de distribuciones  $F_1, \dots, F_k \in \mathcal{F}$ ,*

$$t_0 \in \left[ \bigcap_{1 \leq i \leq k} S_0(F_i) \right]'$$

(b) *El orden dado por*

$$F_1 \prec F_2 \text{ si y sólo si } \lim_{t \rightarrow t_0} \frac{\phi_{F_2}(t)}{\phi_{F_1}(t)} = 0$$

*es un orden total sobre  $\mathcal{F}$ ,*

*entonces la clase  $\mathcal{H}_{\mathcal{F}}$  de todas las mezclas finitas de distribuciones de  $\mathcal{F}$  es identificable.*

Este corolario puede simplificar considerablemente la comprobación de la identificabilidad de familias que tienen asociadas una aplicación lineal que transforma cada distribución en funciones reales con dominios de la forma  $(a(F), +\infty)$ ,

$(-\infty, b(F))$  o  $(-\infty, +\infty)$ , ya que  $+\infty$  y  $-\infty$  pueden jugar el papel de  $t_0$  en el corolario.

## 2.6 Estudio de la identificabilidad de algunas familias de distribuciones finitas

En esta sección se recogen algunas aplicaciones del Teorema 2.5.1 para deducir la identificabilidad de algunas familias de mixturas finitas.

### 2.6.1 Identificabilidad de mixturas de distribuciones de las familias Log-gamma y Log-gamma inversa

A continuación, se aprovecha la versatilidad del Teorema 2.5.1 para comprobar la identificabilidad de mixturas finitas de dos familias de distribuciones, las cuales ya han sido estudiadas por Henna (1994), las de distribuciones log-gamma y las de distribuciones log-gamma inversa. Como se verá, las condiciones son más fáciles de verificar que las que este autor propone.

Sea  $\mathcal{F}_S$  la familia de distribuciones Log-gamma, definidas de la forma siguiente:

$$\mathcal{F}_S = \left\{ F : F(x; \mu, \sigma, k) = \int_{-\infty}^{\frac{x-\mu}{\sigma}} \frac{\exp\{ks - \exp(s)\}}{\Gamma(k)} ds, \mu \in \mathbb{R}, \sigma, k > 0 \right\}.$$

**Teorema 2.6.1.** *La clase de todas las mixturas finitas de  $\mathcal{F}_S$  es identificable.*

*Demostración.* Sea  $\mathcal{H}_{\mathcal{F}_S}$  la clase de todas las mixturas finitas de  $\mathcal{F}_S$ . Se considera la aplicación  $M : \mathcal{H}_{\mathcal{F}_S} \mapsto \Phi$  que transforma una función de distribución  $F \in \mathcal{H}_{\mathcal{F}_S}$  en su función generatriz de momentos, i. e.:

$$M(F)(t) = \phi_F(t) = \int e^{tx} dF(x).$$

Por lo que  $M$  es una aplicación lineal que hace corresponder:

$$M[F(\cdot; \mu, \sigma, k)] = \phi_{(\mu, \sigma, k)}(t) = e^{\mu t} \Gamma(\sigma t + k) / \Gamma(k) \quad t \in (-k/\sigma, +\infty).$$

En este caso,  $S_0(F(\cdot; \mu, \sigma, k)) = (-k/\sigma, +\infty)$ , y por tanto  $+\infty$  es un punto de acumulación de la intersección de un número finito de dominios de definición. Por el Corolario 2.5.1, únicamente hay que probar que

$$F(\cdot; \mu_1, \sigma_1, k_1) \prec F(\cdot; \mu_2, \sigma_2, k_2) \Leftrightarrow \lim_{t \rightarrow +\infty} \frac{\phi_{(\mu_2, \sigma_2, k_2)}(t)}{\phi_{(\mu_1, \sigma_1, k_1)}(t)} = 0$$

define un orden total en  $\mathcal{F}_S$ .

Previamente, es conveniente recordar la fórmula de Stirling, que establece que  $\Gamma(z)$  es asintóticamente equivalente a  $\sqrt{2\pi z}(z/e)^z$  para  $z \rightarrow +\infty$  (es decir el límite del cociente es 1). Con esto, se obtiene que para  $i = 1, 2$ ,  $\phi_{(\mu_i, \sigma_i, k_i)}(t)$  es asintóticamente equivalente a:

$$\begin{aligned} & \frac{e^{\mu_i t}}{\Gamma(k_i)} \sqrt{2\pi} (\sigma_i t + k_i - 1)^{\sigma_i t + k_i - \frac{1}{2}} \exp\{-\sigma_i t - k_i - 1\} = \\ & = \frac{e^{\mu_i t}}{\Gamma(k_i)} \sqrt{2\pi} (\sigma_i t)^{\sigma_i t + k_i - \frac{1}{2}} \left[1 + \frac{k_i - 1}{\sigma_i t}\right]^{\sigma_i t + k_i - \frac{1}{2}} \exp\{-\sigma_i t - k_i - 1\}, \end{aligned} \quad (2.2)$$

y como

$$\lim_{t \rightarrow +\infty} \left[1 + \frac{k_i - 1}{\sigma_i t}\right]^{\sigma_i t + k_i - \frac{1}{2}} = e^{k_i - 1},$$

entonces, (2.2) es asintóticamente equivalente a

$$\frac{\sqrt{2\pi}}{\Gamma(k_i)} \exp\{(\mu_i t - \sigma_i t) + (\sigma_i t + k_i - \frac{1}{2})(\log \sigma_i + \log t)\}$$

y por tanto cuando  $t \rightarrow +\infty$

$$\frac{\phi_{(\mu_2, \sigma_2, k_2)}(t)}{\phi_{(\mu_1, \sigma_1, k_1)}(t)}$$

es asintóticamente equivalente a

$$\begin{aligned} & \frac{\Gamma(k_1) \sigma_2^{1/2} e^{k_1} \exp\{(\sigma_2 t + k_2) \log(\sigma_2 t + k_2)\}}{\Gamma(k_2) \sigma_1^{1/2} e^{k_2} \exp\{(\sigma_1 t + k_1) \log(\sigma_1 t + k_1)\}} \exp\{[(\mu_2 - \mu_1) - (\sigma_2 - \sigma_1)]t\} = \\ & = C \exp\{[(\mu_2 - \mu_1) - (\sigma_2 - \sigma_1) + (\sigma_2 \log \sigma_2 - \sigma_1 \log \sigma_1)]t + \\ & \quad + (\sigma_2 - \sigma_1)t \log t + (k_2 - k_1) \log t\} \end{aligned}$$

con  $C$  una constante positiva.

Así pues se tiene que

$$\lim_{t \rightarrow +\infty} \frac{\phi_{(\mu_2, \sigma_2, k_2)}(t)}{\phi_{(\mu_1, \sigma_1, k_1)}(t)} = 0 \Leftrightarrow \begin{cases} 0 < \sigma_2 < \sigma_1 \\ 0 \\ \sigma_2 = \sigma_1, & \mu_2 < \mu_1 \\ 0 \\ \sigma_2 = \sigma_1, & \mu_2 = \mu_1, & k_2 < k_1. \end{cases}$$

Lo cual es efectivamente un orden total en el espacio paramétrico asociado a la familia de distribuciones  $\mathcal{F}_S$ , por lo que induce un orden total en la propia familia.  $\square$

Del mismo modo se demuestra la identificabilidad de  $\mathcal{F}_R$ , la familia de distribuciones log-gamma inversa, dada por:

$$\mathcal{F}_R = \left\{ R : R(x; \mu, \sigma, k) = \int_{-\infty}^{\frac{x-\mu}{\sigma}} \frac{\exp\{-ks - \exp(-s)\}}{\Gamma(k)} ds, \mu \in \mathbb{R}, \sigma, k > 0 \right\}.$$

**Teorema 2.6.2.** *La clase de todas las mixturas finitas de  $\mathcal{F}_R$  es identificable.*

*Demostración.* La demostración de este teorema es semejante a la anterior considerando de nuevo  $M$ , que asigna a cada mixtura de  $\mathcal{F}_R$  su función generatriz de momentos

$$M[F(\cdot; \mu, \sigma, k)] = \phi_{(\mu, \sigma, k)}(t) = e^{\mu t} \Gamma(-\sigma t + k) / \Gamma(k) \quad t \in (-\infty, k/\sigma)$$

y  $t(F) = -\infty$ .

No obstante, la identificabilidad de esta familia se deduce directamente de la identificabilidad de la familia anterior al observar que cada elemento de  $\mathcal{F}_R$  se puede representar en función de un elemento de  $\mathcal{F}_S$ , así, si en

$$R(x; \mu, \sigma, k) = \int_{-\infty}^{\frac{x-\mu}{\sigma}} \frac{\exp\{-ks - \exp(-s)\}}{\Gamma(k)} ds$$

se realiza el cambio de variable  $[s = -t]$  se obtiene la expresión

$$\int_{-\frac{x+\mu}{\sigma}}^{+\infty} \frac{\exp\{kt - \exp(t)\}}{\Gamma(k)} dt$$

y teniendo en cuenta que se trata de una función de densidad, se obtendrá

$$1 - \int_{-\infty}^{\frac{-x+\mu}{\sigma}} \frac{\exp\{kt - \exp(t)\}}{\Gamma(k)} dt$$

que coincide con

$$1 - F(-x; -\mu, \sigma, k)$$

□

### 2.6.2 Identificabilidad de mixturas de distribuciones Gamma, Lognormal y Weibull

En esta sección se verá la identificabilidad de la familia de mixturas mixtas,  $\mathcal{P}$ , a través de la aplicación del Teorema 2.5.1 a la familia  $\mathcal{H}_{\mathcal{F}_U}$ , que la contiene.

**Teorema 2.6.3.** *La clase,  $\mathcal{H}_{\mathcal{F}_U}$ , de todas las mixturas finitas de las distribuciones de la familia  $\mathcal{F}_U = \mathcal{F}_L \cup \mathcal{F}_G \cup \mathcal{F}_W$  es identificable.*

*Demostración.* Sea  $M : \mathcal{F}_U \mapsto \Phi$  la aplicación que a cada elemento de  $\mathcal{F}_U$  le asigna la función generatriz de momentos del  $\log X$ , es decir:

$$M(F)(t) = \phi_F(t) = \int e^{t \log x} dF(x).$$

La aplicación  $M$  es un isomorfismo lineal que hace corresponder:

$$\begin{aligned} \phi_{F_1(\mu, \sigma)}(t) &= \exp\left(t\mu + \frac{t^2\sigma^2}{2}\right) & t \in (-\infty, +\infty) & \text{si } F_1 \in \mathcal{F}_L \\ \phi_{F_2(a, b)}(t) &= \frac{b^t \Gamma(t+a)}{\Gamma(a)} & t \in (-a, +\infty) & \text{si } F_2 \in \mathcal{F}_G \\ \phi_{F_3(c, d)}(t) &= d^t \Gamma\left(\frac{t}{c} + 1\right) & t \in (-c, +\infty) & \text{si } F_3 \in \mathcal{F}_W \end{aligned}$$

Considerando  $t(F) = +\infty$ , para aplicar el Corolario 2.5.1, basta demostrar la condición (b) de éste. Para ello se verá que cualesquiera dos funciones de esta familia son comparables. De nuevo será de utilidad la fórmula de Stirling.

(i) Si  $F(\mu_1, \sigma_1), G(\mu_2, \sigma_2) \in \mathcal{F}_L$  entonces

$$\frac{\phi_{F(\mu_1, \sigma_1)}(t)}{\phi_{G(\mu_2, \sigma_2)}(t)} = \exp\left\{(\mu_1 - \mu_2)t + (\sigma_1^2 - \sigma_2^2)\frac{t^2}{2}\right\}$$

por lo que

$$G \prec F \text{ si y sólo si } [\sigma_1 < \sigma_2] \text{ ó } [\sigma_1 = \sigma_2 \text{ y } \mu_1 < \mu_2].$$

(ii) Si  $F(a_1, b_1), G(a_2, b_2) \in \mathcal{F}_G$  entonces

$$\frac{\phi_{F(a_1, b_1)}(t)}{\phi_{G(a_2, b_2)}(t)} \text{ es asintóticamente equivalente a}$$

$$\frac{\Gamma(a_2)}{\Gamma(a_1)} \exp \{ (a_1 - a_2) \log t + (\log(b_1/b_2))t \}$$

por lo que

$$G \prec F \text{ si y sólo si } [b_1 < b_2] \text{ ó } [b_1 = b_2 \text{ y } a_1 < a_2].$$

(iii) Si  $F(c_1, d_1), G(c_2, d_2) \in \mathcal{F}_W$  entonces

$$\frac{\phi_{F(c_1, d_1)}(t)}{\phi_{G(c_2, d_2)}(t)} \text{ es asintóticamente equivalente a}$$

$$\left[ \frac{c_2}{c_1} \right]^{1/2} \exp \left\{ t \log \left[ \frac{d_1 (c_2)^{1/c_2}}{d_2 (c_1)^{1/c_1}} \right] + \left[ \frac{1}{c_1} - \frac{1}{c_2} \right] t \log t \right\}$$

por lo que

$$G \prec F \text{ si y sólo si } [c_1 > c_2] \text{ ó } [c_1 = c_2 \text{ y } d_1 < d_2].$$

(iv) Si  $F(a, b) \in \mathcal{F}_G$  y  $G(\mu, \sigma) \in \mathcal{F}_L$  entonces

$$\frac{\phi_{F(a, b)}(t)}{\phi_{G(\mu, \sigma)}(t)} \text{ es asintóticamente equivalente a}$$

$$\frac{\sqrt{2\pi}}{\Gamma(a)} \exp \left\{ \left( a - \frac{1}{2} \right) \log t + (\log b - \mu - 1)t + t \log t - \frac{\sigma^2}{2} t^2 \right\}$$

por lo que su límite para  $t \rightarrow +\infty$  es 0, y por tanto, con independencia de los valores de los parámetros,

$$G \prec F.$$

(v) Si  $F(c, d) \in \mathcal{F}_W$  y  $G(\mu, \sigma) \in \mathcal{F}_L$  entonces

$$\frac{\phi_{F(c, d)}(t)}{\phi_{G(\mu, \sigma)}(t)} \text{ es asintóticamente equivalente a}$$

$$\sqrt{\frac{2\pi}{c}} \exp \left\{ \frac{1}{2} \log t + \left[ \log d - \mu - \frac{1 + \log c}{c} \right] t + \frac{1}{c} t \log t - \frac{\sigma^2}{2} t^2 \right\}$$

por lo que de nuevo su límite para  $t \rightarrow +\infty$  es 0, así que independientemente de los parámetros,

$$G \prec F.$$

(vi) Si  $F(a, b) \in \mathcal{F}_G$  y  $G(c, d) \in \mathcal{F}_W$  entonces

$\frac{\phi_{F(a,b)}(t)}{\phi_{G(c,d)}(t)}$  es asintóticamente equivalente a

$$\frac{c^{1/2}}{\Gamma(a)} \exp \left\{ (a-1) \log t + \left[ \frac{1 + \log c}{c} + \log(b/d) - 1 \right] t + \left[ 1 - \frac{1}{c} \right] t \log t \right\}$$

por lo que

$$G \prec F \text{ si y sólo si } [c < 1] \text{ ó } [c = 1 \text{ y } b < d] \text{ ó } [c = 1, b = d \text{ y } a < 1]$$

y

$$F \prec G \text{ si y sólo si } [c > 1] \text{ ó } [c = 1 \text{ y } b > d] \text{ ó } [c = 1, b = d \text{ y } a > 1].$$

□

**Corolario 2.6.1.**  $\mathcal{P}$  es identificable.

Este resultado se deduce directamente del teorema anterior por ser  $\mathcal{P}$  un subconjunto de mixturas contenido en la clase generada por  $\mathcal{F}_U$ .

# Capítulo 3

## Mixturas finitas de distribuciones: Teoremas de Consistencia

### 3.1 Introducción

En este capítulo se pretende comprobar que en el conjunto  $\mathcal{P}$ , de mixturas mixtas de las familias  $\mathcal{F}_L$ ,  $\mathcal{F}_G$  y  $\mathcal{F}_W$  recogido en la Definición 2.3.1, las soluciones de las ecuaciones de verosimilitud cumplen las condiciones de consistencia dadas en el primer capítulo.

La comprobación de estas condiciones no es, en general, fácil. De hecho, puede resultar muy tediosa ya que es necesario calcular derivadas de hasta tercer orden del logaritmo de la mixtura con respecto a cada uno de sus parámetros que, en el caso que aquí se trata, son ocho. Para facilitar este trabajo se proponen, en este capítulo, algunos resultados que permiten la comprobación de dichas condiciones para las mixturas de uniones de ciertas familias paramétricas de funciones de densidad como son las familias exponenciales que seguidamente se definen.

**Definición 3.1.1.** *Dadas las funciones:*

$$a : D \subset \mathbb{R}^n \rightarrow [0, +\infty),$$

$$t : D \subset \mathbb{R}^n \rightarrow \mathbb{R}^d \text{ y}$$

$$b : \Theta \rightarrow (0, +\infty),$$



con  $\Theta \subset \mathbb{R}^d$  abierto y  $D \subset \mathbb{R}^n$  medible, tales que las funciones

$$f(\cdot, \theta) : x \in D \rightarrow a(x)b(\theta)e^{\theta^T t(x)},$$

son funciones de densidad  $\forall \theta \in \Theta$ , el conjunto dado por

$$\mathcal{E}(a, t, b) = \{f(\cdot, \theta) : \theta \in \Theta\}.$$

se denomina familia exponencial regular asociada a las funciones  $a, t$  y  $b$ .

Las familias de distribuciones Lognormal y Gamma son ejemplos de familias que, reparametrizadas convenientemente, son exponenciales regulares. Sin embargo, teniendo en cuenta que este trabajo también se centra en las mixturas mixtas de  $\mathcal{P}$ , se necesita una familia más amplia que incluya a las distribuciones Weibull por lo que, a continuación, se define una nueva familia de distribuciones que contiene, en particular, a las familias exponenciales, y dentro de la cual se pueden incluir también a la familia de distribuciones Weibull.

**Definición 3.1.2.** Dadas las funciones:

$$a : D \subset \mathbb{R}^n \rightarrow [0, +\infty),$$

$$t : D \subset \mathbb{R}^n \rightarrow \mathbb{R}^{d_1},$$

$$b : \Theta \rightarrow (0, +\infty) \text{ y}$$

$$c : \Delta_2 \subset \mathbb{R}^{d_2} \mapsto (0, +\infty)$$

con  $\Theta = \Delta_1 \times \Delta_2 \times \Delta_3 \subset \mathbb{R}^{d_1} \times \mathbb{R}^{d_2} \times \mathbb{R}^{d_1}$  abierto y  $D \subset \mathbb{R}^n$  medible, tales que las funciones

$$f(\cdot, \theta) : x \in D \rightarrow a(x)b(\theta) \exp\{\alpha^T t(x) + c(\beta)e^{\lambda^T t(x)}\}$$

son funciones de densidad  $\forall \theta = (\alpha, \beta, \lambda) \in \Theta$ , el conjunto dado por

$$\mathcal{W}(a, t, b, c) = \{f(\cdot, \theta) : \theta = (\alpha, \beta, \lambda) \in \Theta\}$$

se denomina familia tipo  $\mathcal{W}$  asociada a las funciones  $a, t, b$  y  $c$ .

Se observa que el conjunto  $\mathcal{P}$ , objeto de estudio en este trabajo, es un subconjunto de las mixturas de la unión de familias tipo  $\mathcal{W}$ .

Con idea de facilitar la comprensión y demostración de los teoremas, el capítulo se ha estructurado como a continuación se detalla. En la segunda sección se presentan una serie de resultados de integrabilidad que son necesarios en las sucesivas demostraciones que aparecen en el resto de las secciones. Desde la sección tercera

hasta la sexta, el interés se centra en ciertos teoremas que permiten la comprobación de las condiciones de consistencia **C1**, **C3** y **C4**, recogidas en el apartado 1.4 para, en primer lugar, las mixturas de familias exponenciales, en segundo lugar, las mixturas de uniones finitas de familias exponenciales, en tercer lugar, las mixturas de familias tipo  $\mathcal{W}$ , y por último, para las mixturas finitas de familias tipo  $\mathcal{W}$ .

En la sección séptima, se comprueba que  $\mathcal{P}$  verifica dichas condiciones de consistencia, aplicando los resultados anteriores y que también verifica la condición **C2** referente a la matriz de información.

## 3.2 Resultados de integrabilidad

En esta sección, se dan diferentes proposiciones, corolarios y lemas que facilitarán las demostraciones de los teoremas que se abordan en las siguientes secciones.

El objetivo de la siguiente proposición es mostrar que, bajo ciertas condiciones, la derivada paramétrica de una función integrable sigue siendo integrable.

**Proposición 3.2.1.** *Sea  $I \subset \mathbb{R}$  un intervalo compacto,  $D \subset \mathbb{R}^n$  medible y sea  $f : D \times I \rightarrow \mathbb{R}$  tal que:*

- (1)  $x \in D \mapsto f(x, \theta)$  es integrable  $\forall \theta \in I$ ;  
 (2)  $\exists \frac{\partial f}{\partial \theta}(x; \theta) \quad \forall \theta \in I$ , e.c.t.  $x \in D$ , verificando:

(2a)  $(x, \theta) \mapsto \frac{\partial f}{\partial \theta}(x, \theta)$  es medible, y

(2b) e.c.t.  $x \in D$ ,

$$\theta \in I \mapsto \frac{\partial f}{\partial \theta}(x, \theta)$$

es integrable en  $I$ ;

- (3) el conjunto  $Z = \{x : \exists \theta_1, \theta_2 \in I / \frac{\partial f}{\partial \theta}(x, \theta_1) > 0 \text{ y } \frac{\partial f}{\partial \theta}(x, \theta_2) < 0\}$  es de medida nula.

Entonces, e.c.t.  $\theta \in I$ , la función

$$x \in D \mapsto \frac{\partial f}{\partial \theta}(x, \theta)$$

es integrable.

*Demostración.* Se considera la siguiente partición del conjunto  $D$ ,

$$D = A^+ \cup A^- \cup Z$$

donde

$$A^+ = \{x \in D : \frac{\partial f}{\partial \theta}(x, \theta) \geq 0 \quad \forall \theta \in I\}$$

y

$$A^- = \{x \in D : \frac{\partial f}{\partial \theta}(x, \theta) \leq 0 \quad \forall \theta \in I\}.$$

Si  $x \in A^+$ , por el Teorema Fundamental del Cálculo y la hipótesis (2), se verifica que

$$\int_I \left| \frac{\partial f}{\partial \theta}(x, \theta) \right| d\theta = \int_I \frac{\partial f}{\partial \theta}(x, \theta) d\theta = f(x, b) - f(x, a) \leq |f(x, b)| + |f(x, a)|$$

siendo  $I = [a, b]$ .

Si  $x \in A^-$ , análogamente

$$\int_I \left| \frac{\partial f}{\partial \theta}(x, \theta) \right| d\theta = - \int_I \frac{\partial f}{\partial \theta}(x, \theta) d\theta = f(x, a) - f(x, b) \leq |f(x, b)| + |f(x, a)|.$$

En cualquier caso, para todo  $x \in D \setminus Z$ ,

$$\int_I \left| \frac{\partial f}{\partial \theta}(x, \theta) \right| d\theta \leq |f(x, b)| + |f(x, a)|.$$

Como las funciones  $x \mapsto f(x, a)$  y  $x \mapsto f(x, b)$  son integrables por (1), se tiene

$$\int_D \left( \int_I \left| \frac{\partial f}{\partial \theta}(x, \theta) \right| d\theta \right) dx < +\infty.$$

Aplicando el Teorema de Tonelli, que asegura que una función medible es integrable si una integral reiterada de su valor absoluto es finita, se tiene que

$$(x, \theta) \in D \times I \mapsto \frac{\partial f}{\partial \theta}(x, \theta)$$

es integrable. Basta aplicar el teorema de Fubini para afirmar que las integrales reiteradas existen y coinciden y, por tanto, se obtiene que la función

$$x \in D \mapsto \frac{\partial f}{\partial \theta}(x, \theta)$$

es integrable para casi todo  $\theta \in I$ . □

*Nota 3.2.1.* La tercera condición es una hipótesis esencial para la prueba. A continuación, se presenta una función que, cumpliendo el resto de las hipótesis, no verifica la tesis del teorema.

Se considera la función

$$f(x, \theta) = \frac{1 + \operatorname{sen}(\theta x)}{x^2} \text{ con } x \geq 1, \quad \theta \in [0, \frac{\pi}{2}].$$

La función

$$x \in [1, +\infty) \rightarrow \frac{1 + \operatorname{sen}(\theta x)}{x^2}$$

está acotada por  $\frac{2}{x^2}$ , que es integrable (verifica (1)).

Además, existe la derivada parcial

$$\frac{\partial f}{\partial \theta} = \frac{x \cos(\theta x)}{x^2} = \frac{\cos(\theta x)}{x} \quad \theta \in [0, \frac{\pi}{2}], \forall x \in [1, +\infty)$$

y es integrable (respecto de  $\theta \in [0, \frac{\pi}{2}]$ ) para todo  $x \in [1, +\infty)$ , por ser continua (verifica (2)).

Nótese que  $A^+ = \{1\}$ , ya que

$$\frac{\partial f}{\partial \theta}(x, \theta) \geq 0 \quad \forall \theta \in [0, \frac{\pi}{2}] \Leftrightarrow \cos(\theta x) \geq 0 \quad \forall \theta \in [0, \frac{\pi}{2}] \Leftrightarrow x \leq 1.$$

y  $A^- = \emptyset$ , pues  $\frac{\partial f}{\partial \theta}(x, 0) = 1$  para todo  $x \in [1, +\infty)$ .

Luego en este caso,  $Z = (1, +\infty)$  no verifica la tercera condición y de hecho la función  $x \rightarrow \frac{\partial f}{\partial \theta} = \frac{\cos(\theta x)}{x}$  no es integrable respecto a  $x$  en  $[1, +\infty)$  para ningún valor de  $\theta \in [0, \frac{\pi}{2}]$ .  $\square$

*Nota 3.2.2.* La tesis de la proposición es para casi todo  $\theta \in I$ , ya que en general no es cierto para todo. Por ejemplo, la función

$$f(x, \theta) = \frac{1}{x^\theta \log^2 x} \text{ con } x \in [e, +\infty), \theta \in [1, +\infty)$$

verifica las condiciones de la proposición, sin embargo,

$$\frac{\partial f}{\partial \theta}(x, \theta) = -\frac{\log x}{x^\theta \log^2 x} = \frac{-1}{x^\theta \log x}$$

sólo es integrable si  $\theta \in (1, +\infty)$ .  $\square$

Los siguientes resultados se centran en el estudio de la integrabilidad de funciones relacionadas con la familia exponencial la cual se utilizará de puente para las mixturas objeto de esta memoria. Para ello es necesario definir la siguiente familia de funciones:

**Definición 3.2.1.** Dada  $t : \mathbb{R}^n \rightarrow \mathbb{R}^d$  con  $t(x) = (t_1(x) \dots t_d(x))^T$  y  $h$  un entero no negativo, se notará por  $\Pi_h(t)$  a la familia de funciones dada por

$$\Pi_h(t) = \{T : D \subset \mathbb{R}^n \rightarrow \mathbb{R} : T(x) = \prod_{i=1}^d t_i^{\nu_i}(x), \nu_i \in \mathbb{Z}^+, \sum_{i=1}^d \nu_i \leq h\},$$

y se le denominará familia de funciones producto de a lo más  $h$  componentes de  $t(x)$ .

**Proposición 3.2.2.** Sea  $\Theta \subset \mathbb{R}^d$  abierto y  $D \subset \mathbb{R}^n$  medible. Sea  $f : D \times \Theta \rightarrow \mathbb{R}$  tal que  $f(x; \theta) = a(x)e^{\theta^T t(x)}$  donde  $a : D \subset \mathbb{R}^n \rightarrow \mathbb{R}$  y  $t : D \subset \mathbb{R}^n \rightarrow \mathbb{R}^d$  son dos funciones medibles tales que para todo  $\theta \in \Theta$  la función

$$x \in D \mapsto f(x; \theta)$$

es integrable en  $D$ . Entonces la función  $T(x)f(x; \theta)$  es integrable en  $D$ ,  $\forall \theta \in \Theta, \forall h \in \mathbb{Z}^+ \cup \{0\}, \forall T \in \Pi_h(t)$ .

*Demostración.* Cualquier producto  $T(x)f(x, \theta)$  con  $T \in \Pi_h(t)$  es de la forma

$$\left[ \prod_{i=1}^d t_i^{\nu_i}(x) \right] f(x, \theta),$$

que se puede escribir como

$$\frac{\partial^{h'}}{\partial \theta_1^{\nu_1} \dots \partial \theta_k^{\nu_k}} f(x, \theta)$$

donde  $h' = \sum_{i=1}^d \nu_i \leq h$ .

Al derivar respecto de una de las componentes del parámetro se obtiene una función de la forma

$$t_i(x)f(x, \theta),$$

que puede considerarse como una nueva función del tipo de  $f$  con  $a(x)t_i(x)$  jugando el papel de  $a(x)$ . Así, para demostrar el resultado, bastará comprobar que estas derivadas respecto de cualesquiera de las componentes son integrables e ir aplicando sucesivamente el mismo razonamiento tantas veces como sea necesario. Sin pérdida de generalidad se demostrará para la última componente del parámetro que se denotará por  $\lambda$ .

### 3.2. RESULTADOS DE INTEGRABILIDAD

Dado  $\theta^* = (\beta^*, \lambda^*) \in \mathbb{R}^{d-1} \times \mathbb{R}$  fijo, se considera la función

$$f_{\beta^*}(x, \lambda) : D \times [\lambda_1, \lambda_2] \rightarrow \mathbb{R}$$

$$f_{\beta^*}(x, \lambda) = f(x, \beta^*, \lambda)$$

donde  $\lambda_1, \lambda_2 \in \mathbb{R}$  tales que  $\lambda^* \in [\lambda_1, \lambda_2]$  y  $(\beta^*, \lambda) \in \Theta \quad \forall \lambda \in [\lambda_1, \lambda_2]$ . Esta función, trivialmente, verifica la condición (1) de la Proposición 3.2.1. Además la derivada de  $f_{\beta^*}(x, \lambda)$  respecto de  $\lambda$  es

$$a(x)t_d(x) \exp\{(\beta^*)^T t_{(d)}(x) + \lambda t_d(x)\},$$

con  $t_{(d)}(x) = (t_1(x) \dots t_{d-1}(x))^T$ , que existe y es medible por serlo  $a(x)$  y  $t(x)$ .

Para comprobar (2) quedaría por demostrar que la función

$$\lambda \in [\lambda_1, \lambda_2] \rightarrow a(x)t_d(x) \exp\{(\beta^*)^T t_{(d)}(x) + \lambda t_d(x)\}$$

es integrable en  $[\lambda_1, \lambda_2]$ , y lo es por ser continua respecto de  $\lambda$  en un intervalo cerrado.

De este modo, se obtiene por la Proposición 3.2.1 que para casi todo  $\lambda \in [\lambda_1, \lambda_2]$ , la función

$$x \in D \rightarrow \frac{\partial f_{\beta^*}}{\partial \lambda}(x, \lambda)$$

es integrable.

Sin embargo, la integrabilidad en casi todo  $[\lambda_1, \lambda_2]$  no asegura la integrabilidad en el punto  $\lambda^*$ . Para realizar la comprobación en el punto  $\lambda^*$  basta con elegir  $\lambda', \lambda'' \in \Theta$  tales que  $\lambda^* \in [\lambda', \lambda''] \subset [\lambda_1, \lambda_2]$  y de forma que tanto  $\frac{\partial f_{\beta^*}}{\partial \lambda}(x, \lambda')$  como  $\frac{\partial f_{\beta^*}}{\partial \lambda}(x, \lambda'')$  sean integrables respecto de  $x$ .

Usando, ahora, la desigualdad

$$\begin{aligned} \left| \frac{\partial f_{\beta^*}}{\partial \lambda_i}(x, \lambda^*) \right| &= a(x)t_d(x) e^{\beta^* t_{(d)} + t_d(x)\lambda^*} \\ &\leq |a(x)t_d(x) e^{\beta^* t_{(d)} + t_d(x)\lambda'}| + |a(x)t_d(x) e^{\beta^* t_{(d)} + t_d(x)\lambda''}| \end{aligned}$$

se tiene la integrabilidad, (realmente es menor que uno de los sumandos, dependiendo de si  $t_d(x) < 0$  o  $t_d(x) \geq 0$ ).

□

Notación. Se utilizará la expresión  $Q_d(\theta^*, r) = \prod_{i=1}^d [\theta_i^* - r, \theta_i^* + r]$  para denotar al hipercubo de centro  $\theta^*$  y semiamplitud  $r$ .

**Proposición 3.2.3.** *Sea  $\Theta = \Delta_1 \times \Delta_2 \times \Delta_3 \subset \mathbb{R}^{d_1} \times \mathbb{R}^{d_2} \times \mathbb{R}^{d_3}$  abierto y  $D \subset \mathbb{R}^n$  medible. Sean  $a : D \subset \mathbb{R}^n \rightarrow \mathbb{R}$  y  $t : D \subset \mathbb{R}^n \rightarrow \mathbb{R}^{d_1}$  dos funciones medibles tales que para todo  $\theta = (\alpha, \beta, \lambda) \in \Theta$  la función*

$$f(\cdot, \alpha, \beta, \lambda) : x \in D \mapsto a(x) \exp\{\alpha_t t(x) + c(\beta) e^{\lambda^T t(x)}\}$$

es integrable en  $D$  y donde  $c : \Delta_2 \rightarrow (0, +\infty)$  es cualquier función que tiene al menos una derivada parcial de primer orden continua y no nula en  $\Delta_2$ .

Entonces, para todo  $\theta \in \Theta$ , para todo  $h$  y para todo  $T \in \Pi_h(t)$  se verifica:

1) Las funciones  $T(x) e^{m\lambda^T t(x)} f(x, \alpha, \beta, \lambda)$ , con  $m$  entero no negativo, son integrables en  $D$

2) Dado  $r > 0$  tal que  $Q_{d_1}(\alpha, 3r) \subset \Delta_1$ , las funciones

$$T(x) e^{m\lambda_1^T t(x)} f(x, \alpha, \beta, \lambda),$$

con  $0 \leq m \leq 3$ , son integrables en  $D$  para todo  $\lambda_1 \in Q_{d_1}(\lambda, r)$ .

*Demostración.* Para demostrar el apartado 1) se supondrá que la derivada parcial de primer orden que es continua y no nula en  $\Delta_2$  es  $\frac{\partial c}{\partial \beta_1}$ .

Por un razonamiento análogo al de la demostración de la Proposición 3.2.2, derivando respecto de las componentes de  $\alpha$  se obtiene que, para todo  $i$ , las funciones

$$\frac{\partial f}{\partial \alpha_i}(x, \alpha, \beta, \lambda) = t_i(x) f(x, \alpha, \beta, \lambda),$$

son integrables.

Derivando respecto de la componente  $\beta_1$ , para la cual la derivada de  $c$  es continua y distinta de 0, y por tanto  $c$  es monótona en esa componente, se obtiene, por la Proposición 3.2.1, que la función

$$\frac{\partial f}{\partial \beta_1}(x, \alpha, \beta, \lambda) = e^{\lambda^T t(x)} \frac{\partial c}{\partial \beta_1}(\beta) f(x, \alpha, \beta, \lambda)$$

es integrable para casi todo  $\beta_1$ , y usando la monotonía de  $c$  y un razonamiento análogo al realizado en la demostración de la Proposición 3.2.2 se obtiene que es

integrable para todo  $\beta_1$ . De donde la función

$$e^{\lambda^T t(x)} f(x, \alpha, \beta, \lambda)$$

también lo es.

Para demostrar que las funciones

$$T(x) e^{m\lambda^T t(x)} f(x, \alpha, \beta, \lambda)$$

con  $T \in \Pi_h(t)$  son integrables, basta aplicar el mismo argumento a las derivadas parciales de  $f(x, \alpha, \beta, \lambda)$  respecto de  $\beta_1$   $m$  veces y respecto de las componentes de  $\alpha$  correspondientes a los factores de  $T$ , divididas previamente por  $\frac{\partial c}{\partial \beta_1} (\neq 0)$ , el número de veces que sea necesario.

Para demostrar 2), basta ver que las funciones  $T(x) e^{m\lambda_1^T t(x)} f(x, \alpha, \beta, \lambda)$  son de la forma:

$$T(x) a(x) \exp\{(\alpha + m\lambda_1)^T t(x) + c(\beta) e^{\lambda^T t(x)}\}.$$

Sumando y restando  $m\lambda^T t(x)$  en el exponente, se obtiene:

$$\begin{aligned} T(x) a(x) \exp\{(\alpha + m\lambda_1 - m\lambda)^T t(x) + m\lambda^T t(x) + c(\beta) e^{\lambda^T t(x)}\} &= \\ = T(x) e^{m\lambda^T t(x)} a(x) \exp\{(\alpha + m(\lambda_1 - \lambda))^T t(x) + c(\beta) e^{\lambda^T t(x)}\} &= \\ = T(x) e^{m\lambda^T t(x)} f(x, \alpha + m(\lambda_1 - \lambda), \beta, \lambda) \end{aligned}$$

donde  $(\lambda_1 - \lambda)m < 3r$  y por tanto  $\alpha + m(\lambda_1 - \lambda) \in \Delta_1$ , por lo que aplicando el apartado 1) a esta nueva función se obtiene el resultado propuesto.

□

### 3.3 Condiciones de consistencia para mixturas de familias exponenciales

En esta sección se aborda el problema de comprobar que, bajo ciertas hipótesis muy generales, las mixturas de una misma familia exponencial regular cumplen las condiciones C1, C3 y C4 de consistencia presentadas en el apartado 1.4. y que se recuerdan a continuación.



C1. Para todo  $\Psi \in \Omega$ , c.p.t.  $x \in \mathbb{R}^p$  y para  $i, j, s = 1, \dots, \nu$ , las derivadas parciales,  $\partial q / \partial \psi_j$ ,  $\partial^2 q / \partial \psi_j \partial \psi_i$  y  $\partial^3 q / \partial \psi_j \partial \psi_i \partial \psi_s$ , existen y verifican las desigualdades:

$$\left| \frac{\partial q(x; \Psi)}{\partial \psi_j} \right| \leq h_j(x), \quad \left| \frac{\partial^2 q(x; \Psi)}{\partial \psi_j \partial \psi_i} \right| \leq h_{ij}(x), \quad \left| \frac{\partial^3 \log q(x; \Psi)}{\partial \psi_j \partial \psi_i \partial \psi_s} \right| \leq h_{ijs}(x)$$

donde  $h_j$  y  $h_{ij}$  son integrables y  $h_{ijs}$  satisfacen:

$$\int h_{ijs}(x) q(x; \Psi^*) dx < +\infty$$

C3. Para cada  $\Psi \in \Omega$  y para  $r > 0$  suficientemente pequeño,

$$\int \log q_r^*(x; \Psi) q(x; \Psi^*) dx < +\infty$$

C4.

$$\int |\log q(x; \Psi^*)| q(x; \Psi^*) dx < +\infty.$$

donde

$$q_r(x; \Psi) = \sup_{\Psi' \in N_r(\Psi)} q(x; \Psi') \quad \text{y} \quad q_r^*(x; \Psi) = \max\{1, q_r(x; \Psi)\}.$$

En Redner y Walker [46] se recogen algunas hipótesis bajo la cuales las mezclas de funciones de una familia exponencial regular cumplen algunas de las condiciones de consistencia. Sin embargo, no todas estas familias las verifican como se ilustra en el ejemplo que se presenta a continuación.

*Ejemplo 3.3.1.* Se considera la familia de funciones de densidad perteneciente a la familia exponencial regular dada por

$$\mathcal{F} = \left\{ f(x, \theta) = \frac{\theta - 1}{x(\log \frac{1}{x})^\theta} : x \in (0, 1/e], \quad \theta \in (1, +\infty) \right\}.$$

Esta es una familia exponencial regular donde  $a(x) = \frac{1}{x}$ ,  $b(\theta) = \theta - 1$  y  $t(x) = -\log \log \frac{1}{x}$ .

La mezcla,

$$p(x, \Psi) = \frac{1}{2} f(x, \frac{3}{2}) + \frac{1}{2} f(x, 4),$$

no verifica la condición C3 de consistencia. Para su comprobación y por comodidad, se denotarán a las componentes por  $f_1$  y  $f_2$  respectivamente.

Para comprobar que no existe un  $\Omega$  conteniendo a  $\Psi^* = (\frac{1}{2}, \frac{3}{2}, 4)$  tal que, para todo  $\Psi \in \Omega$ , se cumpla la condición **C3**, se verá que no se verifica en el punto  $\Psi = \Psi^*$ .

$$\begin{aligned} \int_0^{\frac{1}{e}} \log p_r^*(x; \Psi^*) p(x; \Psi^*) dx &= \int_0^{\frac{1}{e}} \log \left[ \max \left\{ 1, \sup_{\Psi' \in N_r(\Psi^*)} p(x, \Psi') \right\} \right] p(x, \Psi^*) dx \\ &\geq \int_0^{\frac{1}{e}} \log [\max \{1, p(x, \Psi^*)\}] p(x, \Psi^*) dx. \end{aligned}$$

Por ser  $\lim_{x \rightarrow 0} p(x, \Psi^*) = +\infty$ , existe un  $x_0 \in (0, \frac{1}{e})$  tal que  $p(x, \Psi^*) > 1$  para todo  $x \in (0, x_0)$ , y como el integrando es positivo, la expresión anterior es mayor o igual que

$$\int_0^{x_0} \log [\max \{1, p(x, \Psi^*)\}] p(x, \Psi^*) dx = \int_0^{x_0} \log [p(x, \Psi^*)] p(x, \Psi^*) dx.$$

Por otra parte, como el logaritmo es creciente, se tiene que

$$\begin{aligned} \int_0^{x_0} \log \left[ \frac{1}{2} f_1 + \frac{1}{2} f_2 \right] \left( \frac{1}{2} f_1 + \frac{1}{2} f_2 \right) dx &\geq \int_0^{x_0} \log \left[ \frac{f_1}{2} \right] f_1 dx \\ &= \int_0^{x_0} f_1 \log f_1 dx - \int_0^{x_0} f_1 \log 2 dx. \end{aligned}$$

La segunda integral trivialmente es finita y la primera es de la forma

$$\begin{aligned} &\int_0^{x_0} a(x) b \left( \frac{3}{2} \right) e^{\frac{3}{2}t(x)} \log \left[ a(x) b \left( \frac{3}{2} \right) e^{\frac{3}{2}t(x)} \right] dx \\ &= b \left( \frac{3}{2} \right) \left[ \int_0^{x_0} \log a(x) a(x) e^{\frac{3}{2}t(x)} dx + \log b \left( \frac{3}{2} \right) \int_0^{x_0} a(x) e^{\frac{3}{2}t(x)} dx \right] \\ &\quad + b \left( \frac{3}{2} \right) \left[ \frac{3}{2} \int_0^{x_0} t(x) a(x) e^{\frac{3}{2}t(x)} dx \right]. \end{aligned}$$

La segunda de las integrales que aparecen en la expresión anterior es finita pues es la integral de  $\frac{f_1}{b(3/2)}$  en un intervalo más pequeño que su dominio, la tercera integral es finita por la Proposición 3.2.2, y la primera integral es

$$\int_0^{x_0} \frac{\frac{3}{2} - 1}{x(\log \frac{1}{x})^{\frac{3}{2}}} \log \frac{1}{x} dx = \int_0^{x_0} \frac{\frac{1}{2}}{x(\log \frac{1}{x})^{\frac{1}{2}}} dx = +\infty.$$

Por tanto  $p(x, \Psi)$  no cumple la condición **C3** en  $\Psi^*$ .

Notación. Se denota por  $V_d$  al conjunto

$$V_d = \{\varepsilon \in \mathbb{R}^d / \varepsilon_i = \pm 1 \quad \forall i = 1, \dots, d\}$$

Además, al igual que en la sección anterior, se denotará por  $Q_d(\theta^*, r)$  al hipercubo  $\prod_{i=1}^d [\theta_i^* - r, \theta_i^* + r]$  y para cada  $\varepsilon \in V_d$  se denota por  $\theta(\varepsilon) \in \mathbb{R}^d$ , con coordenadas  $\theta_i^* + \varepsilon_i r$ , los vértices de dicho hipercubo.

**Lema 3.3.1.** *Sea  $t: D \subset \mathbb{R}^n \rightarrow \mathbb{R}^d$ ,  $\theta^* \in \mathbb{R}^d$  y  $r > 0$ . Entonces para todo  $x \in D$  y  $\forall \theta \in Q_d(\theta^*, r)$ ,*

$$e^{\theta^T t(x)} \leq \sum_{\varepsilon \in V_d} e^{\theta^T(\varepsilon)t(x)}.$$

*Demostración.* Dado  $x \in D$ , se considera  $\varepsilon(x) = [\varepsilon_1(x) \dots \varepsilon_d(x)]^T$  tal que

$$\varepsilon_i(x) = \begin{cases} 1 & \text{si } t_i(x) \geq 0 \\ -1 & \text{si } t_i(x) < 0, \end{cases}$$

que, obviamente, pertenece a  $V_d$ .

Entonces  $\forall \theta \in Q_d(\theta^*, r)$

$$t_i(x)\theta_i \leq t_i(x)[\theta_i^* + \varepsilon_i(x)r] \quad \forall i = 1, \dots, d \Rightarrow$$

$$\theta^T t(x) \leq [\theta(\varepsilon(x))]^T t(x) \Rightarrow$$

$$\exp\{\theta^T t(x)\} \leq \exp\{[\theta(\varepsilon(x))]^T t(x)\} \leq \sum_{\varepsilon \in V_d} \exp\{(\theta(\varepsilon))^T t(x)\}.$$

□

**Teorema 3.3.1.** *Sea  $\mathcal{E}(a, t, b)$  una familia exponencial regular,  $k \in \mathbb{N}$  y sea  $\psi^* = (\pi^*, \theta_1^*, \dots, \theta_k^*) \in \Omega = \mathcal{C}_k \times \Theta^k$ . Si se verifican:*

(H1)  $b$  continua

(H2)  $b$  admite derivadas parciales hasta de tercer orden continuas en  $\Theta$ ,

entonces cualquier mixtura finita definida a través de  $k$  componentes de  $\mathcal{E}(a, t, b)$  verifica la condición de consistencia C1 en un cierto compacto  $\hat{\mathcal{C}}_k \times \hat{\Theta}_1 \times \dots \times \hat{\Theta}_k \subset \mathcal{C}_k \times \Theta^k$  que contiene a  $\psi^*$ .

*Demostración.* Dado que  $\Theta$  es abierto, según la definición de la familia  $\mathcal{E}$ ,

$$\exists r > 0 : Q_d(\theta_i^*, r) \subset \Theta \quad \forall i = 1, \dots, k$$

Así, eligiendo  $\hat{\mathcal{C}}_k \subset \mathcal{C}_k$  un compacto que contenga a  $\pi^*$ , se tiene que el subconjunto  $\hat{\mathcal{C}}_k \times Q_d(\theta_1^*, r) \times \dots \times Q_d(\theta_k^*, r) \subset \mathcal{C}_k \times \Theta^k$  es un compacto que contiene a  $\psi^*$ .

Dada una mixtura  $p(x; \psi) = \sum_{i=1}^k \pi_i f(x, \theta_i)$  de  $\mathcal{E}(a, \ell, \bar{\ell})$ , considerándola como función sobre dicho compacto, se ha de demostrar que:

$$(a) \left| \frac{\partial p(x, \psi)}{\partial \psi_j} \right| \leq h_j(x) \quad \forall j$$

$$(b) \left| \frac{\partial^2 p(x, \psi)}{\partial \psi_j \partial \psi_s} \right| \leq h_{js}(x) \quad \forall j, s$$

donde  $h_j(x), h_{js}(x)$  son funciones integrables e independientes del parámetro  $\psi$ , y que

$$(c) \left| \frac{\log p(x; \psi)}{\partial \psi_j \partial \psi_s \partial \psi_u} \right| \leq h_{jsu}(x) \quad \forall j, s, u$$

donde  $h_{jsu}$  son funciones independientes de  $\psi$  tales que

$$\int h_{jsu}(x) p(x; \psi^*) < +\infty.$$

Por la linealidad de la derivada, la desigualdad triangular y dada la simetría para cada índice  $i = 1, \dots, k$ , para demostrar (a) y (b) bastará comprobar que  $\forall \theta \in Q_d(\theta^*, r)$  se verifica que:

$$(1) |f(x, \theta)| \leq g(x)$$

$$(2) \left| \frac{\partial f(x, \theta)}{\partial \theta_j} \right| \leq g_j(x) \quad \forall j$$

$$(3) \left| \frac{\partial^2 f(x, \theta)}{\partial \theta_j \partial \theta_s} \right| \leq g_{js}(x) \quad \forall j, s$$

con  $g, g_j, g_{js}$  funciones integrables respecto a  $x$  e independientes de los parámetros.

Dadas las propiedades exigidas a la función  $b$ , existe  $M > 0$  tal que para todo  $\theta \in \cup_{i=1}^k Q_d(\theta_i^*, r)$

$$|b(\theta)|, \quad \left| \frac{\partial b(\theta)}{\partial \theta_i} \right|, \quad \left| \frac{\partial^2 b(\theta)}{\partial \theta_i \partial \theta_j} \right|, \quad \left| \frac{\partial^3 b(\theta)}{\partial \theta_i \partial \theta_j \partial \theta_u} \right| \leq M.$$

A continuación, se procede a demostrar los puntos (1), (2) y (3).

(1) Por el Lema 3.3.1, para todo  $\theta \in \bigcup_{i=1}^k Q_d(\theta_i^*, r)$  se tiene

$$|f(x, \theta)| \leq M \sum_{i=1}^k \sum_{\varepsilon \in V_d} a(x) e^{(\theta_i(\varepsilon))Tt(x)} = g(x),$$

siendo  $\theta_i(\varepsilon) = \theta_i^* + r\varepsilon$ , con  $g(x)$  integrable por la Proposición 3.2.2 e independiente de  $\theta$ .

(2) Por la regla de derivación del producto y la desigualdad triangular, para todo  $\theta \in \bigcup_{i=1}^k Q_d(\theta_i^*, r)$

$$\begin{aligned} \left| \frac{\partial f(x, \theta)}{\partial \theta_j} \right| &\leq \left| a(x) \frac{\partial b(\theta)}{\partial \theta_j} e^{\theta^T t(x)} \right| + \left| a(x) b(\theta) t_j(x) e^{\theta^T t(x)} \right| \leq \\ &\leq M \left| a(x) e^{\theta^T t(x)} \right| + M \left| a(x) t_j(x) e^{\theta^T t(x)} \right| \leq \dots \end{aligned}$$

usando el Lema 3.3.1

$$\dots \leq M \sum_{i=1}^k \left[ \sum_{\varepsilon \in V_d} \left| a(x) e^{(\theta_i(\varepsilon))Tt(x)} \right| + \sum_{\varepsilon \in V_d} \left| a(x) t_j(x) e^{(\theta_i(\varepsilon))Tt(x)} \right| \right] = g_j(x)$$

con  $g_j(x)$  integrable, por la Proposición 3.2.2, e independiente de  $\theta$ .

(3) Este punto se demuestra por un razonamiento similar al anterior y aplicando propiedades de derivación, el Lema 3.3.1 y la Proposición 3.2.2.

Finalmente para demostrar (c), se debe tener en cuenta que  $\frac{\partial^3 \log p}{\partial \psi_i \partial \psi_j \partial \psi_u}$  se puede expresar como:

$$\frac{1}{p} \frac{\partial^3 p}{\partial \psi_i \partial \psi_j \partial \psi_u} - \frac{1}{p^2} \left[ \frac{\partial^2 p}{\partial \psi_i \partial \psi_j} \frac{\partial p}{\partial \psi_u} + \frac{\partial^2 p}{\partial \psi_j \partial \psi_u} \frac{\partial p}{\partial \psi_i} + \frac{\partial^2 p}{\partial \psi_i \partial \psi_u} \frac{\partial p}{\partial \psi_j} \right] +$$

$$+ \frac{2}{p^3} \frac{\partial p}{\partial \psi_i} \frac{\partial p}{\partial \psi_j} \frac{\partial p}{\partial \psi_u}. \quad (3.1)$$

Los valores absolutos de los cocientes entre las derivadas parciales de  $p$  de hasta tercer orden y  $p$ :

$$\left| \frac{1}{p} \frac{\partial^3 p}{\partial \psi_i \partial \psi_j \partial \psi_u} \right|, \quad \left| \frac{1}{p} \frac{\partial^2 p}{\partial \psi_i \partial \psi_j} \right| \quad \text{y} \quad \left| \frac{1}{p} \frac{\partial p}{\partial \psi_i} \right|$$

están acotados por una suma de cocientes de la forma:

$$\left| \frac{1}{p} \frac{\partial^3 f}{\partial \theta_i \partial \theta_j \partial \theta_u} \right|, \quad \left| \frac{1}{p} \frac{\partial^2 f}{\partial \theta_i \partial \theta_j} \right|, \quad \left| \frac{1}{p} \frac{\partial f}{\partial \theta_i} \right| \quad \text{ó} \quad \frac{f}{p}.$$

En consecuencia, dicha suma se puede escribir como suma de funciones de la forma

$$\left| \frac{f}{p} T(x) B(\theta) \right|$$

con  $T \in \Pi_h(t)$  y  $B(\theta)$  una función continua del vector paramétrico  $\theta$  independiente de  $x$  y acotada en  $\bigcup_{i=1}^k Q_d(\theta_i^*, r)$ .

Sea  $M'$  una cota superior de  $|B(\theta)|$  en dicho conjunto unión. Acotando los denominadores de estas expresiones mediante la desigualdad

$$p = \sum_{i=1}^k \pi_i f(x, \theta_i) \geq \alpha_0 f(x, \theta_h),$$

donde  $f(x, \theta_h)$  es, en cada caso, la que aparece en el numerador de la expresión anterior y considerando

$$\alpha_0 = \min \left\{ \pi_i, 1 - \sum_{j=1}^k \pi_j / \pi \in \hat{C}_k, i = 1, \dots, k \right\}$$

la suma de dichas expresiones será menor que una suma de funciones de la forma

$$\left| \frac{1}{\alpha_0} T(x) M' \right|$$

con  $T \in \Pi_h(t)$ .

Por tanto la expresión (3.1) está acotada por suma de funciones de la forma

$$\left| \frac{1}{\alpha_0} T(x) M' \right| \quad \text{ó} \quad \left| \frac{1}{\alpha_0^2} T(x) M'^2 \right| \quad \text{ó} \quad \left| \frac{1}{\alpha_0^3} T(x) M'^3 \right|$$

que son independientes del parámetro  $\psi$ .

Se considera  $h_{iju}$  dicha suma. Por la Proposición 3.2.2,  $T(x)p(x, \psi^*)$  es integrable, y por tanto  $h_{iju}(x)p(x, \psi^*)$  también.  $\square$

A continuación, se proponen unas condiciones suficientes que aseguran que cualquier mezcla de funciones de la misma familia exponencial regular cumple las condiciones de consistencia **C3** y **C4**.

**Teorema 3.3.2.** *Sea  $\mathcal{E}(a, t, b)$  una familia exponencial regular,  $k \in \mathbb{N}$  y  $\psi^* = (\pi^*, \theta_1^*, \dots, \theta_k^*) \in \Omega = \mathcal{C}_k \times \Theta^k$ , si se verifica:*

(H1)  $b$  continua

$$(H3) \int |\log a(x)| f(x, \theta_j^*) dx < +\infty \quad \forall j = 1, \dots, k,$$

entonces cualquier mezcla finita de  $k$  elementos de  $\mathcal{E}(a, t, b)$  cumple las condiciones **C3** y **C4** en un compacto  $\hat{\Omega} \subset \Omega$  tal que  $\psi^* \in \hat{\Omega}$ .

*Demostración.* Según la Proposición 1.4.1, para ver **C3** basta ver que se verifica **C3'**, es decir que  $\forall i, \forall j$  y  $\forall \theta_i$  en un cierto subconjunto de  $\Theta$  y  $r > 0$  suficientemente pequeño se verifica:

$$\int \log f_r^*(x, \theta_i) f(x, \theta_j^*) dx < +\infty, \quad (3.2)$$

siendo  $f_r^*(x, \theta_i) = \max\{1, \sup_{\theta' \in N_r(\theta_i)} f(x, \theta')\}$ , donde  $N_r(\theta_i)$  es la bola de centro  $\theta_i$  y radio  $r$ .

Se considera  $r > 0$  tal que  $Q_d(\theta_i^*, 2r) \subset \Theta$  y se probará que se verifica (3.2)  $\forall \theta_i \in Q_d(\theta_i^*, r)$ .

Ahora bien, como se deduce de la demostración del Lema 3.3.1,

$$\forall x, \quad \exists \varepsilon \in V_d / e^{\theta^T t(x)} \leq e^{\theta(\varepsilon)^T t(x)} \quad \forall \theta \in Q_d(\theta_i, 2r),$$

siendo  $\theta(\varepsilon) = \theta_i^* + 2r\varepsilon$ .

Por otra parte, por ser  $b$  continua, existe  $M$  tal que  $|b(\theta)| \leq M \quad \forall \theta \in Q_d(\theta_i^*, 2r)$ . Así, teniendo en cuenta que la bola  $N_r(\theta_i) \subset Q_d(\theta_i^*, 2r) \quad \forall \theta_i \in Q_d(\theta_i^*, r)$ , se tiene que:

$$f_r^*(x, \theta_i) \leq \max\{1, Ma(x)e^{\theta(\varepsilon)^T t(x)}\} \quad \forall \theta_i \in Q_d(\theta_i^*, r),$$

y por tanto, considerando  $M > 1$ ,

$$\log f_r(x, \theta_i) \leq \max\{0, \log(Ma(x)e^{\theta(\varepsilon)^T t(x)})\}$$

$$\begin{aligned} &\leq \log M + \max\{0, \log a(\mathbf{x})\} + \max\{0, \boldsymbol{\theta}(\boldsymbol{\varepsilon})^T \mathbf{t}(\mathbf{x})\} \\ &\leq \log M + |\log a(\mathbf{x})| + C \sum_{s=1}^d |t_s(\mathbf{x})| \quad \forall \boldsymbol{\theta}_i \in Q_d(\boldsymbol{\theta}_i^*, r) \end{aligned}$$

donde  $C$  es una constante que acota en valor absoluto a las componentes de los puntos del hipercubo  $Q_d(\boldsymbol{\theta}_i^*, 2r)$ .

Como  $t_s(\mathbf{x})f(\mathbf{x}, \boldsymbol{\theta}_j^*)$ ,  $\forall j = 1, \dots, k$ , es integrable por la Proposición 3.2.2 y  $(\log a(\mathbf{x}))f(\mathbf{x}, \boldsymbol{\theta}_j^*)$ ,  $\forall j = 1, \dots, k$ , también lo es por hipótesis, se tiene que  $\log f_r^*(\mathbf{x}, \boldsymbol{\theta}_i)f(\mathbf{x}, \boldsymbol{\theta}_j^*)$  es integrable como se quería demostrar.

Para ver  $\mathbf{C4}$ , por la Proposición 1.4.2, basta probar  $\mathbf{C4}'$ , es decir que

$$\int |\log f(\mathbf{x}, \boldsymbol{\theta}_i^*)| f(\mathbf{x}, \boldsymbol{\theta}_j^*) dx < +\infty$$

Ahora bien

$$|\log f(\mathbf{x}, \boldsymbol{\theta}_i^*)| f(\mathbf{x}, \boldsymbol{\theta}_j^*) \leq \left( |\log a(\mathbf{x})| + |\log b(\boldsymbol{\theta}_i^*)| + \sum_{s=1}^d |\boldsymbol{\theta}_s^{*i} t_s(\mathbf{x})| \right) f(\mathbf{x}, \boldsymbol{\theta}_j^*),$$

y todas las funciones del miembro derecho de la desigualdad son integrables, por lo que su suma también.

□

### 3.4 Condiciones de consistencia para mixturas de uniones de familias exponenciales

Hasta ahora se dispone de resultados que aseguran la verificación de  $\mathbf{C1}$ ,  $\mathbf{C3}$  y  $\mathbf{C4}$  en las mixturas formadas por funciones de la misma familia exponencial regular. En esta sección, se proponen resultados para asegurar dicha verificación en el caso en que intervengan en la mixtura funciones pertenecientes a distintas familias exponenciales regulares. Para ello, la familia de funciones producto de componentes de una función vectorial se generaliza a funciones producto de componentes de más de una función vectorial, tal y como se recoge en la siguiente definición.

**Definición 3.4.1.** Dadas las funciones  $t_i : \mathbb{R}^n \rightarrow \mathbb{R}^{d_i}$  para  $i = 1, \dots, k$  con  $\mathbf{t}_i(\mathbf{x}) = (t_{i1}(\mathbf{x}) \dots t_{id_i}(\mathbf{x}))^T$  y  $h$  un entero no negativo, se notará por  $\Pi_h(\mathbf{t}_1, \dots, \mathbf{t}_k)$



a la familia de funciones dada por

$$\{T : D \subset \mathbb{R}^n \rightarrow \mathbb{R} : T(x) = \prod_{i=1}^k \prod_{j=1}^{d_i} t_{ij}^{\nu_{ij}}(x), \nu_{ij} \in \mathbb{Z}^+ \cup \{0\}, \sum_{i=1}^k \sum_{j=1}^{d_i} \nu_{ij} \leq h\}$$

y se le denominará familia de funciones producto de a lo más  $h$  componentes de las funciones  $t_1(x), \dots, t_k(x)$ .

**Teorema 3.4.1.** Sean  $k \in \mathbb{N}$ ,  $\mathcal{E} = \mathcal{E}(a_i, t_i, b_i)$   $i \in \{1, \dots, k\}$  familias exponenciales regulares definidas sobre  $\Theta_i \subset \mathbb{R}^{d_i}$ . Sea  $\psi^* = (\pi^*, \theta_1^*, \dots, \theta_k^*) \in \Omega = \mathcal{C}_k \times \Theta_1 \times \dots \times \Theta_k$ . Si se verifican:

(H1) <sub>$i$</sub>  las funciones  $b_i$  son continuas  $\forall i \in \{1, \dots, k\}$ ,

(H2) <sub>$i$</sub>  las funciones  $b_i$  tienen derivadas parciales de hasta tercer orden continuas en  $\Theta_i$   $\forall i \in \{1, \dots, k\}$ ,

$$(H4) \int |T(x)| f_j(x, \theta_j^*) dx < +\infty, T \in \Pi_3(t_i, t_i, t_s), \forall j, i, l, s \in \{1, \dots, k\},$$

entonces cualquier mixtura finita de la forma  $p(x, \psi) = \sum_{i=1}^k \pi_i f_i(x, \theta_i)$  con  $f_i(x, \theta_i) \in$

$\mathcal{E}(a_i, t_i, b_i)$  verifica la condición de consistencia C1 en un cierto compacto  $\hat{\Omega} \subset \Omega$  que contiene a  $\psi^*$ .

*Nota 3.4.1.* En relación con el Teorema 3.3.1, este resultado para la unión de familias exponenciales, además de las hipótesis (H1) y (H2) para cada familia, introduce una nueva hipótesis (H4) de enlace entre las familias.

*Demostración.* Siguiendo la estructura de la demostración del Teorema 3.3.1, los puntos (1), (2) y (3) para que se verifiquen las desigualdades (a) y (b) están demostrados por ser las componentes de la mixtura,  $f_i(x, \theta_i)$ , funciones que pertenecen a alguna familia exponencial regular,  $\mathcal{E}_i(a_i, t_i, b_i)$ .

Para demostrar (c) es válido también el razonamiento anterior por el cual el valor absoluto de  $\frac{\partial^3 \log p}{\partial \psi_u \partial \psi_v \partial \psi_w}$  está acotado por una suma de expresiones de la forma

$$\left| \frac{1}{\alpha_0} T(x) M \right| \quad \text{con } T \in \Pi_3(t_i)$$

ó

$$\left| \frac{1}{\alpha_0^2} T(x) M \right| \quad \text{con } T \in \Pi_3(t_i, t_l)$$

ó

$$\left| \frac{1}{\alpha_0^3} T(x)M \right| \text{ con } T \in \Pi_3(t_i, t_l, t_s)$$

siendo  $\alpha_0 = \min\{\pi_i, 1 - \sum_{j=1}^k \pi_j/\pi \in \hat{\mathcal{C}}_k, i = 1, \dots, k\}$ .

Se considera  $h_{uvw}(x)$  la suma anterior. Por la hipótesis del teorema, se sabe que

$$\int |T(x)f_j(x, \theta_j^*)| dx < +\infty$$

para  $1 \leq j \leq k$ , y  $T \in \Pi_3(t_i, t_l, t_s)$ , de donde

$$\int h_{uvw}(x)f_j(x, \theta_j^*)dx < +\infty$$

para  $1 \leq j \leq k$ , y por tanto multiplicando por  $\pi_j^*$  y sumando en  $k$ , se tiene que  $h_{uvw}(x)p(x, \psi^*)$  es integrable.

□

**Teorema 3.4.2.** Sean  $k \in \mathbb{N}$ ,  $\mathcal{E} = \mathcal{E}(a_i, t_i, b_i)$  con  $i \in \{1, \dots, k\}$  familias exponenciales regulares definidas sobre  $\Theta_i \subset \mathbb{R}^{d_i}$  y sea  $\psi^* = (\pi^*, \theta_1^*, \dots, \theta_k^*) \in \Omega = \mathcal{C}_k \times \Theta_1 \times \dots \times \Theta_k$ . Si para todo  $1 \leq i, j \leq k$  se verifican

(H1)<sub>i</sub>  $b_i$  continuas,

(H3)<sub>i</sub>  $\int |\log a_i(x)|f_j(x, \theta_j^*)dx < +\infty$  y

(H5)  $\int |t_{is}(x)|f_j(x, \theta_j^*)dx < +\infty$  con  $t_{is}(x)$  cualquier componente de  $t_i$ ,

entonces cualquier mixtura finita de la forma  $p(x, \psi) = \sum_{i=1}^k \pi_i f_i(x, \theta_i) \in$

$\mathcal{E}(a_i, t_i, b_i)$  verifica las condiciones de consistencia C3 y C4 en un cierto compacto  $\hat{\Omega} \subset \Omega$  que contiene a  $\psi^*$ .

*Nota 3.4.2.* En relación con el Teorema 3.3.2, este caso incluye una nueva hipótesis (H5) de enlace entre las familias.

*Demostración.* La idea de esta demostración es similar a la del Teorema 3.2.2. Según la Proposición 1.4.1 bastará ver que  $\forall i, j, \forall \theta_i$  en un cierto compacto  $\hat{\Theta}_i \subset \Theta_i$

y  $r > 0$  suficientemente pequeño se verifica:

$$\int (\log f_{ir}^*(x, \theta_i)) f_j(x, \theta_j^*) dx < +\infty,$$

siendo  $f_{ir}^*(x, \theta_i) = \max\{1, \sup_{\theta' \in N_r(\theta_i)} f_i(x, \theta')\}$  y  $f_i \in \mathcal{E}_i, f_j \in \mathcal{E}_j$ .

Considerando  $r > 0$  tal que  $Q_{d_i}(\theta_i^*, 2r) \subset \Theta_i$ , se probará que se verifica lo anterior  $\forall \theta_i \in Q_{d_i}(\theta_i^*, r)$ .

Ahora bien,  $\forall x$  existe un  $\varepsilon$ , vector de signos, tal que

$$\forall \theta \in Q_{d_i}(\theta_i^*, 2r) \quad e^{\theta^T t_i(x)} \leq e^{\theta(\varepsilon)^T t_i(x)},$$

siendo  $\theta(\varepsilon) = \theta_i^* + 2r\varepsilon$ , y  $\exists M > 1$  tal que  $|b_i(\theta)| \leq M$  para  $\theta \in N_r(\theta_i) \subset Q_{d_i}(\theta_i^*, 2r)$ , así

$$f_{ir}^*(x, \theta_i) \leq \max\{1, M a_i(x) e^{\theta(\varepsilon)^T t_i(x)}\} \quad \forall \theta_i \in Q_{d_i}(\theta_i^*, r)$$

y por tanto

$$\begin{aligned} \log f_{ir}^*(x, \theta_i) &\leq \max\{0, \log(M a_i(x) e^{\theta(\varepsilon)^T t_i(x)})\} \\ &\leq \log M + \max\{0, \log a_i(x)\} + \max\{0, \theta(\varepsilon)^T t_i(x)\} \\ &\leq \log M + |\log a_i(x)| + C \sum_{s=1}^k |t_{is}(x)| \quad \forall \theta_i \in Q_{d_i}(\theta_i^*, r), \end{aligned}$$

donde  $C$  es una constante que acota en valor absoluto a las componentes de los puntos del hipercubo  $Q_{d_i}(\theta_i^*, 2r)$ .

Las funciones  $t_{is}(x) f_j(x, \theta_j^*)$  y  $\log a_i(x) f_j(x, \theta_j^*)$  son integrables por hipótesis, por tanto se tiene que  $\log f_{ir}^*(x, \theta_i) f_j(x, \theta_j^*)$  es integrable, como se quería probar.

Para ver C4, basta probar, por la Proposición 1.4.2, que

$$\int |\log f_i(x, \theta_i^*)| f_j(x, \theta_j^*) dx < +\infty.$$

Ahora bien

$$\begin{aligned} &|\log f_i(x, \theta_i^*)| f_j(x, \theta_j^*) \\ &\leq \left( |\log a_i(x)| + |\log b_i(\theta_i^*)| + \sum_{s=1}^k |\theta_{is}^* t_{is}(x)| \right) f_j(x, \theta_j^*) \end{aligned}$$

y todas estas funciones del miembro derecho de la desigualdad, son integrables, por lo que su suma también.

□

### 3.5 Condiciones de consistencia para mixturas de familias tipo $\mathcal{W}$

Los resultados presentados en las secciones anteriores recogen condiciones suficientes para que las mixturas de familias exponenciales regulares verifiquen las condiciones de consistencia. Las familias de distribuciones Lognormal y Gamma pertenecen a este tipo, cosa que no ocurre con las distribuciones Weibull. En esta sección se generalizan los resultados vistos anteriormente a las familias tipo  $\mathcal{W}$  definidas en la introducción de este capítulo.

**Notación** Se consideran los hipercubos  $Q_{d_1}(\alpha^*, r) = \prod_{i=1}^{d_1} [\alpha_i^* - r, \alpha_i^* + r]$  y  $Q_{d_1}(\lambda^*, r) = \prod_{i=1}^{d_1} [\lambda_i^* - r, \lambda_i^* + r]$ . Para cada  $\varepsilon \in \mathbb{R}^{d_1}$  vector de signos con cada componente  $\varepsilon_i = \pm 1$ , se denota por  $\alpha(\varepsilon) \in \mathbb{R}^{d_1}$  y  $\lambda(\varepsilon) \in \mathbb{R}^{d_1}$ , con coordenadas  $\alpha_i(\varepsilon) = \alpha_i^* + \varepsilon_i r$  y  $\lambda_i(\varepsilon) = \lambda_i^* + \varepsilon_i r$ , respectivamente, los vértices de los hipercubos de semiamplitud  $r$ .

**Lema 3.5.1.** *Sea  $t : D \subset \mathbb{R}^n \rightarrow \mathbb{R}^{d_1}$ ,  $(\alpha^*, \lambda^*) \in \mathbb{R}^{d_1} \times \mathbb{R}^{d_1}$  y  $r > 0$ . Entonces, para todo  $x \in D$  y  $(\alpha, \lambda) \in Q_{d_1}(\alpha^*, r) \times Q_{d_1}(\lambda^*, r)$*

$$\begin{aligned} & \exp\{t(x)\alpha + Ke^{\lambda^T t(x)}\} \leq \\ & \leq \sum_{\varepsilon \in V_{d_1}} (\exp\{\alpha(\varepsilon)^T t(x) + Ke^{\lambda(\varepsilon)^T t(x)}\} + \exp\{\alpha(\varepsilon)^T t(x) + Ke^{\lambda(-\varepsilon)^T t(x)}\}), \end{aligned}$$

con  $K \in \mathbb{R}$  constante y  $V_{d_1} = \{\varepsilon \in \mathbb{R}^{d_1} / \varepsilon_i = \pm 1 \quad \forall i = 1, \dots, d_1\}$ .

*Demostración.* Dado  $x \in D$ , se considera  $\varepsilon(x) = [\varepsilon_1(x) \dots \varepsilon_{d_1}(x)]^T$  tal que

$$\varepsilon_i(x) = \begin{cases} 1 & \text{si } t_i(x) \geq 0 \\ -1 & \text{si } t_i(x) < 0 \end{cases}$$

que, obviamente, pertenece a  $V_{d_1}$ .

Entonces  $\forall i = 1, \dots, d_1$

$$t_i(x)\alpha_i \leq t_i(x)[\alpha_i^* + \varepsilon_i(x)r] \quad \text{y} \quad t_i(x)\lambda_i \leq t_i(x)[\lambda_i^* + \varepsilon_i(x)r] \quad ,$$

por lo que

$$e^{\alpha^T t(x)} \leq e^{\alpha(\varepsilon(x))^T t(x)} \quad \text{y} \quad e^{\lambda^T t(x)} \leq e^{\lambda(\varepsilon(x))^T t(x)}.$$

Además, de la segunda desigualdad, se deduce que

$$Ke^{\lambda^T t(x)} \leq Ke^{\lambda(s(K)\varepsilon(x))^T t(x)}$$

donde  $s(K)$  es el signo de  $K$ . Así, para un  $\varepsilon(x)$  en concreto, se tiene que

$$\exp\{\alpha^T t(x) + Ke^{\lambda^T t(x)}\} \leq \exp\{\alpha(\varepsilon(x))^T t(x) + Ke^{\lambda(s(K)\varepsilon(x))^T t(x)}\}.$$

Por tanto, al verificarse esto para un vértice concreto del hipercubo, si se suman en todos los vértices, al ser todos los términos cantidades positivas, se obtiene que

$$\begin{aligned} & \exp\{\alpha^T t(x) + Ke^{\lambda^T t(x)}\} \\ & \leq \sum_{\varepsilon \in V_{d_1}} \left( \exp\{\alpha(\varepsilon)^T t(x) + Ke^{\lambda(+\varepsilon)^T t(x)}\} + \exp\{\alpha(\varepsilon)^T t(x) + Ke^{\lambda(-\varepsilon)^T t(x)}\} \right). \end{aligned}$$

□

**Teorema 3.5.1.** *Sea  $\mathcal{W}(a, t, b, c)$  una familia tipo  $\mathcal{W}$  y  $\psi^* = (\pi, \theta_1^*, \dots, \theta_k^*)$  en  $\mathcal{C}_k \times \Theta^k$  con  $\theta_i^* = (\alpha_i^*, \beta_i^*, \lambda_i^*) \forall i = 1, \dots, k$ . Si se verifican*

(H1)' *b y c continuas*

(H2)' *b y c tienen derivadas parciales de hasta tercer orden continuas en sus dominios y*

(H6) *c es la función idénticamente nula o tiene una derivada parcial de primer orden no nula en  $\beta_i^*$  para todo  $i = 1, \dots, k$ ,*

*entonces cualquier mezcla finita definida a través de k componentes de  $\mathcal{W}(a, t, b, c)$  verifica la condición de consistencia C1 en un cierto compacto  $\hat{\mathcal{C}}_k \times \hat{\Theta}_1 \times \dots \times \hat{\Theta}_k \subset \mathcal{C}_k \times \Theta^k$  que contiene a  $\psi^*$ .*

*Nota 3.5.1.* Respecto al Teorema 3.3.1 se han sustituido las hipótesis sobre  $b$ , (H1) y (H2), por sus análogas sobre  $\hat{b}$  y  $\hat{c}$ , (H1)' y (H2)', y se ha incluido una nueva hipótesis, (H6), sobre  $\hat{c}$ .

*Demostración.* Dado que  $\Theta$  es abierto, según la definición de la familia  $\mathcal{W}$

$$\exists r > 0 : Q_d(\theta_i^*, r) \subset \Theta \quad \forall i = 1, \dots, k$$

Por tanto, si se elige un compacto que contenga a  $\pi^*$ ,  $\hat{\mathcal{C}}_k \subset \mathcal{C}_k$ , se tendrá que  $\hat{\mathcal{C}}_k \times Q_d(\theta_1^*, r) \times \dots \times Q_d(\theta_k^*, r)$  es un subconjunto compacto de  $\mathcal{C}_k \times \Theta^k$  que contiene a  $\psi^*$ .

Dada una mixtura  $p(x; \psi) = \sum_{i=1}^k \pi_i f(x, \theta_i)$  de densidades pertenecientes a  $\mathcal{W}(a, t, b, c)$ , por la linealidad de la derivada, la desigualdad triangular y dada la simetría para cada índice, de nuevo hay que probar (1), (2), (3) del Teorema 3.2.1 para toda  $f(x, \theta) \in \mathcal{W}$ , y también (c) para la mixtura de dichas funciones.

Para la demostración se puede suponer la función  $b(\theta) \equiv 1$ , si no fuese así la demostración hecha para la función  $\frac{f(x, \theta)}{b(\theta)}$  implica que se verifica para  $f(x, \theta)$ , por ser  $b$  con derivadas parciales de hasta tercer orden continuas y porque la tesis del teorema se asegura para un cierto compacto.

Por ser  $c$  y sus derivadas de hasta tercer orden continuas para todo  $\theta \in \hat{\Theta}$ , los valores absolutos de  $c(\theta)$ , las sucesivas derivadas, y funciones continuas de las mismas que aparecen a lo largo de la prueba, se pueden acotar por una misma constante  $M > 0$ .

(1) Se demuestra usando el Lema 3.5.1 y considerando  $\beta_0 = \arg \max_{\beta} c(\beta)$ , así para todo  $\theta \in \hat{\Theta}$

$$|f(x, \theta)| \leq \sum_{\varepsilon, s \in \{-1, 1\}} a(x) \exp\{t(x)\alpha(\varepsilon) + c(\beta_0) e^{\lambda(\varepsilon s)^T t(x)}\} = g(x),$$

que es integrable e independiente de  $\theta$ .

(2) Derivando respecto a los distintos parámetros la función  $f(x, \alpha, \beta, \lambda)$  se obtienen las expresiones

$$\left| \frac{\partial f(x, \alpha, \beta, \lambda)}{\partial \alpha_i} \right| = |f(x, \alpha, \beta, \lambda) t_i(x)| \quad \forall i$$

$$\left| \frac{\partial f(x, \alpha, \beta, \lambda)}{\partial \beta_j} \right| = \left| f(x, \alpha, \beta, \lambda) \frac{\partial c(\beta)}{\partial \beta_j} e^{\lambda^T t(x)} \right| \quad \forall j$$

$$\left| \frac{\partial f(x, \alpha, \beta, \lambda)}{\partial \lambda_u} \right| = \left| f(x, \alpha, \beta, \lambda) c(\beta) t_u(x) e^{\lambda^T t(x)} \right| \quad \forall u$$

todas ellas son funciones integrables, por la Proposición 3.2.3, y acotadas respectivamente por:

$$g_{1i}(x) = g(x) t_i(x),$$

$$g_{2j}(x) = M \sum_{\varepsilon, s} e^{\lambda^{T(\varepsilon)}t(x)} f(x, (\alpha(\varepsilon), \beta_0, \lambda(s\varepsilon))),$$

y

$$g_{3u}(x) = |t_u(x)| M \sum_{\varepsilon, s} e^{\lambda(\varepsilon)^T t(x)} f(x, (\alpha(\varepsilon), \beta_0, \lambda(s\varepsilon))),$$

que son integrables por el segundo punto del Proposición 3.2.3, y además son independientes de  $\theta$ , lo que demuestra (2).

(3) Se comprueba igualmente viendo que el resto de las derivadas son sumas de productos de las forma

$$f(x, \theta) T(x) e^{m\lambda^T t(x)} C(\beta), \quad m = 1, 2, 3$$

con  $T \in \Pi_3(t)$  una función producto de a lo sumo tres de las componentes de la función vectorial  $t(x)$  y  $C(\beta)$  una función continua acotada por  $M$ . Se sabe que estas sumas son integrables por la Proposición 3.2.3 y se pueden acotar por funciones integrables independientes de  $\theta$ , por el Lema 3.5.1.

Quedaría por probar (c) del Teorema 3.3.1 para lo cual, siguiendo el razonamiento de la demostración de los teoremas 3.3.1 y 3.4.1, se sabe que el valor absoluto de la expresión  $\frac{\partial^3 \log p}{\partial \psi_u \partial \psi_v \partial \psi_w}$  está acotado por una suma de funciones de la forma

$$\left| \frac{1}{\alpha_0^3} T(x) e^{m\lambda^T t(x)} M \right|,$$

con  $\alpha_0$  definida como en el Teorema 3.4.1 y donde cada sumando está acotado a su vez por una suma de expresiones del tipo

$$\sum_{\varepsilon \in V_d} \left| \frac{1}{\alpha_0^3} T(x) e^{m\lambda(\varepsilon)^T t(x)} M \right|$$

que son funciones independientes de los parámetros. Se considera la función que se buscaba,  $h_{uvw}$ , como la suma de estas funciones. Se verifica que el producto  $h_{uvw}(x)p(x, \psi^*)$  es integrable por ser una suma de funciones que, por la Proposición 3.2.3, lo son. Y por tanto (c) queda demostrado. □

**Teorema 3.5.2.** *Sea  $\mathcal{W}(a, t, b, c)$  una familia tipo  $\mathcal{W}$  y  $\psi^* = (\pi, \theta_1^*, \dots, \theta_k^*)$  en  $\mathcal{C}_k \times \Theta^k$  con  $\theta_i^* = (\alpha_i^*, \beta_i^*, \lambda_i^*)$  para todo  $i = 1, \dots, k$ . Si se verifican:*

(H1)' *b y c continuas,*

$$(H3) \int |\log a(x)| f(x, \theta_i^*) dx < +\infty \text{ para todo } i = 1, \dots, k \text{ y}$$

(H6)  $c$  es la función idénticamente nula o tiene al menos una derivada parcial de primer orden no nula en  $\beta_i^*$  para todo  $i = 1, \dots, k$ ,

entonces cualquier mezcla finita definida a través de  $k$  componentes de  $\mathcal{W}(a, t, b, c)$  verifica la condición de consistencia  $C3$  y  $C4$  en un cierto compacto  $\hat{\Omega}$  que contiene a  $\psi^*$ .

*Nota 3.5.2.* Al igual que en el teorema anterior si se compara este teorema con el análogo referente a una familia exponencial, la hipótesis (H1), sobre  $b$ , se ha sustituido por (H1)', sobre  $b$  y  $c$ , y se ha añadido una nueva hipótesis sobre  $c$ , (H6).

*Demostración.* Según la Proposición 1.4.1 basta ver que para todo  $i, j$  y para todo  $\theta_i$  en un cierto subconjunto de  $\Theta$  y  $r > 0$  suficientemente pequeño se verifica:

$$\int \log f_r^*(x, \theta_i) f(x, \theta_j^*) dx < +\infty,$$

siendo  $f_r^*(x, \theta_i) = \max\{1, \sup_{\theta' \in N_r(\theta_i)} f(x, \theta')\}$ .

Tomando  $r > 0$  tal que  $Q_d(\theta_i^*, 2r) \subset \Theta$ , se probará que se verifica lo anterior  $\forall \theta_i \in Q_d(\theta_i^*, r)$ .

De la demostración del Lema 3.5.1 y teniendo en cuenta que  $c(\beta)$  es continua, se deduce que

$$\forall x \quad \exists \varepsilon(x) \in V_d / \quad \forall \theta_i \in Q_d(\theta_i^*, 2r)$$

$$\exp\{\alpha^T t(x) + c(\beta) e^{\lambda^T t(x)}\} \leq \exp\{\alpha(\varepsilon(x))^T t(x) + M e^{\lambda(\varepsilon(x))^T t(x)}\},$$

donde  $\alpha(\varepsilon)$  y  $\lambda(\varepsilon)$  son, respectivamente, los vértices de los hiper cubos centrados en  $\alpha_i^*$  y  $\lambda_i^*$  y de semi amplitud  $2r$  y  $M$  es tal que  $|c(\beta)| \leq M$  (el tomar esta cota que es un valor positivo hace que en la desigualdad los dos vértices  $\varepsilon(x)$  que se eligen puedan ser el mismo). Así, si además se toma  $M > 1$  tal que  $|b(\theta')| \leq M$ , se tiene

$$f(x, \theta) \leq M a(x) \exp\{\alpha(\varepsilon(x))^T t(x) + M e^{\lambda(\varepsilon(x))^T t(x)}\} \quad \theta \in Q_d(\theta_i^*, 2r)$$

y como  $N_r(\theta_i) \subset Q_d(\theta_i^*, 2r)$ , para todo  $\theta_i \in Q_d(\theta_i^*, r)$

$$f_r(x, \theta_i) \leq \max\{1, M a(x) \exp\{\alpha(\varepsilon(x))^T t(x) + M e^{\lambda(\varepsilon(x))^T t(x)}\}\}$$



y, tomando logaritmo, se obtiene

$$\begin{aligned} \log f_r^*(\mathbf{x}, \theta_i) &\leq \max\{0, \log(Ma(\mathbf{x}) \exp\{\alpha(\boldsymbol{\varepsilon}(\mathbf{x}))^T \mathbf{t}(\mathbf{x}) + Me^{\lambda(\boldsymbol{\varepsilon}(\mathbf{x}))^T \mathbf{t}(\mathbf{x})}\})\} \\ &\leq \log M + \max\{0, \log a(\mathbf{x})\} + \max\{0, \alpha(\boldsymbol{\varepsilon}(\mathbf{x}))^T \mathbf{t}(\mathbf{x}) + Me^{\lambda(\boldsymbol{\varepsilon}(\mathbf{x}))^T \mathbf{t}(\mathbf{x})}\} \\ &\leq \log M + |\log a(\mathbf{x})| + C \sum_{i=1}^k |t_i(\mathbf{x})| + M \sum_{\boldsymbol{\varepsilon}} e^{\lambda(\boldsymbol{\varepsilon})^T \mathbf{t}(\mathbf{x})}, \end{aligned}$$

donde  $C$  es una constante que acota en valor absoluto a las componentes de  $\alpha(\boldsymbol{\varepsilon}(\mathbf{x}))$  en  $Q_{d_1}(\boldsymbol{\alpha}^*, 2r)$ .

Por la Proposición 3.2.3, tanto  $t_i(\mathbf{x})f(\mathbf{x}, \theta_j^*)$  como  $e^{\lambda(\boldsymbol{\varepsilon})^T \mathbf{t}(\mathbf{x})}f(\mathbf{x}, \theta_j^*)$  son integrables y, por hipótesis,  $\log a(\mathbf{x})f(\mathbf{x}, \theta_j^*)$  también, demostrándose la integrabilidad de  $\log f_r(\mathbf{x}, \theta_i)f(\mathbf{x}, \theta_j^*)$  como se pretendía.

Para comprobar C4, por la Proposición 1.4.2, basta probar que

$$\int |\log f(\mathbf{x}, \theta_i^*)|f(\mathbf{x}, \theta_j^*)d\mathbf{x} < +\infty$$

ahora bien

$$\begin{aligned} &|\log f(\mathbf{x}, \theta_i^*)|f(\mathbf{x}, \theta_j^*) \leq \\ &\leq \left( |\log a(\mathbf{x})| + |\log b(\theta_i^*)| + \sum_{s=1}^d |\alpha_{i_s}^* t_i(\mathbf{x})| + Me^{(\lambda_i^*)^T \mathbf{t}(\mathbf{x})} \right) f(\mathbf{x}, \theta_j^*) \end{aligned}$$

siendo esta última expresión una suma en la que todos los sumandos son integrables.

□

### 3.6 Condiciones de consistencia para mixturas de uniones de familias tipo $\mathcal{W}$

Las familias tipo  $\mathcal{W}$  surgen de la necesidad de un tipo de familia en el que estuviesen incluidas las tres familias de distribuciones que entran en juego en las mixturas mixtas  $\mathcal{P}$ . Los resultados anteriores están referidos a las mixturas de funciones de una misma familia tipo  $\mathcal{W}$ , en esta sección se generalizan a las mixturas de funciones de distintas familias tipo  $\mathcal{W}$ .

**Teorema 3.6.1.** Sea  $\psi^* = (\pi^*, \theta_1^*, \dots, \theta_k^*) \in \Omega = \mathcal{C} \times \Theta_1 \times \dots \times \Theta_k$  y sean  $\mathcal{W}(a_i, t_i, b_i, c_i)$   $k$  familias tipo  $\mathcal{W}$  definidas respectivamente sobre  $\Theta_i \subset \mathbb{R}^{d_i}$  con  $i \in \{1, \dots, k\}$ . Si se verifican:

$$(H1)'_i \quad b_i \text{ y } c_i \text{ continuas } i = 1, \dots, k,$$

$(H2)'_i$   $b_i$  y  $c_i$  tienen derivadas parciales de hasta tercer orden continuas en sus dominios,

$$(H4)' \quad \forall i, l, s, j = 1, \dots, k$$

$$\forall m_i, m_l, m_s \in \mathbb{Z} \text{ tales que } \begin{cases} m_i + m_l + m_s \leq 3 \\ c_r = 0 \Rightarrow m_r = 0 \end{cases} \quad (r = i, l, s)$$

$$\text{y } \forall T \in \Pi_3(t_i, t_l, t_s):$$

$$\int |T(x)| e^{m_i t_i(x) \lambda_i} e^{m_l t_l(x) \lambda_l} e^{m_s t_s(x) \lambda_s} f_j(x, \theta_j) dx < +\infty \quad \text{y}$$

$(H6)_i$   $c_i$  es la función idénticamente nula, o tiene alguna de sus derivadas parciales de primer orden no nula en  $\beta_i^*$  para todo  $i = \{1, \dots, k\}$ ,

entonces, cualquier mixtura finita de la forma  $p(x, \psi) = \sum_{i=1}^k \pi_i f_i(x, \theta_i)$  con  $f_i(x, \theta_i) \in$

$\mathcal{W}(a_i, t_i, b_i, c_i)$  verifica la condición de consistencia **C1** en un cierto compacto  $\hat{\mathcal{C}}_k \times \hat{\Theta}_1 \times \dots \times \hat{\Theta}_k \subset \mathcal{C}_k \times \Theta_1 \times \dots \times \Theta_k$  que contiene a  $\psi^*$ .

*Nota 3.6.1.* Con respecto al Teorema 3.5.1, para una sólo familia tipo  $\mathcal{W}$ , se ha incluido una hipótesis nueva  $(H4)'$  de enlace análoga a la hipótesis de enlace  $(H4)$  que se incluyó en el Teorema 3.4.1, pero más general. Este teorema generaliza a todos los teoremas anteriores acerca de verificación de la condición de consistencia **C1**.

*Demostración.* (1), (2) y (3) de la demostración del Teorema 3.2.1 se prueban de forma similar al Teorema 3.5.1 y basta demostrar que se verifica (c) del Teorema 3.2.1. Para ello, siguiendo el razonamiento anterior, se tiene que el valor absoluto

de  $\frac{\partial^3 \log p}{\partial \psi_u \partial \psi_v \partial \psi_w}$  está acotado por una suma de expresiones de la forma

$$\sum_{\varepsilon} \left| \frac{M}{\alpha_0} T(\mathbf{x}) e^{m t_i(\mathbf{x}) \lambda_i(\varepsilon)} \right| \quad \text{con } T \in \Pi_3(t_i)$$

ó

$$\sum_{\varepsilon, \varepsilon'} \left| \frac{M}{\alpha_0^2} T(\mathbf{x}) e^{m_i t_i(\mathbf{x}) \lambda_i(\varepsilon)} e^{m_i t_i(\mathbf{x}) \lambda_i(\varepsilon')} \right| \quad \text{con } T \in \Pi_3(t_i, t_i)$$

ó

$$\sum_{\varepsilon, \varepsilon', \varepsilon''} \left| \frac{M}{\alpha_0^3} T(\mathbf{x}) e^{m_i t_i(\mathbf{x}) \lambda_i(\varepsilon)} e^{m_i t_i(\mathbf{x}) \lambda_i(\varepsilon')} e^{m_s t_s(\mathbf{x}) \lambda_s(\varepsilon'')} \right| \quad \text{con } T \in \Pi_3(t_i, t_i, t_s)$$

con  $m_r \in \{0, 1, 2, 3\}$  tales que  $m_i + m_l + m_s \leq 3$ .

Nótese que, si para algún  $r$  se verifica que  $c_r = 0$ , entonces  $m_r = 0$  ya que los factores  $e^{m_r t_r(\mathbf{x}) \lambda_r}$  aparecen de derivar la función  $f_r$  respecto de alguna componente de los vectores paramétricos  $\beta_r$  o  $\lambda_r$  de los que no depende la función.

Todas esas funciones son independientes de los parámetros y multiplicadas por  $p(\mathbf{x}, \psi^*)$  son integrables por la Proposición 3.2.3. Se considera la función  $h_{uvw}$  esta suma de funciones; se verifica que el producto  $h_{uvw}(\mathbf{x}) p(\mathbf{x}, \psi^*)$  es integrable por ser suma finita de funciones que lo son, luego (c) queda demostrado. □

**Teorema 3.6.2.** *Sea  $\psi^* = (\pi^*, \theta_1^*, \dots, \theta_k^*) \in \Omega = \mathcal{C} \times \Theta_1 \times \dots \times \Theta_k$  y sean  $\mathcal{W}(a_i, t_i, b_i, c_i)$   $k$  familias tipo  $\mathcal{W}$  definidas respectivamente sobre  $\Theta_i \subset \mathbb{R}^{d_i}$  con  $i \in \{1, \dots, k\}$ . Si se verifican:*

(H1)<sub>*i*</sub>  $b_i$  son continuas,

$$(H3)_i \int |\log a_i(\mathbf{x})| f_i(\mathbf{x}, \theta_i^*) d\mathbf{x} < +\infty,$$

(H5)'  $\int |t_{is}(\mathbf{x})| e^{m t_i(\mathbf{x}) \lambda_i} f_l(\mathbf{x}, \theta_l^*) d\mathbf{x} < +\infty$   $i, l \in \{1, \dots, k\}$ , con  $t_{is}(\mathbf{x})$  cualquier componente de  $t_i(\mathbf{x})$ , para  $m = 0, 1, 2, 3$  si  $c_i \neq 0$  y para  $m = 0$  si  $c_i = 0$ ,  
y

(H6)<sub>*i*</sub>  $c_i$  son la función idénticamente nula o una función con derivadas parciales de primer orden continuas y alguna no nula en  $\beta_i^*$ ,

entonces, cualquier *mixtura finita* de la forma  $p(x, \psi) = \sum_{i=1}^k \pi_i f_i(x, \theta_i)$  con  $f_i(x, \theta_i) \in \mathcal{W}(a_i, t_i, b_i, c_i)$  verifica las condiciones de consistencia  $\mathcal{C}_3$  y  $\mathcal{C}_4$  en un compacto  $\hat{\Omega} \subset \Omega$  que contiene a  $\psi^*$ .

*Nota 3.6.2.* En este caso, respecto al Teorema 3.5.2, para una sólo familia tipo  $\mathcal{W}$ , se ha incluido una hipótesis (H5)' de enlace más general que la hipótesis de enlace (H5) que se incluyó en el Teorema 3.4.2. De esta forma, este teorema es una generalización de todos los teoremas anteriores acerca de verificación de las condiciones de consistencia  $\mathcal{C}_3$  y  $\mathcal{C}_4$ .

*Demostración.* La demostración es semejante a la de los teoremas 3.4.2 y 3.5.2. □

### 3.7 Condiciones de consistencia para *mixturas finitas* de la unión de las familias Gamma, Lognormal y Weibull

El objetivo de esta sección es comprobar que el conjunto de *mixturas mixtas*  $\mathcal{P}$  verifica las condiciones de consistencia.

#### 3.7.1 Condiciones de consistencia $\mathcal{C}_1$ , $\mathcal{C}_3$ y $\mathcal{C}_4$

Para comprobar que las *mixturas* de  $\mathcal{P}$  cumplen estas condiciones bastará ver que las familias de distribuciones Lognormal, Gamma y Weibull verifican las hipótesis de los teoremas dados en la sección anterior. Para ello, a continuación, se identifican las funciones  $a_i(x)$ ,  $t_i(x)$ ,  $b_i(\theta_i)$ ,  $c_i(\beta_i)$ , en las que el subíndice  $i = 1, 2, 3$  se refiere a las familias de distribuciones Lognormal, ( $\mathcal{F}_L$ ), Gamma, ( $\mathcal{F}_G$ ), y Weibull, ( $\mathcal{F}_W$ ), respectivamente.

La familia Lognormal,

$$f(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma x}} \exp\left\{-\frac{(\log x - \mu)^2}{2\sigma^2}\right\}, \quad x > 0$$

para  $\mu \in \mathbb{R}$ ,  $\sigma > 0$ , puede considerarse una familia del tipo  $\mathcal{W}$ , dependiente únicamente del vector paramétrico  $\alpha$ , con

$$a_1(x) = \frac{1}{x}, \quad t_1(x) = (\log x, -\log^2 x), \quad b_1(\alpha_{11}, \alpha_{12}) = \frac{\sqrt{\alpha_{12}} e^{-\frac{\alpha_{11}^2}{4\alpha_{12}}}}{\sqrt{\pi}} \text{ y } c_1(\beta) \equiv 0$$

para  $\alpha_1 = (\alpha_{11}, \alpha_{12}) = (\frac{\mu}{\sigma^2}, \frac{1}{2\sigma^2}) \in \mathbb{R} \times (0, +\infty)$ .

La familia Gamma,

$$f(x; a, b) = \frac{b^{-a}}{\Gamma(a)} x^{a-1} e^{-x/b}, \quad x > 0$$

para  $a, b > 0$  puede considerarse una familia del tipo  $\mathcal{W}$  con

$$a_2(x) = \frac{1}{x}, \quad t_2(x) = (\log x, x), \quad b_2(\alpha_{21}, \alpha_{22}) = \frac{(-\alpha_{22})^{\alpha_{21}}}{\Gamma(\alpha_{21})} \text{ y } c_2(\beta) \equiv 0$$

para  $\alpha_2 = (\alpha_{21}, \alpha_{22}) = (a, \frac{-1}{b}) \in (0, +\infty) \times (-\infty, 0)$ .

La familia Weibull, es una restricción de la familia  $\mathcal{W}(a_3, t_3, b_3, c_3)$ . En general, dada una familia paramétrica

$$\mathcal{T} = \{f(x, \theta) : x \in S, \theta \in \Theta\}$$

y dada una cierta función  $\Phi : \Theta' \mapsto \Theta$ , puede considerarse una nueva familia de la siguiente forma

$$\mathcal{T}' = \{f(x, \Phi(\theta)) : x \in S, \theta' \in \Theta'\}$$

Una simple aplicación de la regla de la cadena muestra que, teniendo en cuenta el carácter local de las condiciones de consistencia, bajo condiciones de regularidad suficientes para  $\Phi$ , si las mixturas de la familia  $\mathcal{T}$  verifican las condiciones de consistencia también las verifican las mixturas de la familia  $\mathcal{T}'$ .

En este caso, teniendo en cuenta que la densidad de una distribución Weibull es

$$f(x; c, d) = \frac{c}{d^c} x^{c-1} e^{-x^c/d^c}, \quad x > 0$$

para  $c, d > 0$ ,  $\Phi$  viene dada por

$$\begin{aligned} \Phi : (0, \infty) \times (0, \infty) &\mapsto (0, \infty) \times (0, \infty) \times (0, \infty) \\ (c, d) &\mapsto (c, d^c, c) \end{aligned}$$

que es continua e indefinidamente derivable. Así pues, bastará ver que las condiciones se verifican para  $\mathcal{W}(a_3, t_3, b_3, c_3)$  donde

$$a_3(x) = \frac{1}{x}, \quad t_3(x) = \log x, \quad b_3(\alpha_3, \beta, \lambda) = \frac{\lambda}{\Gamma(\frac{\alpha_3}{\lambda})\beta^{\frac{\alpha_3}{\lambda}}} \text{ y } c_3(\beta) = -\frac{1}{\beta},$$

tomando los parámetros  $(\alpha_3, \beta, \lambda) = (c, d^c, c) \in (0, +\infty) \times (0, +\infty) \times (0, +\infty)$ .

**Teorema 3.7.1.** *La familia de mixturas  $\mathcal{P}$  cumple las condiciones de consistencia C1, C3 y C4.*

*Demostración.* Previamente es importante notar que:

(1)  $\log a_i(x) = \log \frac{1}{x}$  para  $i = 1, 2, 3$ , que en valor absoluto coincide con  $\log x$ .

(2) Si  $T(x) \in \Pi_3(t_j, t_l, t_s)$  con  $j, l, s \in \{1, 2, 3\}$  entonces es de la forma  $(x)^p \log^h x$  con  $p, h \in \mathbb{Z}^+ \cup \{0\}$ .

(3) Para el caso de distribuciones Lognormales y Gammas las funciones  $c_i$  son la función idénticamente nula y por tanto  $m_i = 0$  para  $i = 1, 2$ , lo que implica que la única  $t_i(x)$  que aparece en la exponencial sea la correspondiente a la distribución Weibull, es decir  $t_3(x) = \log x$ .

Así, teniendo en cuenta quienes son las funciones  $t(x)$ , para ver que se verifican las hipótesis de integrabilidad sobre  $x$  de los teoremas 3.6.1 y 3.6.2, bastará comprobar que:

$$x^p (\log x)^h e^{k \log x} f_l(x, \theta_l)$$

son integrables para  $p, h \in \mathbb{Z}^+ \cup \{0\}$ ,  $k \in [0, +\infty)$  y  $f_l(x, \theta_l)$  una función de densidad de alguna de las tres familias.

**Caso 1** Si  $k = 0$  y  $p = 0$  esas funciones son integrables ya que  $\log^h x$  se puede obtener a partir del producto de componentes de  $t(x)$  en cualquiera de los tres casos y aplicar la Proposición 3.2.3.

**Caso 2**  $p$  ó  $k$  no nulos.

Si  $p \neq 0$  el factor  $x^p$  se puede escribir de la forma  $e^{p \log x}$  y por tanto únicamente quedaría ver que

$$(\log x)^h e^{k \log x} f_l(x, \theta_l)$$

son integrables para  $h \in \mathbb{Z}^+ \cup \{0\}$ ,  $k \in [0, +\infty)$  y  $f_l(x, \theta_l)$  una función de densidad de alguna de las tres familias.

Para  $l=1$ , es decir,  $f_l$  es la función de densidad de una distribución Lognormal,

$$\begin{aligned} (\log x)^h e^{k \log x} a_1(x) b_1(\alpha_1) \exp\left\{(\alpha_{11} \quad \alpha_{12}) \begin{pmatrix} \log x \\ -\log^2 x \end{pmatrix}\right\} = \\ (\log x)^h a_1(x) b_1(\alpha_1) \exp\left\{(\alpha_{11} + k \quad \alpha_{12}) \begin{pmatrix} \log x \\ -\log^2 x \end{pmatrix}\right\} = \\ \frac{b_1(\alpha_{11}, \alpha_{12})}{b_1(\alpha_{11} + k, \alpha_{12})} \log^h x f_1(x, (\alpha_{11} + k, \alpha_{12})) \end{aligned}$$

que son integrables dado que se reducen al Caso 1, ya que  $\alpha_{11} + k$  pertenece al rango del primer parámetro de la distribución Lognormal, y el primer factor no depende de  $x$ .

Para  $l=2$ , es decir,  $f_l$  es la función de densidad de una distribución Gamma, de la misma forma anterior se ve que

$$(\log x)^h e^{k \log x} f_l(x, \theta_l) = \frac{b_2(\alpha_{21}, \alpha_{22})}{b_2(\alpha_{21} + k, \alpha_{22})} (\log x)^h f_2(x, (\alpha_{21} + k, \alpha_{22}))$$

son integrables pues  $\alpha_{21} + k$  pertenece al rango de  $\alpha_{21}$  que es  $(0, +\infty)$  y por tanto, nuevamente, se reduce al primer caso.

Y por último, para el caso  $l=3$ , es decir, que  $f_l$  sea la función de densidad de una distribución Weibull,

$$(\log x)^h e^{k \log x} f_l(x, \theta_l) = \frac{b_3(\alpha_3, \beta_3, \alpha_3)}{b_3(\alpha_3 + k, \beta_3, \alpha_3)} \log^h x f_3(x, (\alpha_3 + k, \beta_3, \alpha_3))$$

donde  $f_3(x, (\alpha_3 + k, \beta_3, \alpha_3))$  pertenece a la familia de funciones de densidad de la distribución Weibull al ser  $\alpha_3 + k$  perteneciente a  $(0, +\infty)$ .  $\square$

### 3.7.2 Condición de consistencia C2

En esta subsección se procede a comprobar que el conjunto de mixturas mixtas  $\mathcal{P}$  verifica la segunda condición de consistencia que seguidamente se recuerda.

(C2) La matriz información de Fisher,  $I(\psi)$ , dada por:

$$I(\psi) = \int [\nabla_{\psi} \log q(x; \psi)] [\nabla_{\psi} \log q(x; \psi)]^T q(x; \psi) dx,$$

existe y es definida positiva en  $\psi^*$ , donde  $\nabla_\psi$  denota el gradiente de primeras derivadas parciales de la función con respecto a las componentes de  $\psi$ .

*Nota 3.7.1.* Para que esté bien definida deberán existir las derivadas parciales de  $\log q(x; \psi)$ , esto es equivalente a que las componentes de la mixtura sean derivables respecto de  $\psi$ , y además los productos

$$\left( \frac{\partial \log q(x; \psi)}{\partial \psi_i} \right) \left( \frac{\partial \log q(x; \psi)}{\partial \psi_j} \right) q(x, \psi)$$

deben ser integrables respecto de  $x$  para todo  $i, j$ .  $\square$

**Teorema 3.7.2.** Sea  $\mathcal{F} = \{f(\cdot, \theta) : \forall \theta = (\theta_1, \dots, \theta_d) \in \Theta\}$  una familia de funciones de densidad para las que las derivadas parciales respecto de  $\theta_i$  existen y son integrables respecto de  $x$  para todo  $i \in \{1, \dots, d\}$  y tales que

$$\int \frac{\left( \frac{\partial f(x, \theta)}{\partial \theta_i} \right)^2}{f(x, \theta)} dx < +\infty$$

entonces la matriz información de Fisher de cualquier mixtura finita de elementos de  $\mathcal{F}$  está bien definida.

*Demostración.* Para que la matriz información exista, según lo recogido en la Nota 3.7.2, deben ser integrables respecto de  $x$  cada uno de los productos

$$\left( \frac{\partial \log q(x; \psi)}{\partial \psi_i} \right) \left( \frac{\partial \log q(x; \psi)}{\partial \psi_j} \right) q(x, \psi)$$

que pueden ser expresados por:

$$\frac{\left( \frac{\partial q(x; \psi)}{\partial \psi_i} \right) \left( \frac{\partial q(x; \psi)}{\partial \psi_j} \right)}{q(x, \psi)}$$

Dependiendo de si las componentes de  $\theta$  con respecto a las que se deriva son los coeficientes de la mixtura o son parámetros de las componentes, el cociente anterior será una suma de funciones de la forma

$$\frac{f_i(x, \theta_i) f_j(x, \theta_j)}{q(x, \psi)}$$

ó

$$\frac{f_i(x; \theta_i) \pi_j \left( \frac{\partial f_j(x; \theta_j)}{\partial \theta_{jr}} \right)}{q(x, \psi)}$$

ó

$$\frac{\pi_i \pi_j \left( \frac{\partial f_i(x; \theta_i)}{\partial \theta_{is}} \right) \left( \frac{\partial f_j(x; \theta_j)}{\partial \theta_{jr}} \right)}{q(x, \psi)}$$



Usando la acotación

$$q(x, \psi) \geq \pi_i f_i(x, \theta_i) \quad \forall i = 1, \dots, k, \quad (3.3)$$

en el primer y segundo caso, esos sumandos en valor absoluto se pueden acotar por funciones integrables con el sumando de la mixtura que interese y, así, se obtendrá que los valores absolutos de los dos primeros cocientes son, respectivamente, menores o iguales que

$$\left| \frac{f_j(x, \theta_j)}{\pi_i} \right| \quad \text{y} \quad \left| \frac{\pi_j \left( \frac{\partial f_j(x; \theta_j)}{\partial \theta_{jr}} \right)}{\pi_i} \right|$$

que son funciones integrables.

Utilizando la desigualdad

$$ab \leq \frac{1}{2}(a^2 + b^2) \quad \forall a, b \in \mathbb{R}$$

se obtiene

$$\frac{\pi_i \pi_j \left( \frac{\partial f_i(x; \theta_i)}{\partial \theta_{is}} \right) \left( \frac{\partial f_j(x; \theta_j)}{\partial \theta_{jr}} \right)}{q(x, \psi)} \leq \frac{1}{2} \pi_i \pi_j \frac{\left[ \left( \frac{\partial f_i(x; \theta_i)}{\partial \theta_{is}} \right)^2 + \left( \frac{\partial f_j(x; \theta_j)}{\partial \theta_{jr}} \right)^2 \right]}{q(x, \psi)}$$

y usando la acotación de (3.3) se tiene que

$$\frac{\pi_i \pi_j \left( \frac{\partial f_i(x; \theta_i)}{\partial \theta_{is}} \right) \left( \frac{\partial f_j(x; \theta_j)}{\partial \theta_{jr}} \right)}{q(x, \psi)} \leq \frac{1}{2} \left[ \frac{\left( \frac{\partial f_i(x; \theta_i)}{\partial \theta_{is}} \right)^2}{f_i(x; \theta_i)} + \frac{\left( \frac{\partial f_j(x; \theta_j)}{\partial \theta_{jr}} \right)^2}{f_j(x; \theta_j)} \right]$$

y por lo tanto, aplicando las hipótesis del teorema, se tiene su integrabilidad.  $\square$

**Corolario 3.7.1.** Sean  $\mathcal{E}_i(a_i, t_i, b_i)$  familias exponenciales regulares definidas, respectivamente, sobre  $\Theta_i \subset \mathbb{R}^{d_i}$  con  $i \in \{1, \dots, k\}$  y tales que las funciones  $b_i$  tienen derivadas parciales de primer orden continuas en  $\Theta_i$ . Sea  $\psi^* = (\pi^*, \theta_1^*, \dots, \theta_k^*) \in \Omega = \mathcal{C} \times \Theta_1 \times \dots \times \Theta_k$ .

Entonces, la matriz información de Fisher de  $p(x, \psi) = \sum_{i=1}^k \pi_i f_i(x, \theta_i)$  con  $f_i(x, \theta_i) \in \mathcal{E}_i(a_i, t_i, b_i)$  está bien definida en un cierto compacto  $\tilde{\Omega} \subset \Omega$  que contiene a  $\psi = (\pi, \theta_1, \dots, \theta_k)$ .

*Demostración.* Dada  $f(x, \theta_i) \in \mathcal{E}_i(a_i, t_i, b_i)$  para algún  $i \in \{1, \dots, k\}$  se tiene que una derivada parcial será de la forma

$$\frac{\partial f(x; \theta_i)}{\partial \theta_{is}} = t_{is}(x) a_i(x) \frac{\partial b_i(\theta_i)}{\partial \theta_{is}} e^{\theta_i^T t_i(x)},$$

que al elevarla al cuadrado y dividirla por la propia función, quedaría

$$(t_{is}(x))^2 a_i(x) \left( \frac{\partial b_i(\theta_i)}{\partial \theta_{is}} \right)^2 e^{\theta_i^T t_i(x)} \frac{1}{b_i(\theta_i)}$$

y esta función es integrable respecto de  $x$  para  $\theta_i$  en un cierto compacto por la Proposición 3.2.2.

□

**Corolario 3.7.2.** Sea  $\psi^* = (\pi^*, \theta_1^*, \dots, \theta_k^*) \in \Omega = \mathcal{C} \times \Theta_1 \times \dots \times \Theta_k$  y sean  $\mathcal{W}_i(a_i, t_i, b_i, c_i)$   $k$  familias tipo  $\mathcal{W}$  definidas, respectivamente, sobre  $\Theta_i \subset \mathbb{R}^{d_i}$  con  $i \in \{1, \dots, k\}$  tales que las  $b_i$  tienen derivadas parciales de primer orden continuas en  $\Theta_i$  y las  $c_i$  son la función idénticamente nula o una función con derivadas parciales de primer orden continuas y alguna no nula en  $\beta_i^*$ . Entonces, la matriz información de Fisher de cualquier mezcla finita de la forma  $p(x, \psi) = \sum_{i=1}^k \pi_i f_i(x, \theta_i)$  con  $f_i(x, \theta_i) \in \mathcal{W}(a_i, t_i, b_i, c_i)$  está bien definida en un cierto compacto  $\hat{\Omega} \subset \Omega$  que contiene a  $\psi^* = (\pi^*, \theta_1^*, \dots, \theta_k^*)$ .

*Demostración.* Dada  $f(x, \theta_i) \in \mathcal{W}(a_i, t_i, b_i, c_i)$  para algún  $i \in \{1, \dots, k\}$ , por las condiciones del teorema se puede suponer  $b(\theta_i) = 1$ , si no fuese así la demostración hecha para  $\frac{f(x, \theta_i)}{b(\theta_i)}$  implica que se verifica para todo  $\frac{f(x, \theta_i)}{b(\theta_i)}$ , ya que, al ser continua su derivada primera y estar considerando un cierto compacto, está acotada y no afecta a la integrabilidad respecto de  $x$ .

Derivando respecto de alguna de las componentes de  $\theta_i$  se obtienen funciones de alguna de estas formas

$$t_{is}(x) a_i(x) \exp\{\alpha_i^T t_i(x) + c_i(\beta_i) e^{\lambda_i^T t_i(x)}\}$$

ó

$$a_i(x) \frac{\partial c_i(\beta_i)}{\partial \beta_{is}} e^{\lambda_i^T t_i(x)} \exp\{\alpha_i^T t_i(x) + c_i(\beta_i) e^{\lambda_i^T t_i(x)}\}$$

ó

$$t_{is}(x) a_i(x) e^{\lambda_i^T t_i(x)} \exp\{\alpha_i^T t_i(x) + c_i(\beta_i) e^{\lambda_i^T t_i(x)}\}$$

que al cuadrado y divididas por  $f(x, \theta_i)$  son de la forma

$$(t_{is}(x))^2 f(x, \theta_i)$$

ó

$$\left( \frac{\partial c_i(\beta_i)}{\partial \beta_{is}} \right)^2 e^{2\lambda_i^T t_i(x)} f(x, \theta_i)$$

ó

$$(t_{is}(x))^2 e^{2\lambda_i^T t_i(x)} f(x, \theta_i)$$

que son todas integrables, por la Proposición 3.2.3. □

**Corolario 3.7.3.** *La matriz información de una mezcla finita de funciones de la familia  $\mathcal{F}_L \cup \mathcal{F}_G \cup \mathcal{F}_W$  está bien definida.*

*Nota 3.7.2.* Demostrada la existencia de la matriz información, ésta es, por la propia definición, semidefinida positiva. Habrá que ver que es definida positiva, es decir, habrá que probar que para todo  $\xi \neq 0, \xi \in \mathbb{R}^n, \xi I(\psi) \xi^T > 0$ . Para ello se da la siguiente proposición.

**Proposición 3.7.1.** *Sea  $f(x, \psi)$  una función de densidad, la matriz información de Fisher en  $\psi^0, I(\psi^0)$ , es definida positiva si y sólo si el conjunto de funciones  $\{h_1(x, \psi^0), h_2(x, \psi^0), \dots, h_\nu(x, \psi^0)\}$  es linealmente independiente e.c.t. el conjunto  $S = \{x \in \mathbb{R}^n : f(x; \psi) \neq 0\}$ , donde  $h_i(x, \psi^0) = \frac{\partial f(x, \psi)}{\partial \psi_i} |_{\psi = \psi^0}$ .*

*Demostración.* Denotando por  $M(x, \psi)$  a  $[\nabla_\psi \log f(x; \psi)][\nabla_\psi \log f(x; \psi)]^T$ , sus componentes son

$$m_{ij} = \left( \frac{\partial \log f(x; \psi)}{\partial \psi_i} \right) \left( \frac{\partial \log f(x; \psi)}{\partial \psi_j} \right).$$

De esta forma,

$$\xi I(\psi) \xi^T = 0 \Leftrightarrow \sum_{i,j} m_{ij} \xi_i \xi_j f(x, \psi) = 0 \quad p.c.t \quad x \in \mathbb{R}^n \Leftrightarrow$$

$$\Leftrightarrow \exists Z \subset \mathbb{R} \text{ de medida nula} / (\nabla_\psi \log f(x; \psi) \xi)^2 f(x, \psi) = 0 \quad \forall x \notin Z$$

$$\Leftrightarrow \exists Z \subset \mathbb{R} \text{ de medida nula} / \forall x \in S \setminus Z, \left( \frac{1}{f} \nabla_\psi f(x; \psi) \xi \right)^2 = 0,$$

donde  $S = \{x \in \mathbb{R}^n : f(x; \psi) \neq 0\}$ .

Es decir que  $I(\Phi)$  será definida positiva si y sólo si

$$\forall x \in S \setminus Z \quad \nabla_\psi f(x; \psi) \xi = 0 \Rightarrow \xi = 0,$$

lo que equivale a afirmar que las derivadas parciales de  $f(x; \psi)$  respecto a las componentes de  $\psi$  son funciones linealmente independientes en  $S \setminus Z$ . □

**Teorema 3.7.3.** *La matriz información de Fisher asociada a una mezcla finita  $p(\cdot, \phi) \in \mathcal{P}$  es definida positiva.*

*Demostración.* Fijado el valor  $\psi$  del parámetro, habrá que comprobar que las siguientes funciones de  $x$  son linealmente independientes en  $(0, +\infty)$ :

$$h_1(x) = f_1(x, \psi) - f_3(x, \psi)$$

$$h_2(x) = f_2(x, \psi) - f_3(x, \psi)$$

$$h_3(x) = \pi_1 f_1(x, \psi) \frac{(\log x - \mu)}{\sigma^2}$$

$$h_4(x) = \pi_1 f_1(x, \psi) \left( \frac{(\log x - \mu)^2}{\sigma^3} - \frac{1}{\sigma} \right)$$

$$h_5(x) = \pi_2 f_2(x, \psi) (\log x - \log b - \text{Poly}(a))$$

$$h_6(x) = \pi_2 f_2(x, \psi) \frac{x}{b^2}$$

$$h_7(x) = (1 - \pi_1 - \pi_2) f_3(x, \psi) \left[ \frac{1}{c} + \log \left( \frac{x}{d} \right) - \left( \frac{c}{d} \right)^c \log \left( \frac{x}{d} \right) \right]$$

$$h_8(x) = (1 - \pi_1 - \pi_2) f_3(x, \psi) \left[ \frac{x^{c-1}}{d} - \frac{c}{xd} \right].$$

Para ello, se considera el siguiente orden parcial entre dichas funciones:

$$h_i < h_j \Leftrightarrow \lim_{x \rightarrow \infty} \frac{h_j(x)}{h_i(x)} = 0$$

y

$$h_i \sim h_j \Leftrightarrow \lim_{x \rightarrow \infty} \frac{h_j(x)}{h_i(x)} \in \mathbb{R} \setminus \{0\}.$$

De esta forma las ocho funciones quedarían ordenadas de la siguiente forma:

$$(c < 1)$$

$$h_4 < h_3 < h_1 < h_7 < h_2 < h_8 < h_6 < h_5$$

$$(c > 1) \text{ o } (c = 1 \text{ y } b > d) \text{ o } (c = 1, b < d, \text{ y } a > 2)$$

$$h_4 < h_3 < h_1 < h_6 < h_5 < h_2 < h_7 < h_8$$

$$(c = 1, \text{ y } b < d)$$

$$h_4 < h_3 < h_1 < h_7 < h_2 \sim h_8 < h_6 < h_5$$

$$(c = 1, b = d \quad y \quad 1 < a < 2)$$

$$h_4 < h_3 < h_1 < h_6 < h_7 < h_5 < h_2 < h_8$$

$$(c = 1, b = d \quad y \quad a < 1)$$

$$h_4 < h_3 < h_1 < h_7 < h_6 < h_2 \sim h_8 < h_5$$

$$(c = 1, b = d \quad y \quad a = 2)$$

$$h_4 < h_3 < h_1 < h_6 < h_5 \sim h_7 < h_2 < h_8$$

Para ver que son linealmente independientes, basta demostrar que si:

$$\sum_{i=1}^8 c_i h_i(x) = 0$$

entonces  $c_1 = \dots = c_8 = 0$ .

Siguiendo la misma idea de la demostración del Teorema 2.5.1, se dividen ambos miembros de la igualdad por la función más pequeña atendiendo al orden definido anteriormente,  $h_4(x)$  en cualquier caso,

$$\sum_{i=1}^8 c_i \frac{h_i(x)}{h_4(x)} = 0 \tag{3.4}$$

y tomando límite en el infinito, se obtiene que:

$$c_4 = 0.$$

Se elimina, pues, este sumando de la combinación lineal y se repite el proceso con la siguiente función más pequeña. De esta forma, se obtiene que los  $c_i$  son nulos hasta que, siguiendo las secuencias que marcan el orden en cada caso, se tope con dos funciones equivalentes bajo este orden. Existen tres casos donde esto ocurre, por lo que hay que analizar cada uno de ellos en particular.

- Caso ( $c = 1 \quad y \quad b < d$ )

En este caso por el procedimiento anterior se obtiene que  $c_4, c_3, c_1, c_7$  son 0. Dividiendo por  $h_2(x)$  y tomando límite se obtiene:

$$c_2 + c_8 \lim_{x \rightarrow \infty} \frac{h_8(x)}{h_2(x)} = 0.$$

Teniendo en cuenta que el límite de ese cociente es  $-\frac{1 - \pi_1 - \pi_2}{b} \neq 0$ , se tiene

$$c_2 - \frac{1 - \pi_1 - \pi_2}{b} c_8 = 0 \Leftrightarrow c_8 = \frac{b}{1 - \pi_1 - \pi_2} c_2$$

En consecuencia, en este caso (3.4) implica:

$$c_2 \left( h_2(x) + \frac{b}{1 - \pi_1 - \pi_2} h_8(x) \right) + c_6 h_6(x) + c_5 h_5(x) = 0.$$

Dado que, según el orden definido, se verifica

$$\left( h_2 + \frac{b}{1 - \pi_1 - \pi_2} h_8 \right) < h_6 < h_5$$

si se divide por  $h_2(x) + \frac{h_8(x)}{1 - \pi_1 - \pi_2}$  y se vuelve a tomar límite, se obtiene, que  $c_2 = 0$  y por tanto  $c_8$  también. Por el mismo razonamiento, se comprueba que  $c_5$  y  $c_6$  son nulos.

- Caso ( $c = 1, b = d$  y  $a < 1$ )

Procediendo según lo indicado anteriormente, se obtiene que  $c_4, c_3, c_1, c_7, c_6$  son nulos. Análogamente, al caso anterior, teniendo en cuenta que  $\lim_{x \rightarrow \infty} \frac{h_8(x)}{h_2(x)}$  es también  $-\frac{1 - \pi_1 - \pi_2}{b} \neq 0$  y que

$$\left( h_2 + \frac{b}{1 - \pi_1 - \pi_2} h_8 \right) < h_5$$

se deduce que  $c_2, c_8$  y  $c_5$  son también nulos.

- Caso ( $c = 1, b = d$  y  $a = 2$ )

Nuevamente, siguiendo el procedimiento indicado, se obtiene que  $c_4, c_3, c_1, c_6$  son 0. Y sabiendo que  $\lim_{x \rightarrow \infty} \frac{h_7(x)}{h_5(x)} = -\frac{1 - \pi_1 - \pi_2}{\pi_2} \neq 0$  y que

$$\left( h_5 + \frac{1 - \pi_1 - \pi_2}{\pi_2} h_7 \right) < h_2 < h_8$$

por el razonamiento anterior, se obtiene que  $c_5 = c_7 = c_2 = c_8 = 0$ .

Por tanto, en cualquiera de los casos, estas ocho funciones son linealmente independientes y, en consecuencia la matriz información es definida positiva.

□

En el siguiente capítulo, para la aplicación del algoritmo EM a las mixturas del problema que ocupa esta memoria, será necesario efectuar reparametrizaciones en algunas de las familias, por ese motivo se presenta la siguiente nota.

*Nota 3.7.3.* La reparametrización de la familia no cambia el carácter de definida positiva de la matriz información, ya que si  $\Phi$  es la matriz de la reparametrización que transforma  $q$  en  $q'$

$$(\nabla q)(\nabla q)^T = \Phi(\nabla q')(\nabla q')^T\Phi^T$$

al multiplicar a ambos lados por un  $\xi \neq 0$  se tiene:

$$\xi^T \nabla q(\nabla q)^T \xi = \xi^T \Phi(\nabla q')(\nabla q')^T \Phi^T \xi = \xi'(\nabla q')(\nabla q')^T (\xi')^T,$$

con  $\xi' \neq 0$ , y por tanto el primer miembro de la igualdad será 0 únicamente cuando lo sea el último.

## Capítulo 4

# Estimación de parámetros mediante el algoritmo EM

### 4.1 Introducción

Una vez modelados experimentos aleatorios con datos que presentan asimetría positiva, como es el caso de la variable estancia hospitalaria mediante una mixtura finita, y analizada la identificabilidad del modelo, así como las propiedades de optimalidad de los estimadores, el objetivo que se aborda a continuación es el cálculo de dichos estimadores.

Debido a que las mezclas de distribuciones bajo estudio constituyen una familia paramétrica dependiente de ocho parámetros, el cálculo de los correspondientes estimadores de máxima verosimilitud conlleva resolver un complejo sistema de ecuaciones no lineales. En vista de esta problemática, se ha optado por aproximar las soluciones de dicho sistema mediante la aplicación del algoritmo EM, que se expone en la sección segunda.

Las dos secciones siguientes estudian los elementos necesarios para la aplicación de dicho algoritmo, a saber, los valores iniciales y el criterio de parada.

Debido a la velocidad de convergencia que puede tener el algoritmo EM, se hace necesario introducir un método de aceleración que se presenta en la sección quinta.



En la sección sexta, se muestran los resultados obtenidos de la aplicación del algoritmo EM, usando el paquete informático Mathematica 4.1, a un conjunto de 128 muestras simuladas a partir del modelo. Para finalizar, dicho algoritmo se aplica a conjuntos de datos reales de la variable estancia hospitalaria.

## 4.2 Aplicación del algoritmo EM

Tal y como viene recogido en la introducción de esta memoria y en el Capítulo 1, existen una gran variedad de experimentos (biológicos, industriales, farmacológicos,...) que tienen asociadas colecciones de datos cuyo denominador común es la presencia de una considerable asimetría positiva y, que tales experiencias pueden ser modelizadas a través de mixturas mixtas con densidades de la forma:

$$q(x; \Psi) = \sum_{j=1}^3 \pi_j f_j(x; \theta_j)$$

donde  $\{\pi_i\}_{i=1}^3 \in \mathcal{C}$  y  $f_1$ ,  $f_2$  y  $f_3$  son funciones de densidad Lognormal, Gamma y Weibull, respectivamente, dadas por:

$$f_1(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma x}} \exp\left\{-\frac{(\log x - \mu)^2}{2\sigma^2}\right\}$$

$$f_2(x; a, b) = \frac{b^{-a} x^{a-1} \exp\{-x/b\}}{\Gamma(a)}$$

$$f_3(x; c, d) = \frac{c}{d^c} x^{c-1} \exp\{-x^c/d^c\}$$

Con el objetivo de calcular los estimadores de máxima verosimilitud, se considera una muestra  $\mathbf{x} = (x_1, \dots, x_n)$  de dicha función de densidad. La función de verosimilitud asociada viene dada por:

$$L(\Psi) = \prod_{i=1}^n \left[ \sum_{j=1}^3 \pi_j f_j(x_i; \theta_j) \right],$$

por lo tanto, la función  $L$  depende de los ocho parámetros componentes del vector  $\Psi = (\pi_1, \pi_2, \mu, \sigma, a, b, c, d)$ .

*Nota 4.2.1.* Al igual que para el caso de una mixtura de dos normales, como se apuntó en la Proposición 2.2.1, la función  $L$  no está acotada superiormente. En consecuencia, el interés se centrará en la búsqueda de máximos locales.  $\square$

Como viene recogido en el Capítulo 1, los máximos locales, en el caso de modelos mixtura, no son en general fáciles de encontrar de forma explícita y, el problema

que aquí se trata no es una excepción. Se procederá pues a aplicar el algoritmo EM, basado en la interpretación de los datos mixtura como un conjunto de datos incompletos.

Así, se consideran las observaciones  $x = (x_1, \dots, x_n)$  como un conjunto de datos incompleto  $x$  que proviene de un conjunto de datos completamente categorizados  $y = (y_1, \dots, y_n)$ , es decir, de los que se conoce de cuál de las tres distribuciones proceden, y que se pueden representar como:

$$\{y_i : i = 1, \dots, n\} = \{(x_i, z_i) : i = 1, \dots, n\},$$

donde cada  $z_i = (z_{i1}, z_{i2}, z_{i3})$  es un vector indicador de tres componentes con valor 1 en la posición que indica la categoría correspondiente y ceros en el resto.

Si se denota por  $g(y; \Psi)$  a la función de densidad correspondiente a  $y$ , se tiene que para cada  $y_i$ :

$$g(y_i; \Psi) = \prod_{j=1}^3 \pi_j^{z_{ij}} f_j(x_i; \theta_j)^{z_{ij}}$$

donde  $z_{ij}$  vale 1 en la categoría correspondiente y 0 en el resto, por lo tanto:

$$g(y_i; \Psi) = \pi_j f_j(x_i; \theta_j)$$

siendo  $j$  la categoría de la que se ha extraído  $x_i$ .

La función de verosimilitud para  $(y_1, \dots, y_n)$  viene dada por:

$$g(y_1, \dots, y_n; \Psi) = \prod_{i=1}^n \prod_{j=1}^3 \pi_j^{z_{ij}} f_j(x_i; \theta_j)^{z_{ij}}$$

siendo su logaritmo:

$$l(\Psi) = \sum_{i=1}^n z_i^T \begin{pmatrix} \log \pi_1 \\ \log \pi_2 \\ \log \pi_3 \end{pmatrix} + \sum_{i=1}^n z_i^T \begin{pmatrix} \log f_1(x_i; \theta_1) \\ \log f_2(x_i; \theta_2) \\ \log f_3(x_i; \theta_3) \end{pmatrix}.$$

En esta situación, el algoritmo EM genera desde un valor inicial de  $\Psi$ ,  $\Psi_0$ , una secuencia de estimadores  $\{\Psi^{(m)}\}_m$ . Cada iteración está constituida por los dos pasos siguientes:

**Paso E:** Evalúa  $\mathbb{E}[\log g(y; \Psi) | x, \Psi^{(m)}] = Q(\Psi, \Psi^{(m)})$ .

Paso M: Obtiene  $\Psi = \Psi^{(m+1)}$  que maximiza  $Q(\Psi, \Psi^{(m)})$  donde, recordando que  $z_1, \dots, z_n$  son cantidades desconocidas, se tiene:

$$Q(\Psi, \Psi^{(m)}) = \sum_{i=1}^n (w_i(\Psi^{(m)}))^T \begin{pmatrix} \log \pi_1 \\ \log \pi_2 \\ \log \pi_3 \end{pmatrix} + \sum_{i=1}^n (w_i(\Psi^{(m)}))^T \begin{pmatrix} \log f_1(x_i; \theta_1) \\ \log f_2(x_i; \theta_2) \\ \log f_3(x_i; \theta_3) \end{pmatrix} \quad (4.1)$$

siendo  $w_i(\Psi^{(m)}) = \mathbb{E}(z_i | x_i, \Psi^{(m)})$  para  $i = 1, \dots, n$ .

Teniendo en cuenta que los posibles valores para  $z_i$ , en esta situación, son  $e_1 = (1 \ 0 \ 0)$ ,  $e_2 = (0 \ 1 \ 0)$  y  $e_3 = (0 \ 0 \ 1)$ , para cada  $i$  se obtiene:

$$\mathbb{E}(z_i | x_i, \Psi^{(m)}) = k(e_1 | x_i) e_1 + k(e_2 | x_i) e_2 + k(e_3 | x_i) e_3$$

donde  $k(z|x)$  es la función de probabilidad de  $z$  dado  $x$ .

Así pues, se obtiene el vector de tres componentes

$$\mathbb{E}(z_i | x_i, \Psi^{(m)}) = [k(e_1 | x_i) \quad k(e_2 | x_i) \quad k(e_3 | x_i)]^T.$$

Por definición de probabilidad condicionada,  $k(z_i | x_i) = g(y_i; \Psi) / q(x_i; \Psi)$  y, de esta forma, las tres componentes de  $w_i(\Psi^{(m)})$  son:

$$\begin{aligned} w_{i1}(\Psi^{(m)}) &= \frac{\pi_1^{(m)} f_1(x_i; \mu^{(m)}, \sigma^{(m)})}{q(x_i; \Psi^{(m)})}, \\ w_{i2}(\Psi^{(m)}) &= \frac{\pi_2^{(m)} f_2(x_i; a^{(m)}, b^{(m)})}{q(x_i; \Psi^{(m)})} \quad \text{y} \\ w_{i3}(\Psi^{(m)}) &= \frac{(1 - \pi_1^{(m)} - \pi_2^{(m)}) f_3(x_i; c^{(m)}, d^{(m)})}{q(x_i; \Psi^{(m)})}. \end{aligned}$$

Como se ha comentado en la sección 1.5, el problema M de maximización se separa en dos problemas de optimización, interviniendo en el primero de ellos únicamente los pesos,  $\pi_j$ , mientras que el resto de los parámetros aparecen en el segundo. Así, la solución para los pesos  $\pi_j$ , aplicando los multiplicadores de Lagrange, viene dada por:

$$\pi_j^{(m+1)} = \frac{\sum_{i=1}^n w_{ij}(\Psi^{(m)})}{n} \quad j = 1, 2, 3.$$

El problema se centra ahora en la maximización del segundo de los sumandos de  $Q(\Psi, \Psi^{(m)})$ :

$$\sum_{i=1}^n (w_{i1} \log f_1(x_i; \mu, \sigma) + w_{i2} \log f_2(x_i; a, b) + w_{i3} \log f_3(x_i; c, d)).$$

Derivando respecto de cada parámetro e igualando a 0, se obtiene el siguiente sistema de seis ecuaciones no lineales de incógnitas  $\mu$ ,  $\sigma$ ,  $c$ ,  $b$ ,  $a$  y  $d$ :

$$\begin{cases} \sum_{i=1}^n w_{i1} \frac{\log x_i - \mu}{\sigma^2} = 0 \\ \sum_{i=1}^n w_{i1} \frac{(\log x_i - \mu)^2 - \sigma^2}{\sigma^3} = 0 \end{cases} \quad (4.2)$$

$$\begin{cases} \sum_{i=1}^n w_{i2} [\log x_i - \log b - \text{Poly}[a]] = 0 \\ \sum_{i=1}^n w_{i2} \left[ \frac{x_i}{b^2} - \frac{a}{b} \right] = 0 \end{cases} \quad (4.3)$$

$$\begin{cases} \sum_{i=1}^n w_{i3} \left[ 1 + \log \frac{x_i^c}{d^c} - \frac{x_i^c}{d^c} \log \frac{x_i^c}{d^c} \right] = 0 \\ \sum_{i=1}^n w_{i3} c \frac{x_i^c - d^c}{d^{c+1}} = 0 \end{cases} \quad (4.4)$$

Este sistema consta, a su vez, de tres sistemas de dos ecuaciones con dos incógnitas, dividiéndose el problema de maximización del segundo sumando de  $Q(\Psi, \Psi^{(m)})$  en tres problemas componentes. Además, estos tres sistemas se corresponden, salvo por las ponderaciones  $w_{ij}$ , con los sistemas de ecuaciones resultantes del cálculo de estimadores de máxima verosimilitud de cada una de las funciones por separado. De este modo, la dificultad de resolución de cada uno de los tres sistemas es la misma que la del cálculo de los estimadores para cada uno de los casos.

La resolución del sistema (4.2), en el que intervienen los parámetros de la distribución Lognormal, es directa.

**Proposición 4.2.1.** *La solución del sistema (4.2) viene dada por:*

$$\mu = \frac{\sum_{i=1}^n w_{i1} \log x_i}{\sum_{i=1}^n w_{i1}} \quad \sigma^2 = \frac{\sum_{i=1}^n w_{i1} (\log x_i - \mu)^2}{\sum_{i=1}^n w_{i1}}.$$

No ocurre lo mismo con los otros dos sistemas, que se tendrán que resolver en cada caso haciendo uso de métodos numéricos.

En el caso del sistema (4.3), donde intervienen los parámetros  $a$  y  $b$  de la distribución Gamma, se utiliza el método de Newton-Rapshon para resolver la primera ecuación, ya que la estimación del parámetro  $b$  se obtiene de forma explícita conocido  $a$ :

$$b = \frac{\sum_{i=1}^n w_{i2} x_i}{a \sum_{i=1}^n w_{i2}}.$$

Sustituyendo el valor de  $b$  en la segunda ecuación resulta la siguiente ecuación en  $a$ :

$$\log \left[ \frac{\sum_{i=1}^n w_{i2} x_i}{\sum_{i=1}^n w_{i2}} \right] - \frac{\sum_{i=1}^n w_{i2} \log x_i}{\sum_{i=1}^n w_{i2}} - \log a + \text{Poly}[a] = 0.$$

Por último, en el caso de los parámetros de la distribución Weibull, se resuelve el sistema de ecuaciones no lineales (4.4) mediante el método de Newton-Rapshon para el caso de dos ecuaciones, ya que no es posible reducir el problema a una sola ecuación como en el caso anterior.

Una vez resueltos estos sistemas y calculados los valores de  $\Psi$  en la etapa M, se vuelve al paso E, en el que se actualizan los valores de los  $w_{ij}$  con los nuevos parámetros y de nuevo se plantea otro sistema.

### 4.3 Elección de los valores iniciales

Otro de los problemas que ha de resolverse es la elección de adecuados valores iniciales para el algoritmo, que aseguren la convergencia, con una velocidad aceptable, hacia el mayor máximo local. McLachlan [40] propone algunos métodos para la elección de los valores iniciales utilizando particiones de las observaciones. Sin embargo, estos métodos no parecen apropiados para este caso concreto, debido a que cada una de las componentes de la mixtura podrían ajustar bien el total de las observaciones, y no solamente una partición suya.

En este caso, la elección de los valores iniciales para  $\Psi$ ,  $\Psi^{(0)}$ , conlleva la

elección, por una parte, de los pesos iniciales de las funciones componentes,  $\pi_j^{(0)}$  para  $j = 1, 2, 3$ , y por otra, la elección de los valores iniciales para los parámetros de cada una de dichas funciones,  $\mu^{(0)}, \sigma^{(0)}, a^{(0)}, b^{(0)}, c^{(0)}$  y  $d^{(0)}$ .

#### 4.3.1 Valores iniciales para los pesos

Para la elección de los valores iniciales se pueden considerar diferentes estrategias. Entre ellas, en particular, las dos siguientes.

Teniendo en cuenta que el algoritmo EM busca estimaciones de los parámetros que maximicen la función de verosimilitud, en un primer momento se consideró ponderar el peso de cada una de las funciones de densidad de acuerdo con los valores que presentan sus funciones de verosimilitud, considerando como valores para los parámetros sus valores iniciales calculados en la siguiente subsección. Ya que el valor de la función de verosimilitud depende del tamaño de la muestra, para eliminar esta influencia se calculan sus raíces  $n$ -ésimas. De esta forma, llamando

$$fp_1 = \left( \prod_{i=1}^n f_1(x_i, \mu^0, \sigma^0) \right)^{\frac{1}{n}},$$

$$fp_2 = \left( \prod_{i=1}^n f_2(x_i, a^0, b^0) \right)^{\frac{1}{n}},$$

$$fp_3 = \left( \prod_{i=1}^n f_3(x_i, c^0, d^0) \right)^{\frac{1}{n}}$$

los valores iniciales de los pesos serían

$$\pi_j^0 = \frac{fp_j}{s_f} \quad j = 1, 2, 3,$$

donde  $s_f = fp_1 + fp_2 + fp_3$ .

Los resultados obtenidos mediante esta estrategia coinciden en la gran mayoría de los casos con los de la segunda estrategia expuesta a continuación y que fue la que finalmente se llevó a cabo, a pesar de que precisa de un mayor número de iteraciones.

Esta segunda estrategia consiste en considerar distintos valores iniciales para los pesos, que recorran de forma suficientemente exhaustiva el rango de valores

posibles, obteniéndose un conjunto de soluciones de entre las cuales, el mayor máximo local se elige comparando los valores de la función de verosimilitud. De esta forma se asegura la convergencia evitando caer en regiones donde la verosimilitud es casi constante y, por tanto, pudiendo ocurrir que el algoritmo EM no converja a un máximo local.

En una revisión que hacen del tema Karlis y Xekalaki [27], aparece esta segunda estrategia y se sugiere además utilizar un método mixto, empezando por distintos valores iniciales, haciendo un pequeño número de iteraciones sin examinar necesariamente la convergencia y, entonces, continuar el algoritmo hasta su convergencia desde el punto con mayor verosimilitud después de esas iteraciones iniciales. Esta estrategia ayudaría a reducir la cantidad de tiempo usado en áreas de verosimilitud plana, lejos del mayor máximo local.

### 4.3.2 Valores iniciales de los parámetros de las funciones componentes

Para la búsqueda de los valores iniciales de los parámetros de las funciones componentes, se ha optado por tomar los estimadores que maximizan la verosimilitud dada por cada una de las distribuciones por separado. Por ejemplo, se han calculado los estimadores de máxima verosimilitud de  $\mu$  y  $\sigma$  suponiendo que los datos provienen exclusivamente de una variable Lognormal:

$$\mu^{(0)} = \frac{\sum_{i=1}^n \log x_i}{n} \quad \text{y} \quad \sigma^{(0)} = \sqrt{\frac{\sum_{i=1}^n (\log x_i - \mu^{(0)})^2}{n}}.$$

Sin embargo, al intentar hacer este mismo razonamiento en el caso de las distribuciones Gamma y Weibull, aparece de nuevo el problema de que éstos no pueden obtenerse de forma explícita, por lo que sería necesario aplicar un método numérico para calcularlos, que necesitaría a su vez de unos valores iniciales, volviendo a aparecer el mismo problema. Así, para estos dos casos, se han elegido como valores iniciales estimadores calculados a partir de cada distribución por separado y que no necesitan algoritmos numéricos para su cómputo efectivo. Ambos métodos están recogidos en el Capítulo 1.

En el caso de la distribución Gamma se ha optado por las aproximaciones a

los estimadores de máxima verosimilitud dados por Greenwood y Durand [17]

$$a^{(0)} = s(M) \quad \text{y} \quad b^{(0)} = \frac{\sum_{i=1}^n x_i}{a^{(0)}n},$$

con  $s(M)$  definida como en la sección 1.2.2.

Para la distribución Weibull, se han elegido los estimadores propuestos por el método de estimación de Bain y Antle [5],

$$c^{(0)} = \frac{\prod_{i=1}^n x_{(i)}^{\frac{1}{n}}}{\prod_{i=1}^n \varepsilon_i^{\frac{1}{nd}}} \quad \text{y} \quad d^{(0)} = \frac{\sum_{i=1}^n \log \varepsilon_i \log x_{(i)} - \frac{1}{n} \sum_{i=1}^n \log \varepsilon_i \sum_{i=1}^n \log x_{(i)}}{\sum_{i=1}^n \log^2 x_{(i)} - \frac{1}{n} \left( \sum_{i=1}^n \log x_{(i)} \right)^2},$$

donde  $\varepsilon_i$  están definidos como en la sección 1.2.3.

## 4.4 Criterio de parada

La tarea de elegir un criterio de parada apropiado ha merecido la atención de muchos autores (véase, por ejemplo, Lindstrom y Bates [36]). Las reglas usadas se basan en la diferencia entre los valores de los parámetros y/o de la verosimilitud en etapas consecutivas, indicando falta de progreso más que convergencia. La característica común en todos ellos es que el algoritmo se detiene cuando los valores según los criterios elegidos son menores que una constante especificada. Cuanto más pequeña es dicha constante, más exigente es el criterio. De este modo, si el algoritmo sigue funcionando, se produce posiblemente una mejora sustancial en la verosimilitud.

Debido a que el tamaño de la verosimilitud depende significativamente del tamaño de la muestra, es preferible fijar la parada del algoritmo cuando la variación de los parámetros entre una etapa y la anterior sea menor que cierta constante. Dicha variación se mide a partir de la suma de los valores absolutos de las diferencias entre los valores de cada parámetro en dos etapas consecutivas. Se toma como constante  $10^{-5}$ , atendiendo al valor usado por otros autores como Karlis y Xekalaki [27]. En la sección 4.7 se podrá comprobar que, teniendo en cuenta los valores que toman los parámetros en el caso de la variable estancia hospitalaria, esta constante proporciona un criterio de parada bastante estricto.



## 4.5 Aceleración del algoritmo EM

Como se ha comentado en el capítulo primero, una de las críticas que se puede hacer al algoritmo EM es su velocidad de convergencia. Si además, se está tratando con un número elevado de parámetros, como es este caso, y se añade la complejidad de tener que aplicar métodos iterativos en el paso M de cada iteración, su aplicación puede resultar excesivamente lenta.

Varios autores, como Jamshidian y Jennrich [22], Lange[31], Lansky y Casella [32], Meilijson [41] y Meng y Rubin [42], han estudiado métodos para acelerar el algoritmo EM. Una revisión en profundidad de dichos aceleradores, así como una comparación entre ellos, puede encontrarse en Jamshidian y Jennrich [23].

Como se indica en la última referencia, debido al alto número de parámetros, muchos de estos aceleradores no son recomendables en este caso, algunos de ellos por cuestiones de dificultad computacional, como en el método propuesto por Lange [31], que requiere la inversión de una matriz de alta dimensión en cada etapa. Siguiendo la recomendación dada en Jamshidian y Jennrich [23], se ha optado por el *método de aceleración del gradiente conjugado* presentado por Jamshidian y Jenrich [22].

### 4.5.1 Algoritmo del gradiente conjugado generalizado

La herramienta básica de este algoritmo es el concepto de gradiente generalizado que, a continuación, se presenta.

**Definición 4.5.1.** *Sea  $W$  una matriz  $p \times p$  simétrica definida positiva, y considérese la norma euclídea asociada,*

$$\|\theta\|_W = (\theta^T W \theta)^{\frac{1}{2}}.$$

*Sea  $f(\theta)$  una función regular y  $g(\theta)$  el gradiente de  $f(\theta)$ . Se llama gradiente generalizado de  $f(\theta)$  definido por  $W$  al vector gradiente de  $f(\theta)$  con respecto a esta norma, es decir*

$$\tilde{g}(\theta) = W^{-1}g(\theta).$$

Dado  $\theta^0$ , sea  $d^0 = \tilde{g}^0 = \tilde{g}(\theta^0)$ , el algoritmo del gradiente conjugado generalizado procede secuencialmente según el siguiente esquema:

Etapa k: Para  $k \geq 1$  se determinan el punto  $\theta^k$  y la dirección  $d^k$  como:

$$\theta^k = \theta^{k-1} + \alpha^k d^{k-1} \quad \text{y} \quad d^k = \tilde{g}^k - \beta^k d^{k-1}$$

donde

$$\alpha^k = \sup_{\alpha} f(\theta^{k-1} + \alpha d^{k-1}) \quad \text{y}$$

$$\beta^k = \frac{(\tilde{g}^k)^T (g^k - g^{k-1})}{(d^k)^T (g^k - g^{k-1})}$$

con  $g^k = g(\theta^k)$  y  $\tilde{g}^k = \tilde{g}(\theta^k)$ .

#### 4.5.2 Aplicación del algoritmo del gradiente conjugado a la aceleración del algoritmo EM

El algoritmo EM dado por Dempster et al. [15] pretende encontrar el vector paramétrico  $\tilde{\Psi}$  que maximiza una función de verosimilitud  $L(\Psi)$ ,  $\Psi \in \Omega$ . Para su aplicación se define una función  $Q(\Psi', \Psi)$  que puede verse como una aproximación local de  $\log L(\Psi')$  en un entorno de  $\Psi$ . Sea  $\tilde{\Psi}$  el valor de  $\Psi'$  que maximiza  $Q(\Psi', \Psi)$ .

Como se demuestra en Jamshidian y Jennrich [22], si  $\hat{\Psi}$  es un punto interior de  $\Omega$ :

$$\tilde{\Psi} - \Psi = -(\ddot{Q}(\hat{\Psi}, \hat{\Psi}))^{-1} g(\Psi) + o(\Psi - \hat{\Psi}),$$

donde  $g(\Psi)$  es el gradiente de  $\log L(\Psi')$  en  $\Psi$  y  $\ddot{Q}(\hat{\Psi}, \hat{\Psi})$  es la matriz Hessiana de  $Q(\Psi', \Psi)$  considerada como función de  $\Psi'$ , y evaluada en  $(\Psi', \Psi) = (\hat{\Psi}, \hat{\Psi})$ .

Usualmente, la matriz  $\ddot{Q}(\hat{\Psi}, \hat{\Psi})$  es definida negativa. De esta forma, cuando  $\Psi$  está cerca de  $\hat{\Psi}$ , la diferencia entre dos pasos del algoritmo EM,  $\tilde{\Psi} - \Psi$  es un gradiente generalizado de  $\log L(\Psi')$  respecto de  $-\ddot{Q}(\hat{\Psi}, \hat{\Psi})$ .

Así, la clave del método de aceleración consiste en usar dicha diferencia,  $\tilde{\Psi} - \Psi$ , como el gradiente generalizado  $\tilde{g}(\Psi)$  en el algoritmo presentado en el apartado anterior, donde  $f(\Psi) = \log L(\Psi)$ . Es decir, en este caso,

$$W = -\ddot{Q}(\hat{\Psi}, \hat{\Psi}).$$

Este algoritmo precisa en cada etapa el cálculo de los gradientes de  $\log L(\Psi)$  y la inversa de la matriz  $\ddot{Q}(\Psi', \Psi)$ , que como se verá, es diagonal por cajas, por

lo que el cálculo se simplifica bastante. El paso más complejo que presenta este algoritmo es la búsqueda de  $\alpha^k$ , en cada etapa, ya que se trata de maximizar una función de una sola variable, pero muy compleja. Para solucionar este problema, se han seguido las indicaciones de Jamshidian y Jenrich [22] que proponen aplicar el método de la secante (vease Johnson y Riess [24]), con algunos detalles añadidos para evitar que  $\alpha^k$  pueda ser negativo.

### 4.5.3 Aplicación del algoritmo del gradiente conjugado al modelo de mixturas mixtas $\mathcal{P}$

Sea  $\mathcal{L}(\Psi) = \log L(\Psi)$  la función que se quiere maximizar y sea  $Q(\hat{\Psi}, \hat{\Psi})$  la función definida por (4.1). En estas circunstancias se plantea un problema adicional, ya que uno de los requisitos que se deben cumplir para que el método funcione es que  $W$  sea definida positiva, es decir, que la matriz Hessiana de  $Q(\hat{\Psi}, \hat{\Psi})$  sea una matriz definida negativa. En nuestro caso, aunque esto no sucede, puede conseguirse sin más que reparametrizar las funciones de densidad adecuadamente. Como se indicó en el capítulo 3, (véase la Nota 3.7.3), esta reparametrización no cambia el carácter de definida positiva de la matriz información y, por tanto, no afecta a los resultados de consistencia obtenidos en el capítulo anterior.

En este caso, la matriz  $\ddot{Q}(\hat{\Psi}, \hat{\Psi})$  de dimensión  $8 \times 8$  es diagonal por cajas  $2 \times 2$ , correspondiendo cada una de las 4 cajas a la matriz de segundas derivadas de  $Q(\hat{\Psi}, \hat{\Psi})$  respecto de las parejas de parámetros  $(\pi_1, \pi_2)$ ,  $(\sigma, \mu)$ ,  $(a, b)$  y  $(c, d)$ , respectivamente. Como se comprueba fácilmente, una matriz diagonal por cajas es definida negativa si y sólo si lo es cada caja.

La primera caja es:

$$\sum_{i=1}^n \begin{bmatrix} -\frac{w_{i1}}{\pi_1^2} - \frac{w_{i3}}{(1 - \pi_1 - \pi_2)^2} & -\frac{w_{i3}}{(1 - \pi_1 - \pi_2)^2} \\ -\frac{w_{i3}}{(1 - \pi_1 - \pi_2)^2} & -\frac{w_{i2}}{\pi_2^2} - \frac{w_{i3}}{(1 - \pi_1 - \pi_2)^2} \end{bmatrix}$$

que es definida negativa, pues su primer elemento diagonal es negativo y su determinante

$$\sum_{i,j=1}^n \frac{w_{i1}w_{j2}}{\pi_1^2\pi_2^2} + \sum_{i,j=1}^n \left( \frac{w_{i1}}{\pi_1^2} + \frac{w_{i2}}{\pi_2^2} \right) \frac{w_{j3}}{(1 - \pi_1 - \pi_2)^2}$$

es positivo.

Desgraciadamente, no ocurre lo mismo con el resto de las cajas. Es, por tanto, necesario encontrar una reparametrización adecuada. Los siguientes lemas pro-

porcionan estas reparametrizaciones para cada una de las cajas asociadas a las funciones componentes del modelo.

**Lema 4.5.1.** Para una distribución Lognormal de parámetros  $\mu$  y  $\sigma$ , la reparametrización

$$\alpha = 1/2\sigma^2, \quad \beta = \mu/\sigma^2,$$

tiene asociada una caja definida negativa en la matriz Hessiana de  $Q(\hat{\Psi}, \hat{\Psi})$ .

*Demostración.* La caja asociada en este caso viene dada por

$$\left( \sum_{i=1}^n w_{i1} \right) \begin{bmatrix} -\frac{\alpha + \beta^2}{2\alpha^3} & \frac{\beta}{2\alpha^2} \\ \frac{\beta}{2\alpha^2} & -\frac{1}{2\alpha} \end{bmatrix},$$

que es definida negativa, ya que el primer término diagonal es negativo y el determinante es

$$\frac{1}{4\alpha^3} \left( \sum_{i=1}^n w_{i1} \right)^2 > 0.$$

□

**Lema 4.5.2.** Para una distribución Gamma de parámetros  $a$  y  $b$ , la reparametrización

$$a = a \quad y \quad \tau = 1/b$$

tiene asociada una caja definida negativa en la matriz Hessiana de  $Q(\hat{\Psi}, \hat{\Psi})$ .

*Demostración.* La caja asociada en este caso viene dada por

$$\left( \sum_{i=1}^n w_{i2} \right) \begin{bmatrix} -\text{Poly}'[a] & 1/\tau \\ 1/\tau & -\frac{a}{\tau^2} \end{bmatrix}$$

que es definida negativa, pues su primer elemento diagonal es negativo y su determinante es

$$\left( \sum_{i=1}^n w_{i2} \right)^2 \frac{a\text{Poly}'[a] - 1}{\tau^2},$$

dado que  $\text{Poly}'[a] > 1/a$ , también es positivo.

□

**Lema 4.5.3.** Para una distribución Weibull de parámetros  $c$  y  $d$ , la reparametrización

$$c = c, \quad y \quad \gamma = -\log d,$$

tiene asociada una caja definida negativa en la matriz hessiana de  $Q(\hat{\Psi}, \hat{\Psi})$ .

*Demostración.* La caja asociada,

$$\left( \sum_{i=1}^n w_{i3} \right) \begin{bmatrix} -\frac{1}{c^2} - e^{\gamma x^c} \log^2 x & -e^{\gamma x^c} \log x \\ -e^{\gamma x^c} \log x & -e^{\gamma x^c} \end{bmatrix},$$

tiene determinante  $\frac{e^{\gamma x^c}}{c^2}$ , positivo, y su primer elemento diagonal negativo. Luego, al igual que en los casos anteriores, es definida negativa.  $\square$

Encontradas las reparametrizaciones adecuadas, se define la función  $\mathcal{L}(\Psi)$  en términos de los nuevos parámetros. Sea  $\hat{\Psi} = (\pi_1, \pi_2, \alpha, \beta, a, \tau, c, \gamma)$  al vector de nuevos parámetros y, así, se tiene que la expresión del logaritmo de la verosimilitud en función del nuevo vector paramétrico,  $\mathcal{L}(\hat{\Psi})$ , es:

$$\sum_{i=1}^n \log \left[ \pi_1 \frac{\sqrt{\alpha}}{\sqrt{\pi x_i}} e^{-\alpha(\log x_i - \frac{\beta}{2\alpha})^2} + \pi_2 \frac{\tau^a}{\Gamma(a)} e^{-\tau x_i} x_i^{a-1} + \pi_3 c x_i^{c-1} e^{\gamma - x_i^c e^{\gamma}} \right]$$

donde  $\pi_3 = 1 - \pi_1 - \pi_2$ .

En este caso, el gradiente de  $\mathcal{L}(\hat{\Psi})$  respecto de  $\hat{\Psi}$  viene dado por el vector

$$\left( \sum_{i=1}^n \frac{1}{q(x_i, \hat{\Psi})} \right) \begin{bmatrix} \frac{\sqrt{\alpha}}{\sqrt{\pi x_i}} e^{-\alpha(\log x_i - \frac{\beta}{2\alpha})^2} - c x_i^{c-1} e^{\gamma - x_i^c e^{\gamma}} \\ \frac{\tau^a}{\Gamma(a)} \exp\{-\tau x_i\} x_i^{a-1} - c x_i^{c-1} e^{\gamma - x_i^c e^{\gamma}} \\ \frac{\pi_1 e^{-\alpha \log^2 x_i - \frac{\beta^2}{4\alpha} + \beta \log x_i}}{\sqrt{\pi x_i}} \left[ \frac{1}{2\sqrt{\alpha}} - \sqrt{\alpha} \left( \log^2 x_i - \frac{\beta^2}{4\alpha^2} \right) \right] \\ \frac{\pi_1 e^{-\alpha \log^2 x_i - \frac{\beta^2}{4\alpha} + \beta \log x_i}}{\sqrt{\pi x_i}} \sqrt{\alpha} \left( \log x_i - \frac{\beta}{2\alpha} \right) \\ \frac{\pi_2 \tau^a e^{-\tau x_i} x_i^{a-1} [\log \tau + \log x_i - \text{Poly}(a)]}{\Gamma(a)} \\ \frac{\pi_2 \tau^{(a-1)} e^{-\tau x_i} x_i^{a-1} [a - x_i \tau]}{\Gamma(a)} \\ (1 - \pi_1 - \pi_2) x_i^{c-1} e^{\gamma - x_i^c e^{\gamma}} [1 - e^{\gamma} x_i^c \log x_i + c \log x_i] \\ (1 - \pi_1 - \pi_2) c x_i^{c-1} e^{\gamma - x_i^c e^{\gamma}} [1 - x_i^c e^{\gamma}] \end{bmatrix}$$

Con los resultados obtenidos en esta sección se puede aplicar directamente la aceleración del gradiente conjugado generalizado al modelo planteado.

## 4.6 Simulación

En esta sección, se presentan los resultados de aplicar el algoritmo EM acelerado, propuesto anteriormente, a 128 muestras de tamaño 1000, generadas por mixturas mixtas para distintos valores de cada uno de los ocho parámetros del modelo.

Dada una combinación de parámetros, cada valor de la muestra correspondiente se ha generado mediante los siguientes pasos:

- 1) Se ha dividido el intervalo  $[0, 1]$  en tres partes en función de los pesos,
- 2) Se ha generado un valor aleatorio de dicho intervalo siguiendo una ley uniforme,
- 3) Se ha generado un valor aleatorio a partir de una ley Lognormal, Gamma o Weibull, en función del valor obtenido en 2.

Las distintas combinaciones de parámetros de las tres funciones componentes se han obtenido mediante la asignación de dos valores diferentes a cada uno de ellos, obteniéndose por tanto 64 combinaciones posibles de los mismos. Para la elección de los dos valores de cada parámetro, previamente, se tomaron muestras de valores reales de la estancia hospitalaria y se ajustaron con cada una de las funciones componentes por separado con el fin de conocer los posibles rangos de valores para dichos parámetros.

Se han generado dos muestras para cada una de las 64 combinaciones anteriores, cambiando los valores de los pesos que en los dos casos fueron elegidos al azar. Por tanto, en total se han obtenido 128 muestras.

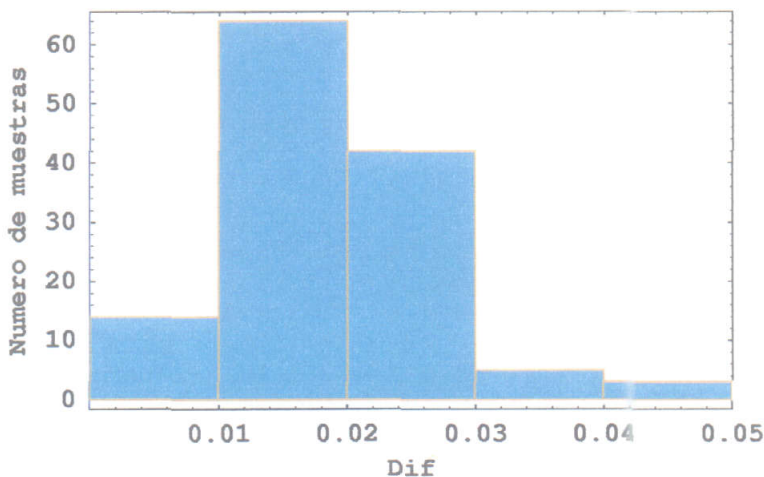
En el Anexo A, para cada muestra se presenta un gráfico con el histograma obtenido de la misma, la función de densidad de la distribución original (en color verde y con línea discontinua) y la función de densidad de la distribución estimada a través del algoritmo (en rojo y con línea continua).

Como medida de la discrepancia entre la función de distribución original,  $F(x; \Psi)$  y la estimada  $F(x; \hat{\Psi})$ , se ha optado por la métrica uniforme (Zolotarev [56]), también conocida como métrica de Kolmogorov (Kolmogorov [29]). Es decir,

$$d[F(x; \Psi), F(x; \hat{\Psi})] = \sup_{x \in \mathbb{R}} |F(x; \Psi) - F(x; \hat{\Psi})|.$$

Esta métrica toma valores entre  $[0, 1]$ , lo que permite tener un intervalo de referencia para interpretar adecuadamente los valores de la misma. En las gráficas se denota por **dif**.

A modo de resumen, de las 128 simulaciones realizadas, se puede concluir que los resultados obtenidos deben ser considerados como satisfactorios, dado que, como se puede observar en la siguiente gráfica, el 93.8% de ellas tienen asociado un valor menor de 0.03 en la métrica uniforme y ninguna de ellas supera el valor 0.05.



## 4.7 Aplicación a datos reales de la estancia hospitalaria

La estancia hospitalaria es un indicador, ampliamente utilizado y fácilmente disponible, de la actividad hospitalaria y se usa para varios propósitos, como la dirección, planificación y control de calidad del servicio hospitalario. El motivo es que constituye un estimador indirecto del consumo de recursos y de la eficiencia en uno de los aspectos del cuidado del paciente hospitalizado: la gestión de camas.

Los Grupos Relacionados con el Diagnóstico (GRD), creados por un equipo

de la Universidad de Yale, constituyen un sistema de clasificación de episodios de hospitalización con definiciones clínicamente reconocibles, y en la que se espera que los pacientes de cada clase consuman una cantidad similar de recursos, como resultado de un proceso de cuidados hospitalarios semejantes. La estancia media se utiliza como indicación del consumo de recursos por su disponibilidad y por la buena relación demostrada con los costes incurridos. Por tanto, podría decirse en parte, que los GRDs se han construido para conseguir grupos homogéneos respecto al consumo de servicios y costos, variables con las que la estancia hospitalaria está estrechamente relacionada. Más aún, la comparación de la estancia media por GRD permite extraer conclusiones de un aspecto determinado de la gestión de cada tipo específico de pacientes.

Generalmente, las distribuciones de la estancia hospitalaria son asimétricas y contienen outliers. Estos dos hechos son obstáculos serios para el uso de la media aritmética como un estimador de la esperanza de la estancia hospitalaria, y limita el uso de técnicas de inferencia estadística basadas en la hipótesis de normalidad.

Actualmente, en la práctica, para abordar estos problemas, se usan varias reglas de ajuste, algunas de ellas basadas en el modelo Lognormal. Esta elección está motivada por la gran cola observada en la mayoría de las distribuciones de la estancia hospitalaria. Como consta en la literatura, sin embargo, hay una gran variedad de diseños de distribución de la estancia hospitalaria, muchos de ellos bastante alejados del modelo Lognormal. Por otra parte, el progreso en Estadística y Computación proporciona nuevas herramientas para el ajuste de modelos robustos, así como el tratamiento automatizado de grandes conjuntos de muestras. En esta perspectiva, Marazzi *y otros* [37] enfocan la modelización de esta variable a través de tres familias de distribuciones: Lognormal, Weibull y Gamma (estas dos últimas caracterizadas por colas menos pesadas).

Marazzi *y otros* [37] usan una base de datos que contiene estancias hospitalarias de cinco países europeos, clasificada según el país, el año y el GRD. En un primer paso, ajustan cada uno de los modelos mencionados a cada muestra, para luego, en un segundo paso, decidir sobre la adecuación a uno sólo de ellos usando dos herramientas: una versión robusta del test de Cox, para modelos no anidados, y una medida, ATAR, usada en los casos en los que el modelo de Cox acepta a más de uno o rechaza a todos. Así, para cada muestra, se llevan a cabo seis tests de Cox, combinando los modelos de dos en dos, entre la hipótesis nula y la alternativa. El proceso concluye seleccionando uno de los tres modelos y rechazando los dos restantes.

Posteriormente, otros autores abordaron el problema a través de las mixturas, pero, exclusivamente, de distribuciones de una misma familia de las tres anterior-



mente citadas. Por ejemplo, para algunos GRDs asociados al área de maternidad, Lee et al. [34] utilizan mixturas finitas de distribuciones Gamma. Yau et al. [55] aplican las mixturas de distribuciones normales para la variable estancia hospitalaria en pacientes neo-natales.

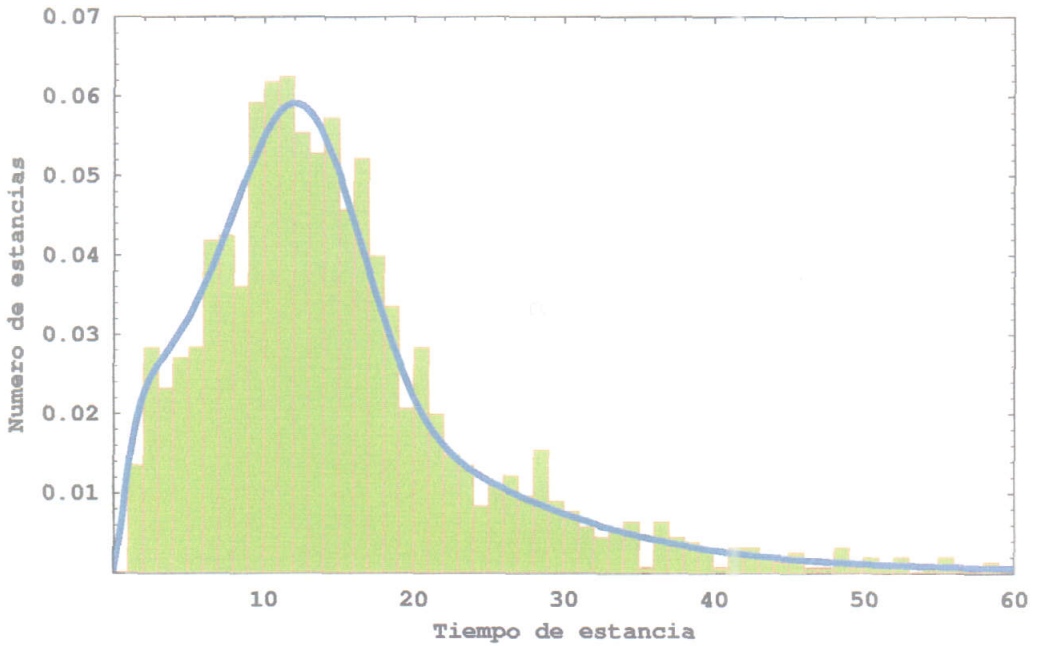
Como indican Lee et al. [34] es más aconsejable la utilización de las distribuciones mixturas finitas para tratar de modelizar la heterogeneidad de la variable estancia hospitalaria. Por otra parte, esta variable, caracterizada por su asimetría positiva, no tiene el mismo comportamiento en todos los GRDs. Los dos trabajos citados anteriormente constituyen una adecuada ilustración de tal afirmación. Por ello, como estrategia general para abordar el problema de la modelización de la estancia hospitalaria en cualquier GRDs, se puede usar la propuesta en esta memoria, es decir, la utilización de las mixturas mixtas de las distribuciones Lognormal, Gamma, y Weibull. Las tres distribuciones pueden ser válidas por sí solas en algunos GRDs, como se indica en Marazzi et al. [37], y sus distribuciones mixturas mixtas constituyen un modelo distribucional muy flexible, capaz de reflejar, de forma acusada, la asimetría positiva de la variable objeto de análisis en la gran mayoría de los GRDs. Es posible que para algunos de estos grupos la aplicación de esta estrategia no proporcione buenos resultados (por ejemplo, para grupos en los que la estancia hospitalaria sea pequeña). No obstante, es válida en aquellos GRDs con estancias altas, los cuales son los que centran un mayor interés tanto en estudios de costes y consumos de servicios hospitalarios como en estudios clínicos y/o farmacológicos.

Para ilustrar la adecuación de este modelo mixtura a la estancia hospitalaria, se presentan, a continuación, los resultados obtenidos de la aplicación del mismo a cinco muestras de datos de la variable estancia hospitalaria obtenidos del Hospital Universitario Virgen Macarena de Sevilla, pertenecientes a distintos GRDs y a periodos de tiempo durante los cuáles estos GRDs no variaron.

Cada una de dichas muestras se ha representado gráficamente mediante un histograma que aparece acompañado de la densidad del modelo mixtura estimado.

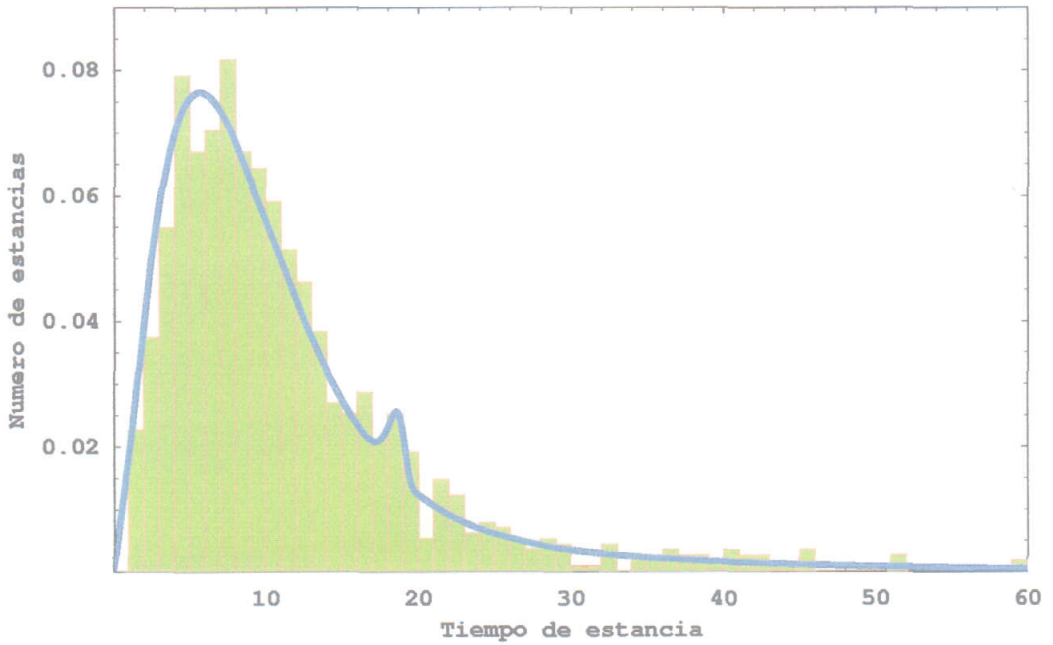
Junto a cada gráfico, se presentan para cada modelo individual, Lognormal, Gamma y Weibull, y para el modelo mixtura, los valores de los parámetros estimados. Asimismo, como medida de discrepancia ( $D$ ) entre la distribución estimada y la distribución empírica, se recogen la métrica uniforme para cada uno de los modelos. Puede observarse que esta medida de discrepancia es considerablemente más pequeña para el modelo mixtura que para cada uno de los modelos individuales. Tanto la programación utilizada para la aplicación del algoritmo, como la confección de los gráficos han sido realizadas con el software *Mathematica 4.1*.

<b>GRD</b>	14
<b>Descripción</b>	Trastornos específicos cerebrovasculares excepto Ictus transitorio
<b>Años</b>	1999, 2000, 2001, 2002
<b>Tamaño de muestra</b>	1579



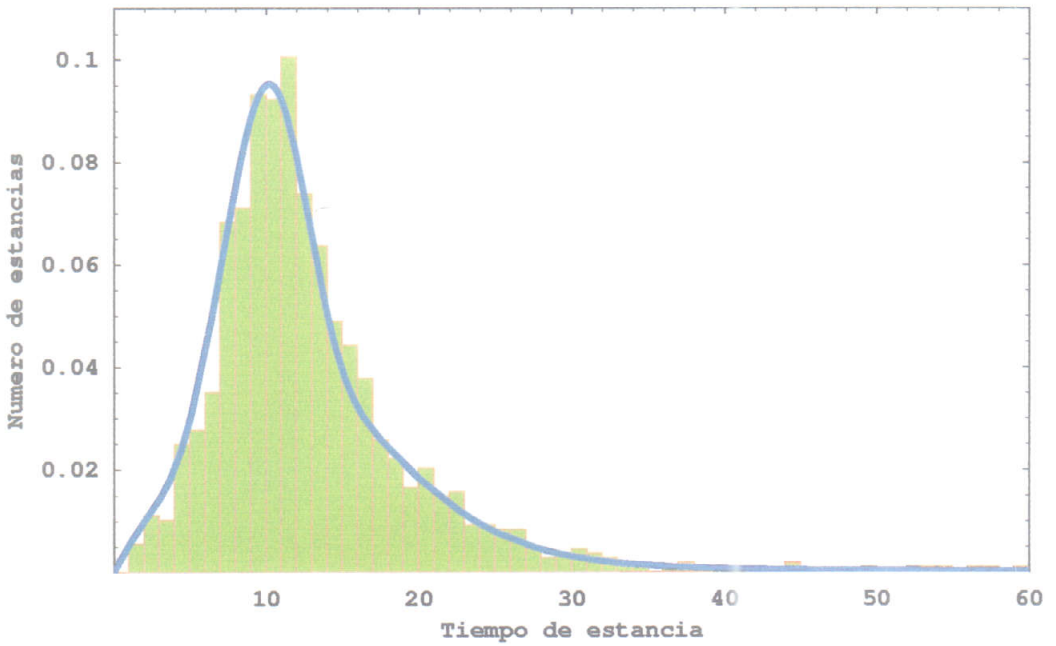
Modelo	Parámetros estimados								D
	$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$	
Lognormal	2.51	0.759							0.105
Gamma			8.35	1.					0.097
Weibull					1.64	17.4			0.097
Mixtura	2.29	1.08	4.96	4.23	3.33	13.5	0.343	0.244	0.038

GRD	88
Descripción	Enfermedad pulmonar obstructiva crónica
Años	1995, 1996, 1997, 1998
Tamaño de muestra	1156



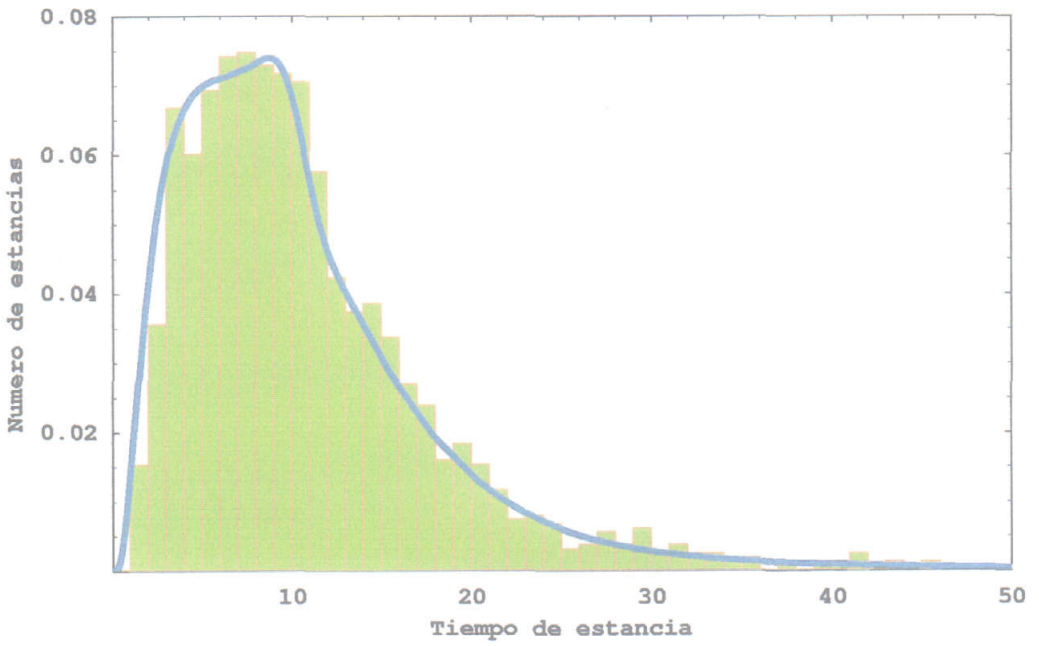
Modelo	Parámetros estimados								D
	$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$	
Lognormal	2.15	0.776							0.069
Gamma			6.18	1.					0.076
Weibull					1.6	12.2			0.085
Mixtura	3.41	0.487	2.61	3.52	35.6	18.7	0.086	0.899	0.047

<b>GRD</b>	122
<b>Descripción</b>	Transtornos circulatorios con infarto agudo de miocardio sin complicaciones cardiovasculares sin defunción
<b>Años</b>	1995, 1996, 1997, 1998
<b>Tamaño de muestra</b>	1086



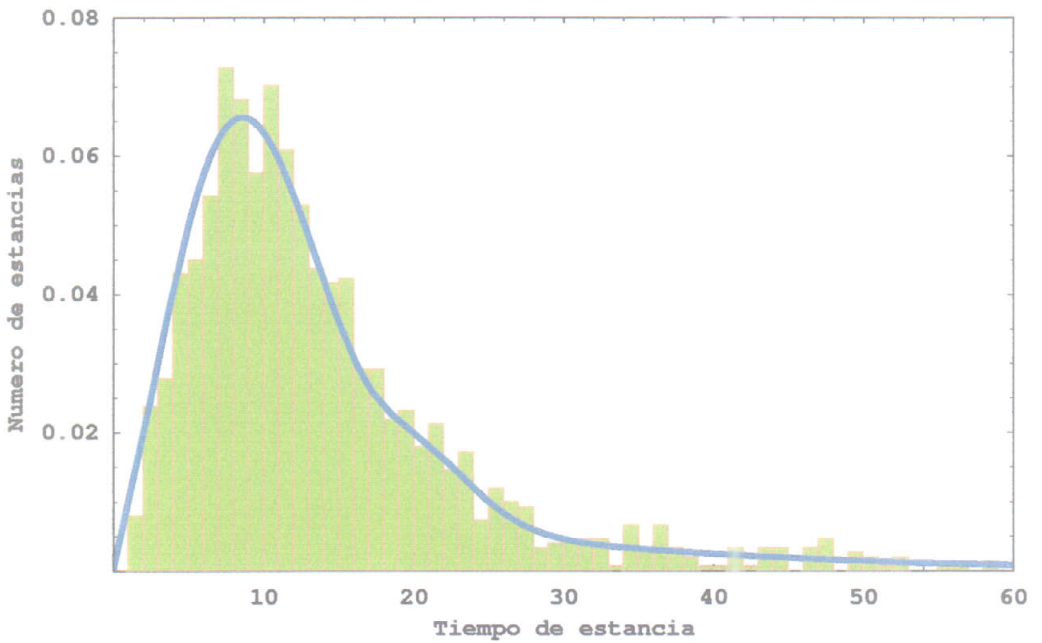
Modelo	Parámetros estimados								D
	$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$	
Lognormal	2.41	0.555							0.101
Gamma			3.89	1.					0.101
Weibull					2.22	14.3			0.124
Mixtura	2.27	1.01	6.22	2.39	4.22	10.6	0.156	0.443	0.052

GRD	127
Descripción	Insuficiencia cardíaca y shock
Años	1999, 2000, 2001, 2002
Tamaño de muestra	1639



Modelo	Parámetros estimados								D
	$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$	
Lognormal	2.13	0.706							0.076
Gamma			4.59	1.					0.058
Weibull					1.77	11.7			0.075
Mixtura	1.99	0.804	4.4	2.66	7.71	9.64	0.578	0.374	0.039

GRD	541
Descripción	Transtornos respiratorios excep. infecciones, bronquitis, asma
Años	1999, 2000, 2001, 2002
Tamaño de muestra	1528



Modelo	Parámetros estimados								D
	$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$	
Lognormal	2.4	0.725							0.068
Gamma			6.85	1.					0.083
Weibull					1.71	15.3			0.096
Mixtura	3.08	0.151	1.92	13.3	2.25	11.1	0.064	0.233	0.038

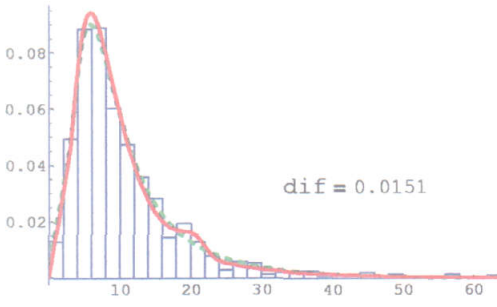
# Anexo A

## Resultados de la simulación: Gráficos

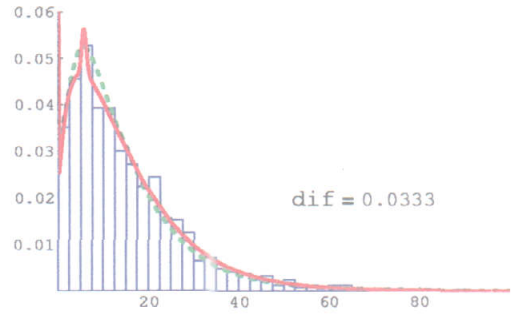
Se incluyen para cada una de las 128 simulaciones citadas en la sección 4.6 los resultados obtenidos. En concreto, para cada muestra generada se presentan:

- función de densidad de la distribución original (línea verde discontinua)
- histograma
- función de densidad de la distribución estimada a través del algoritmo propuesto en esta memoria (línea roja continua)
- diacrepancia entre las distribuciones según la métrica uniforme (dif).

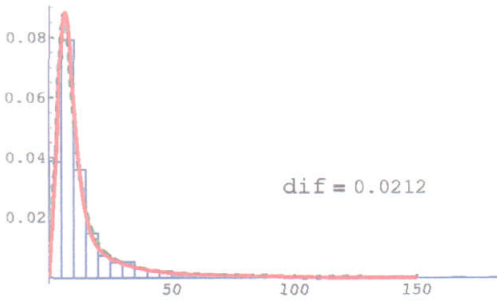
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
2.	0.5	2.	7.	1.	15.	0.554	0.363



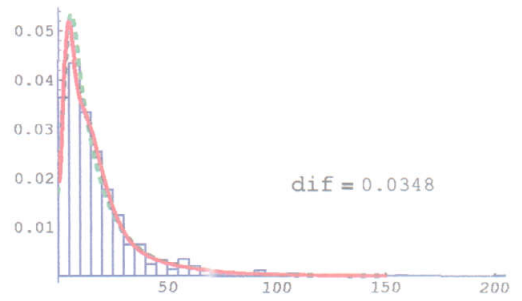
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
2.	0.5	2.	7.	1.	15.	0.048	0.537



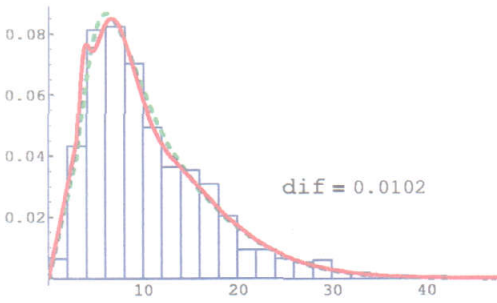
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
2.	0.5	2.	7.	1.	25.	0.588	0.136



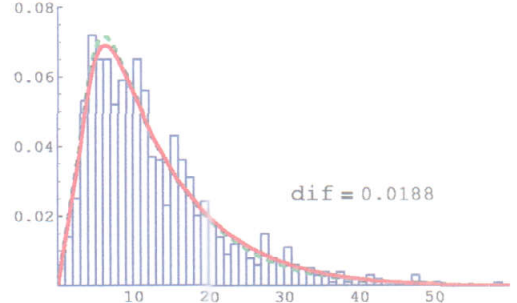
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
2.	0.5	2.	7.	1.	25.	0.145	0.432



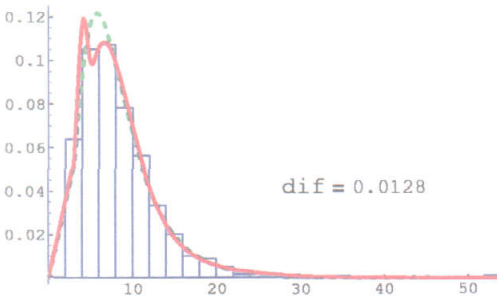
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
2.	0.5	2.	7.	2.	15.	0.538	0.042



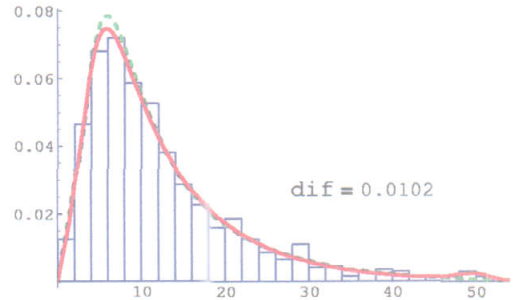
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
2.	0.5	2.	7.	2.	15.	0.293	0.564



$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
2.	0.5	2.	7.	2.	25.	0.996	0.001

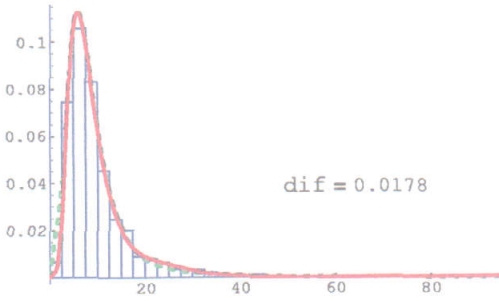


$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
2.	0.5	2.	7.	2.	25.	0.397	0.565

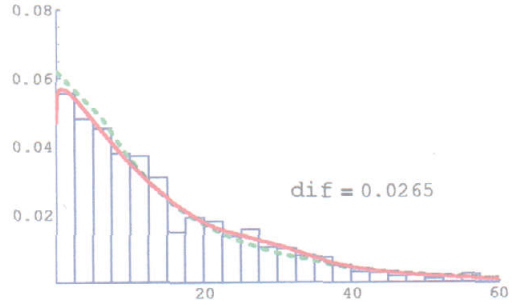




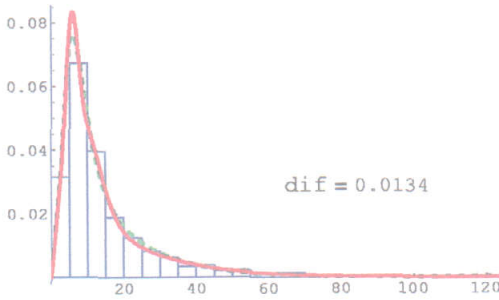
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
2.	0.5	2.	10.	1.	15.	0.896	0.102



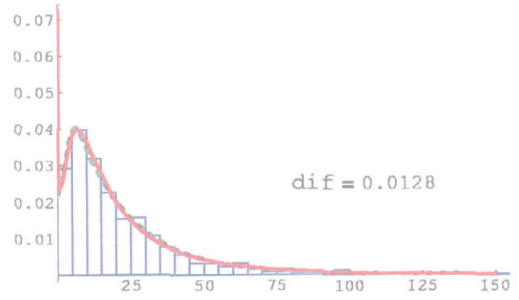
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
2.	0.5	2.	10.	1.	15.	0.057	0.014



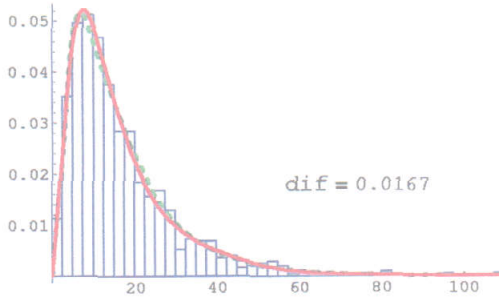
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
2.	0.5	2.	10.	1.	25.	0.488	0.461



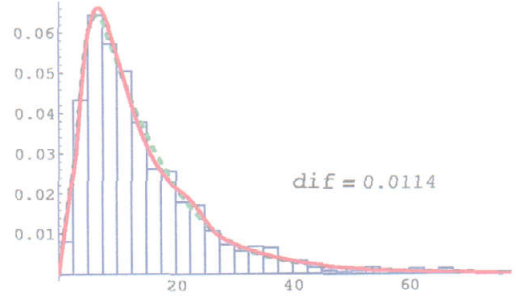
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
2.	0.5	2.	10.	1.	25.	0.093	0.365



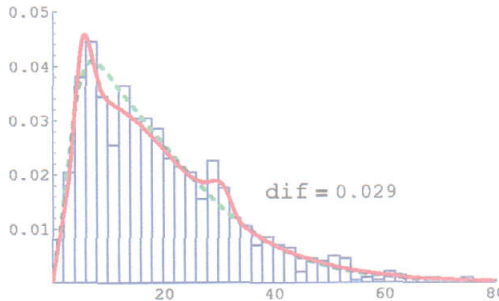
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
2.	0.5	2.	10.	2.	15.	0.178	0.632



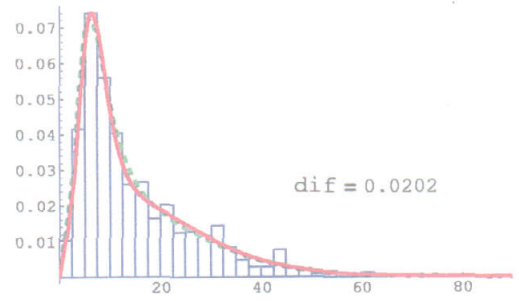
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
2.	0.5	2.	10.	2.	15.	0.303	0.369



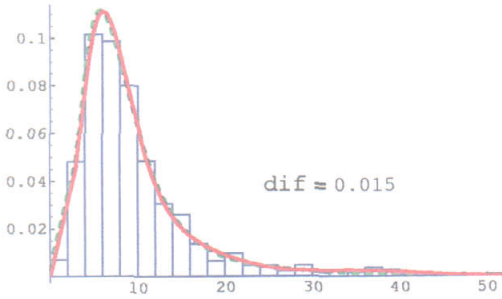
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
2.	0.5	2.	10.	2.	25.	0.1	0.78



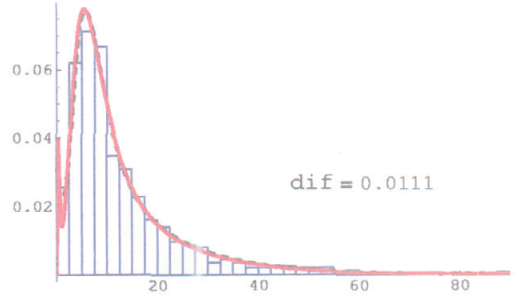
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
2.	0.5	2.	10.	2.	25.	0.485	0.268



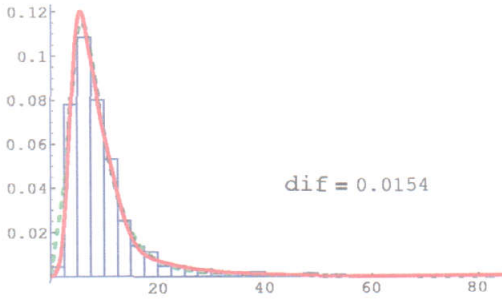
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
2.	0.5	3.	7.	1.	15.	0.874	0.058



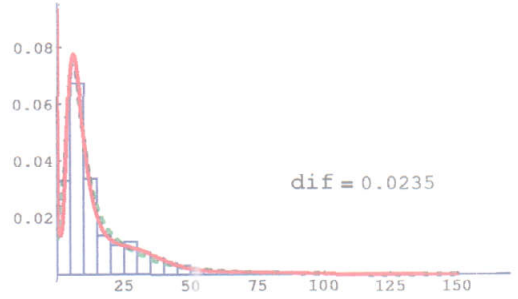
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
2.	0.5	3.	7.	1.	15.	0.472	0.187



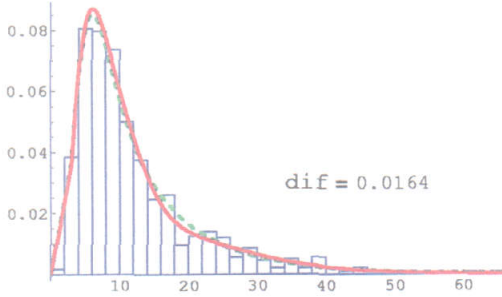
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
2.	0.5	3.	7.	1.	25.	0.925	0.047



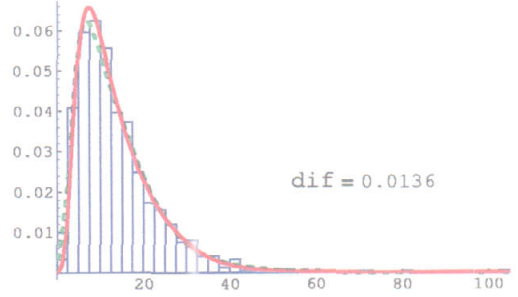
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
2.	0.5	3.	7.	1.	25.	0.498	0.186



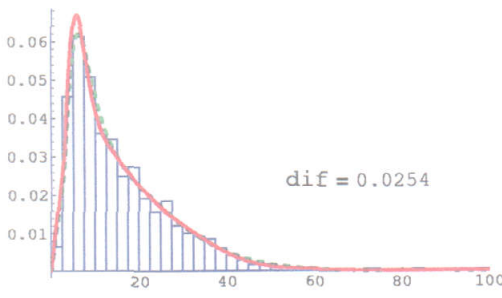
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
2.	0.5	3.	7.	2.	15.	0.593	0.249



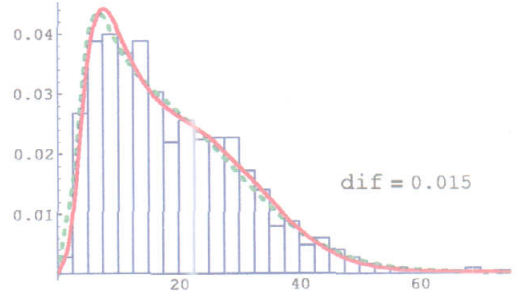
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
2.	0.5	3.	7.	2.	15.	0.292	0.266



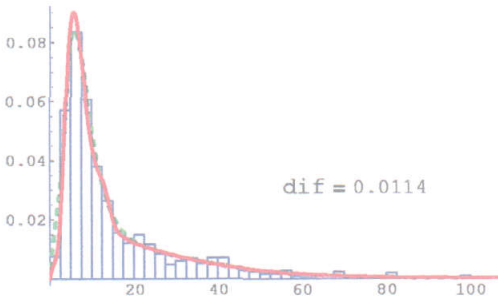
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
2.	0.5	3.	7.	2.	25.	0.405	0.385



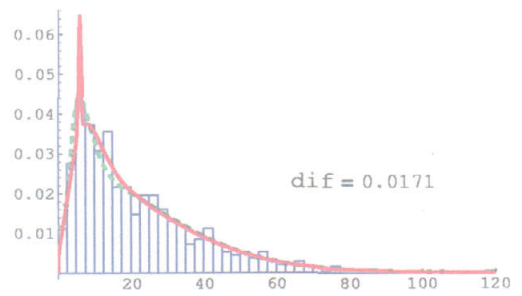
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
2.	0.5	3.	7.	2.	25.	0.232	0.224



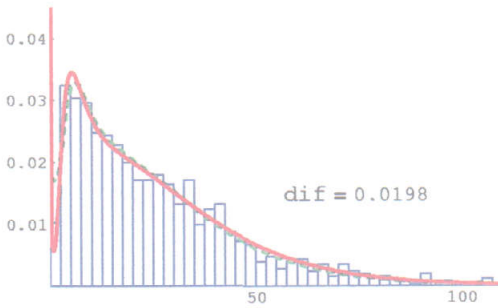
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
2.	0.5	3.	10.	1.	15.	0.645	0.3



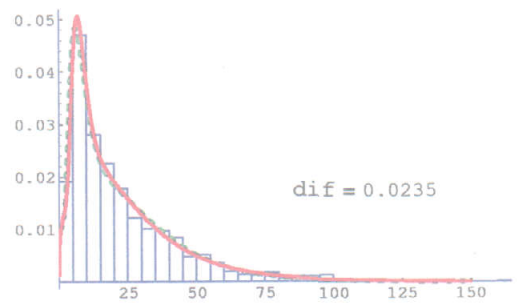
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
2.	0.5	3.	10.	1.	15.	0.252	0.572



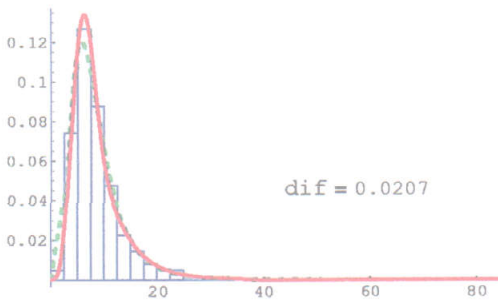
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
2.	0.5	3.	10.	1.	25.	0.121	0.44



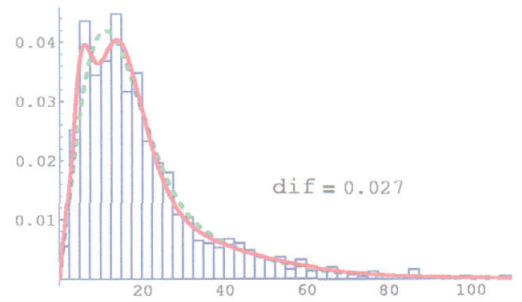
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
2.	0.5	3.	10.	1.	25.	0.295	0.446



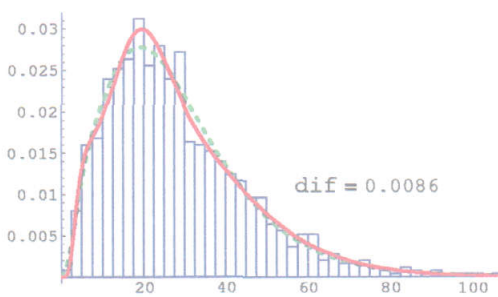
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
2.	0.5	3.	10.	2.	15.	0.976	0.011



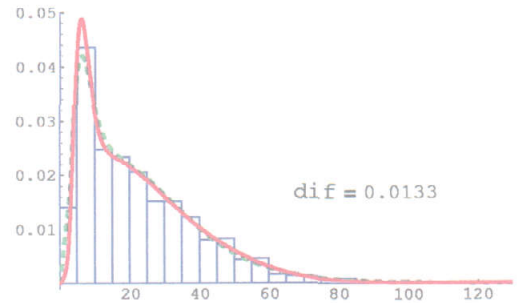
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
2.	0.5	3.	10.	2.	15.	0.023	0.409



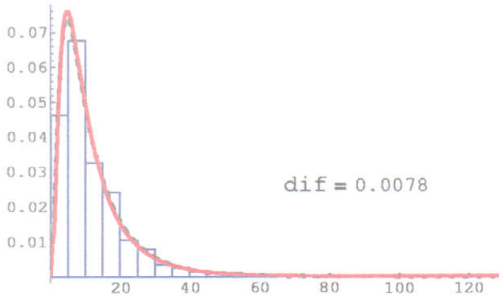
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
2.	0.5	3.	10.	2.	25.	0.027	0.792



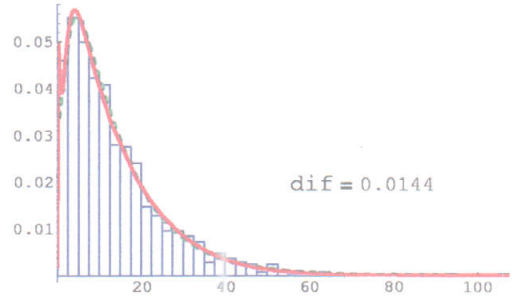
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
2.	0.5	3.	10.	2.	25.	0.276	0.595



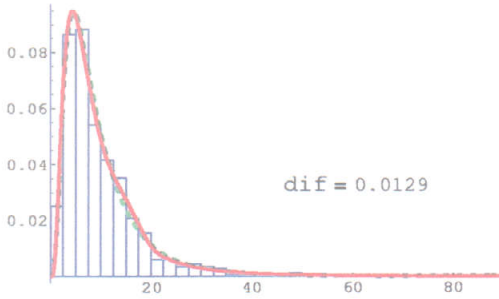
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
2.	0.7	2.	7.	1.	15.	0.512	0.341



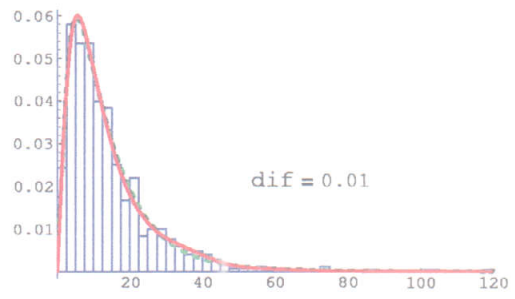
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
2.	0.7	2.	7.	1.	15.	0.132	0.402



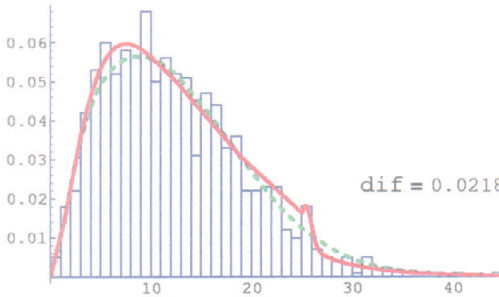
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
2.	0.7	2.	7.	1.	25.	0.921	0.064



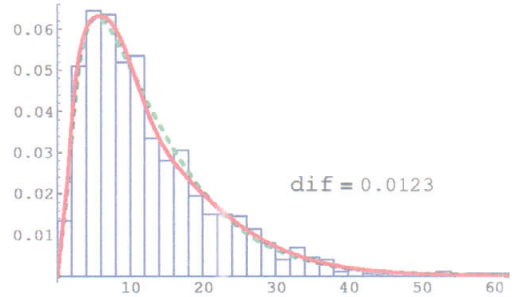
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
2.	0.7	2.	7.	1.	25.	0.25	0.616



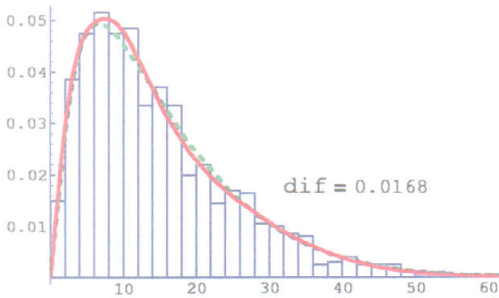
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
2.	0.7	2.	7.	2.	15.	0.158	0.031



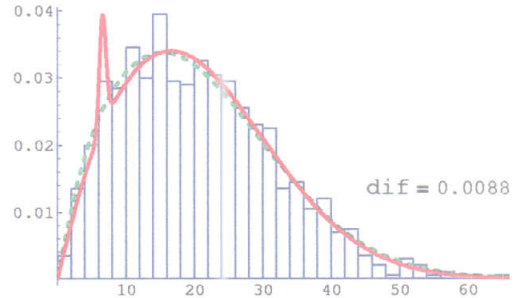
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
2.	0.7	2.	7.	2.	15.	0.29	0.601



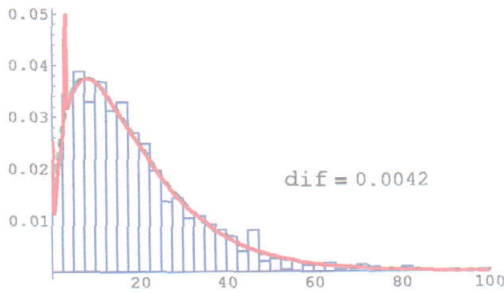
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
2.	0.7	2.	7.	2.	25.	0.078	0.746



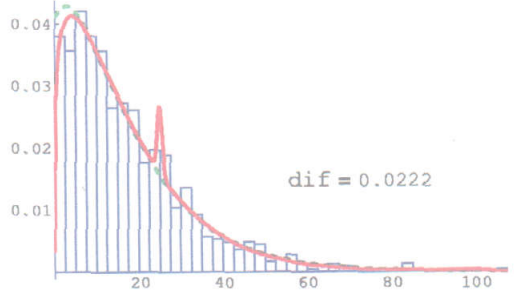
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
2.	0.7	2.	7.	2.	25.	0.011	0.153



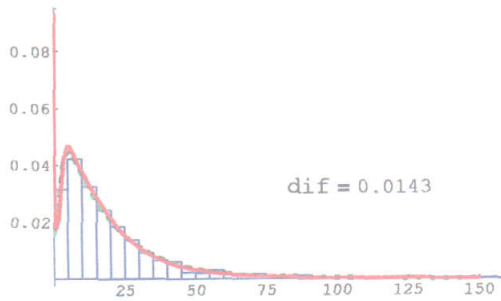
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
2.	0.7	2.	10.	1.	15.	0.024	0.758



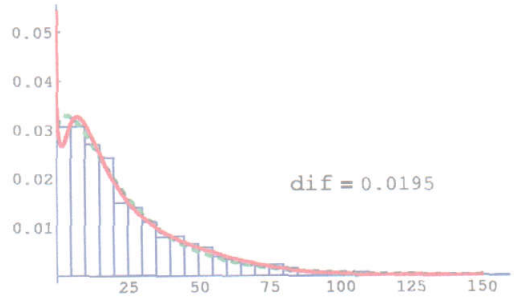
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
2.	0.7	2.	10.	1.	15.	0.009	0.373



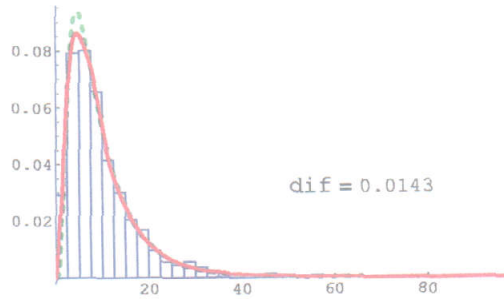
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
2.	0.7	2.	10.	1.	25.	0.2	0.413



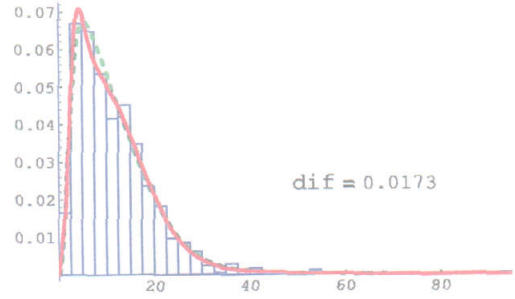
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
2.	0.7	2.	10.	1.	25.	0.006	0.223



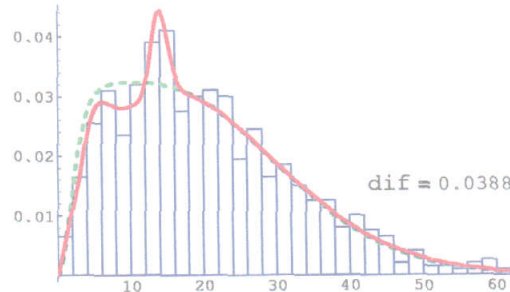
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
2.	0.7	2.	10.	2.	15.	0.921	0.012



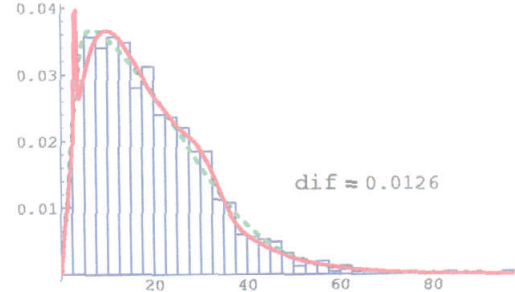
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
2.	0.7	2.	10.	2.	15.	0.48	0.05



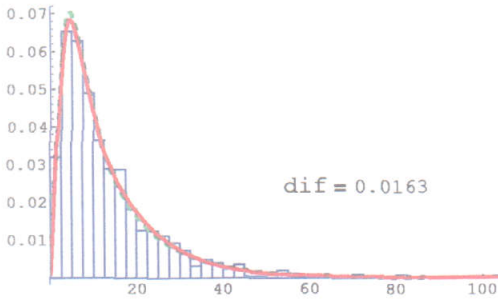
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
2.	0.7	2.	10.	2.	25.	0.157	0.125



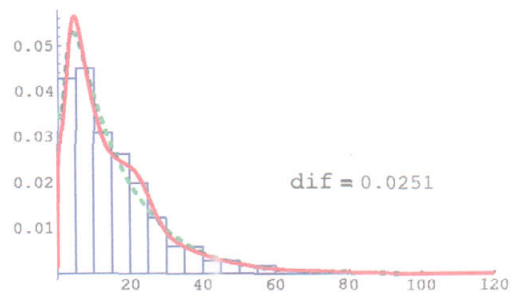
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
2.	0.7	2.	10.	2.	25.	0.172	0.402



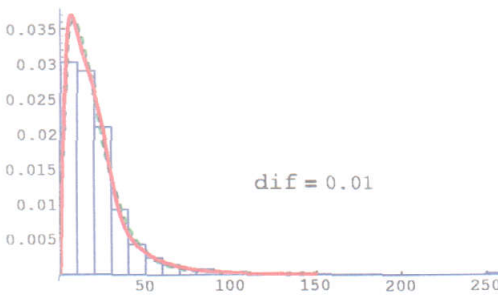
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
2.	0.7	3.	7.	1.	15.	0.527	0.147



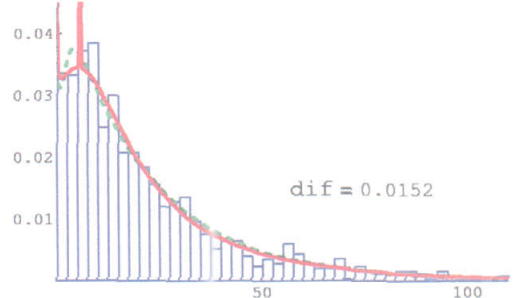
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
2.	0.7	3.	7.	1.	15.	0.235	0.241



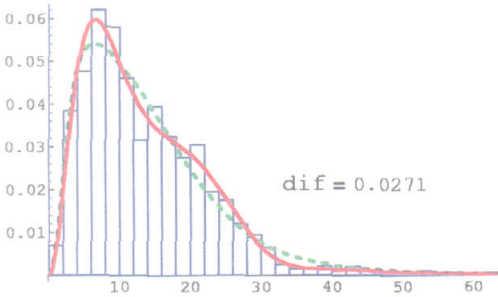
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
2.	0.7	3.	7.	1.	25.	0.162	0.553



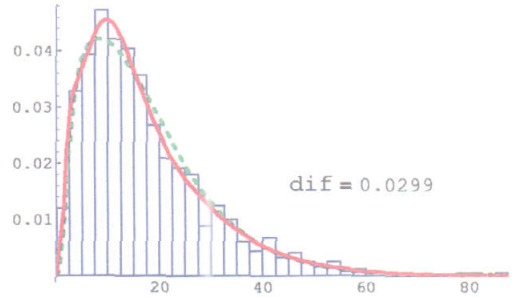
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
2.	0.7	3.	7.	1.	25.	0.093	0.112



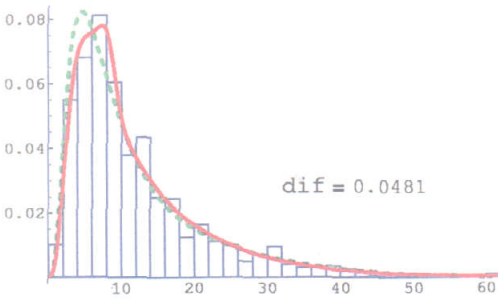
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
2.	0.7	3.	7.	2.	15.	0.319	0.262



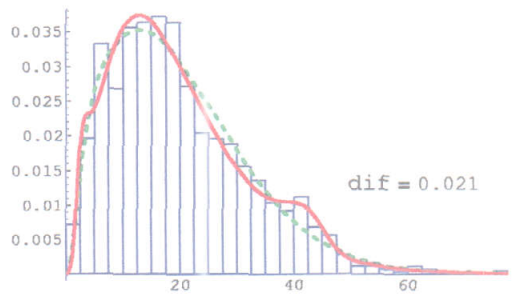
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
2.	0.7	3.	7.	2.	15.	0.231	0.65



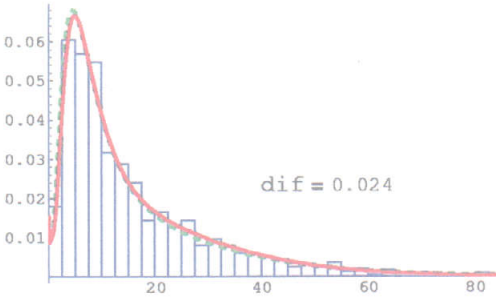
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
2.	0.7	3.	7.	2.	25.	0.806	0.025



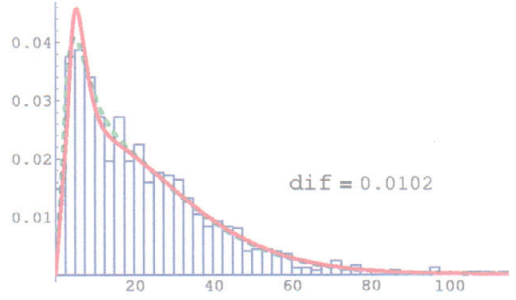
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
2.	0.7	3.	7.	2.	25.	0.122	0.521



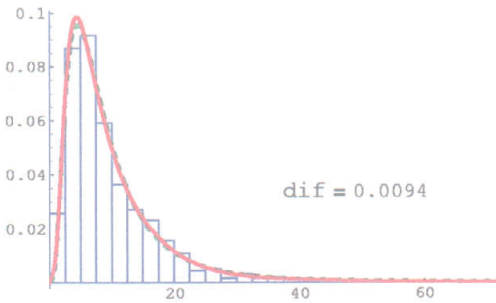
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
2.	0.7	3.	10.	1.	15.	0.597	0.251



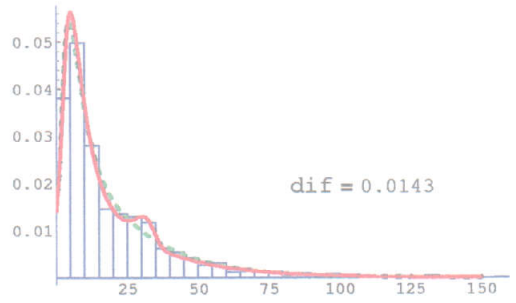
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
2.	0.7	3.	10.	1.	15.	0.316	0.567



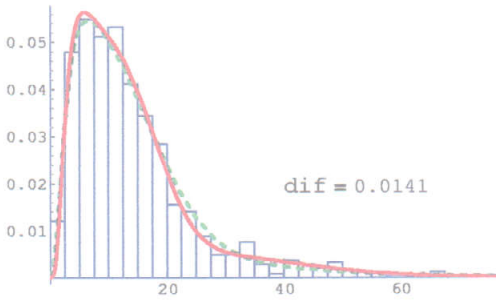
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
2.	0.7	3.	10.	1.	25.	0.959	0.019



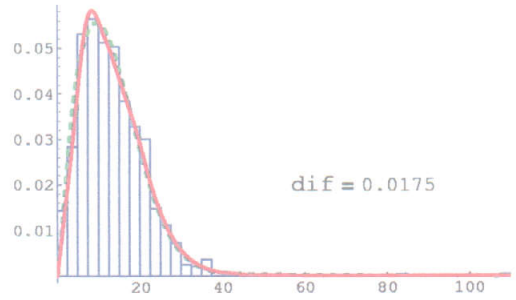
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
2.	0.7	3.	10.	1.	25.	0.386	0.151



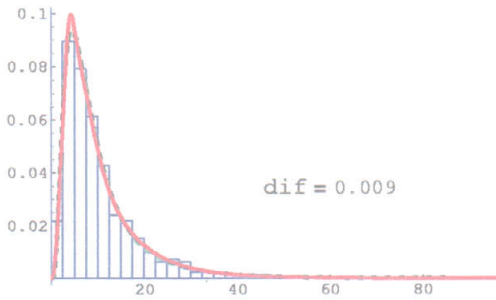
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
2.	0.7	3.	10.	2.	15.	0.299	0.129



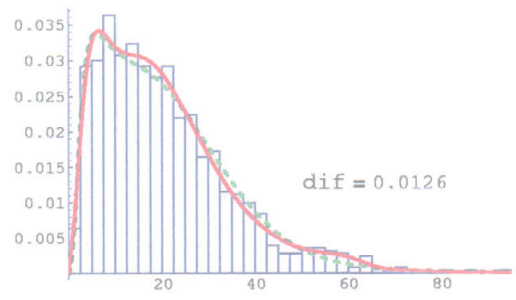
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
2.	0.7	3.	10.	2.	15.	0.113	0.025



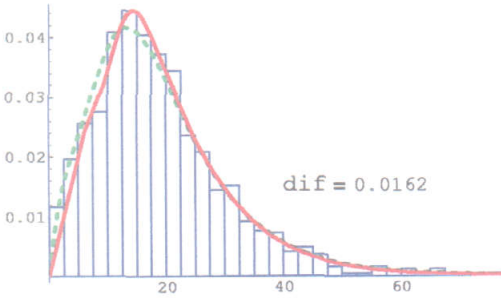
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
2.	0.7	3.	10.	2.	25.	0.942	0.026



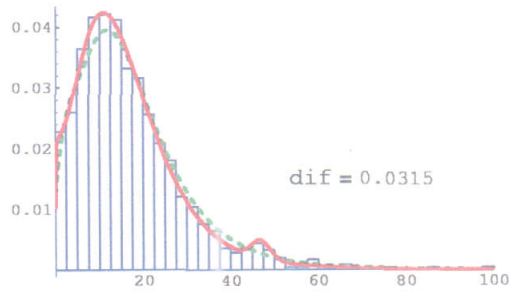
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
2.	0.7	3.	10.	2.	25.	0.238	0.202



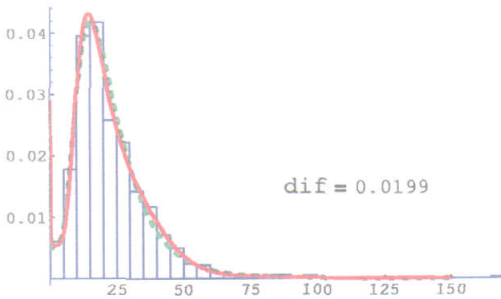
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
3.	0.5	2.	7.	1.	15.	0.567	0.421



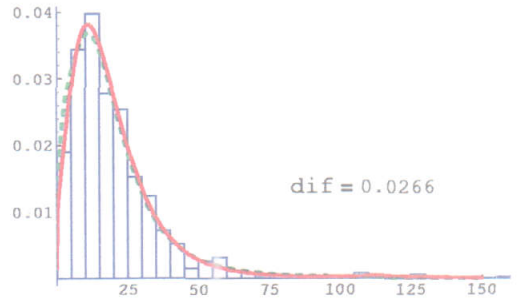
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
3.	0.5	2.	7.	1.	15.	0.394	0.417



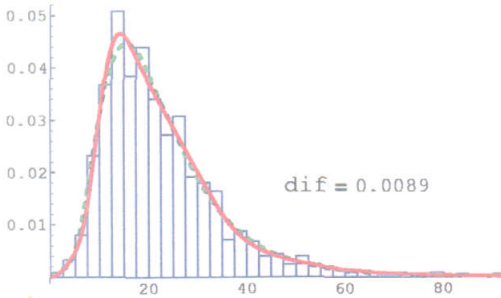
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
3.	0.5	2.	7.	1.	25.	0.844	0.038



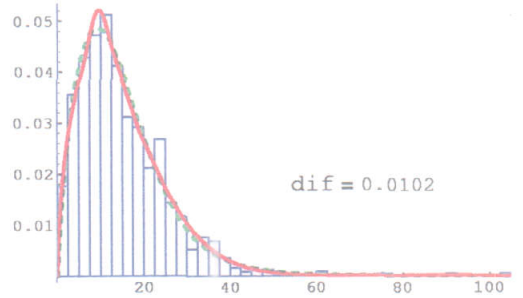
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
3.	0.5	2.	7.	1.	25.	0.285	0.403



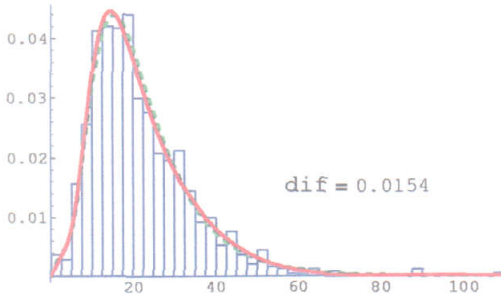
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
3.	0.5	2.	7.	2.	15.	0.958	0.037



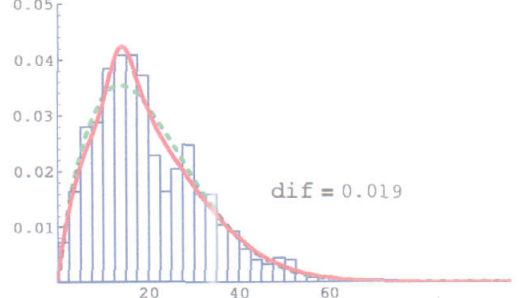
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
3.	0.5	2.	7.	2.	15.	0.131	0.628



$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
3.	0.5	2.	7.	2.	25.	0.889	0.1

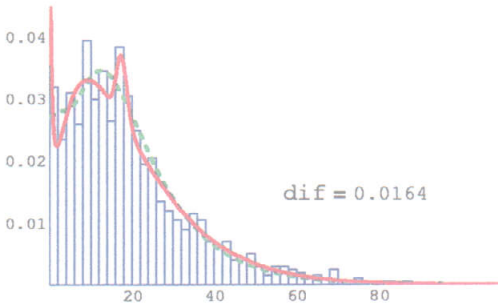


$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
3.	0.5	2.	7.	2.	25.	0.098	0.266

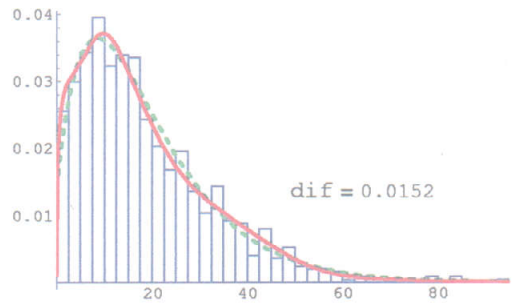




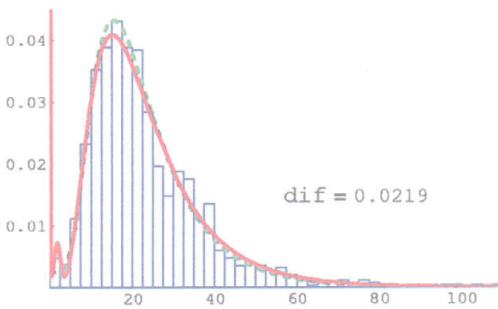
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
3.	0.5	2.	10.	1.	15.	0.331	0.26



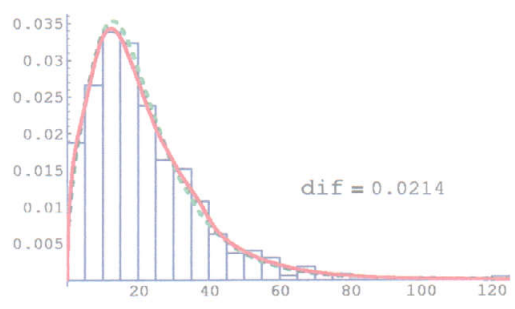
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
3.	0.5	2.	10.	1.	15.	0.011	0.777



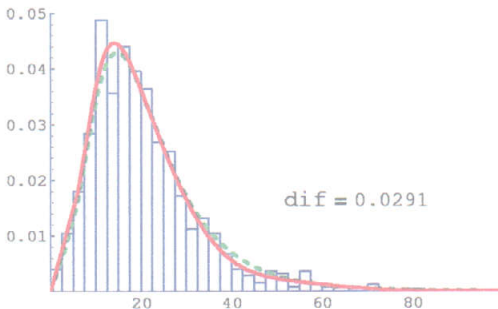
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
3.	0.5	2.	10.	1.	25.	0.905	0.038



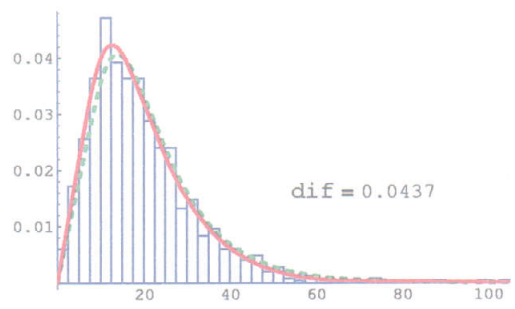
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
3.	0.5	2.	10.	1.	25.	0.247	0.608



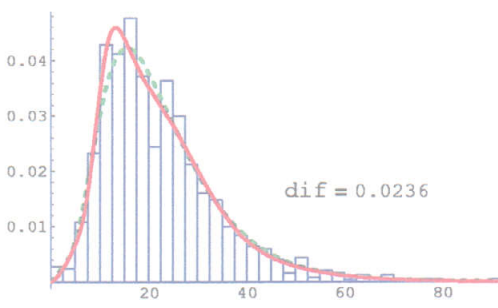
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
3.	0.5	2.	10.	2.	15.	0.708	0.204



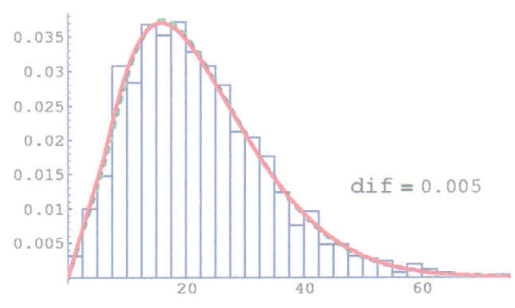
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
3.	0.5	2.	10.	2.	15.	0.482	0.424



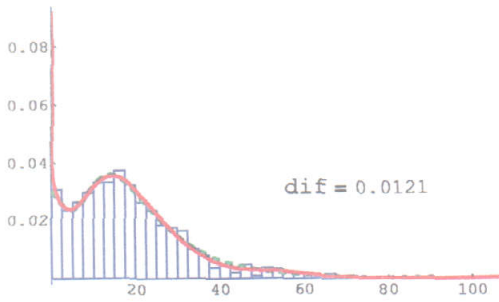
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
3.	0.5	2.	10.	2.	25.	0.757	0.069



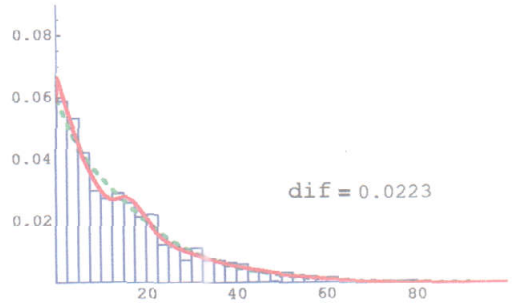
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
3.	0.5	2.	10.	2.	25.	0.329	0.093



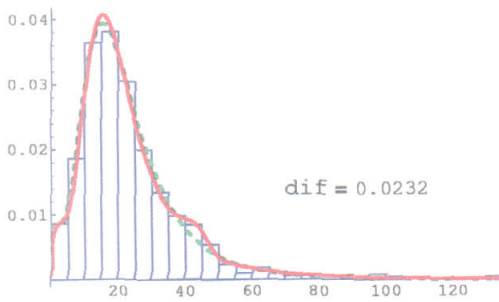
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
3.	0.5	3.	7.	1.	15.	0.52	0.032



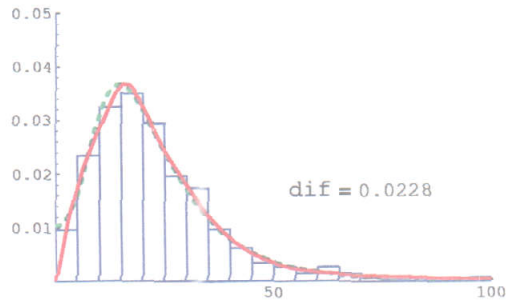
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
3.	0.5	3.	7.	1.	15.	0.057	0.051



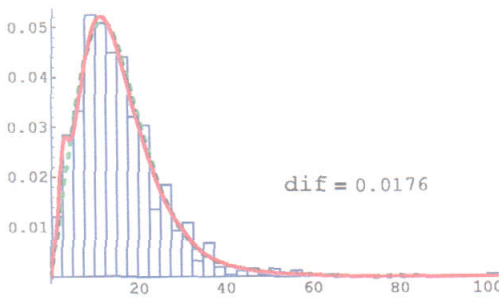
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
3.	0.5	3.	7.	1.	25.	0.656	0.14



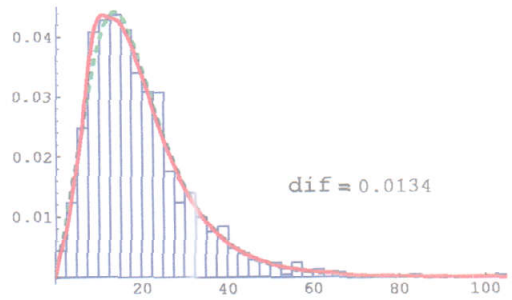
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
3.	0.5	3.	7.	1.	25.	0.379	0.359



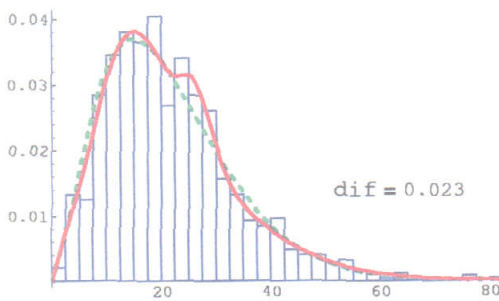
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
3.	0.5	3.	7.	2.	15.	0.128	0.168



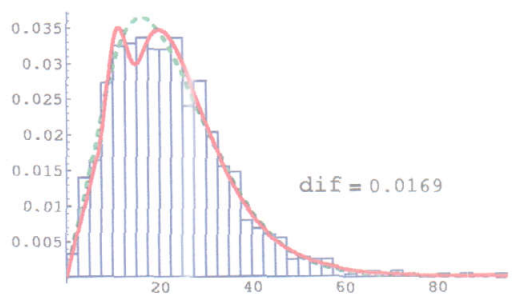
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
3.	0.5	3.	7.	2.	15.	0.295	0.426



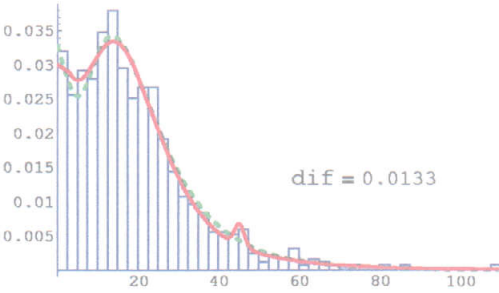
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
3.	0.5	3.	7.	2.	25.	0.006	0.663



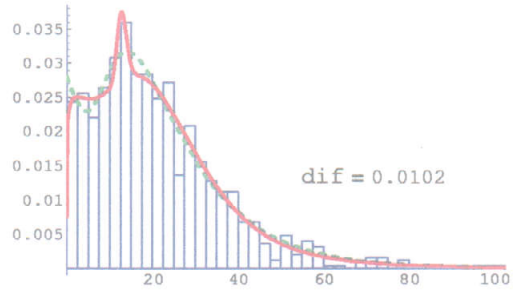
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
3.	0.5	3.	7.	2.	25.	0.111	0.291



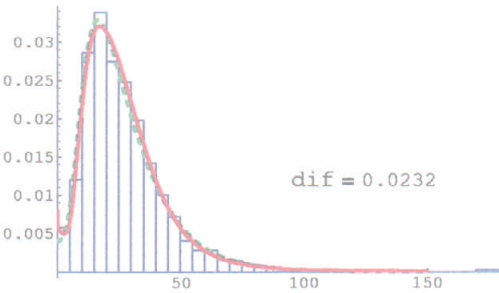
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
3.	0.5	3.	10.	1.	15.	0.46	0.048



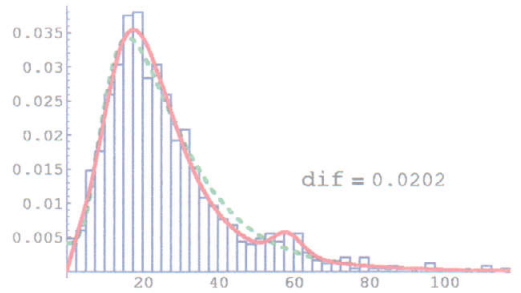
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
3.	0.5	3.	10.	1.	15.	0.337	0.243



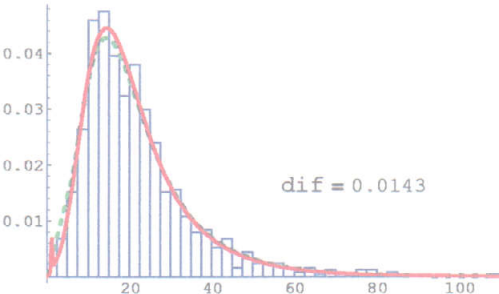
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
3.	0.5	3.	10.	1.	25.	0.391	0.51



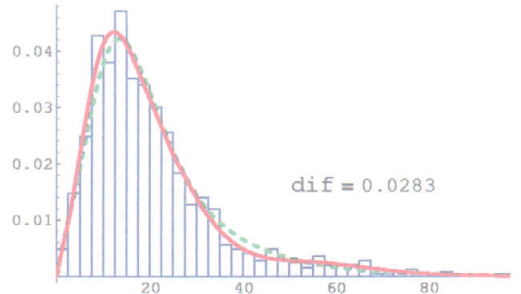
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
3.	0.5	3.	10.	1.	25.	0.466	0.431



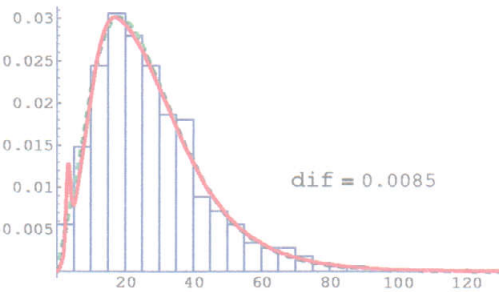
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
3.	0.5	3.	10.	2.	15.	0.581	0.165



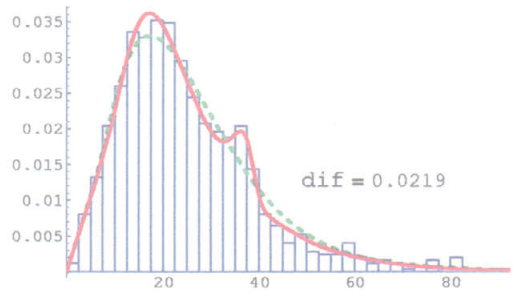
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
3.	0.5	3.	10.	2.	15.	0.341	0.26



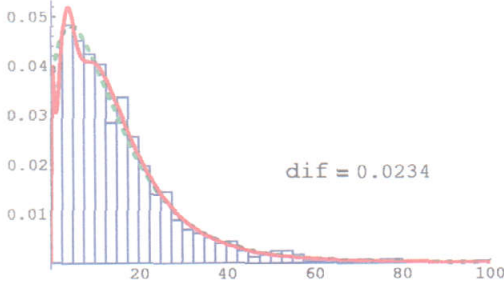
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
3.	0.5	3.	10.	2.	25.	0.08	0.651



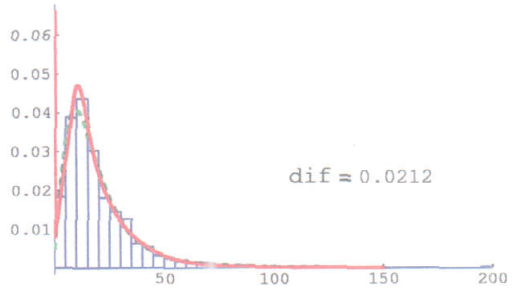
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
3.	0.5	3.	10.	2.	25.	0.197	0.411



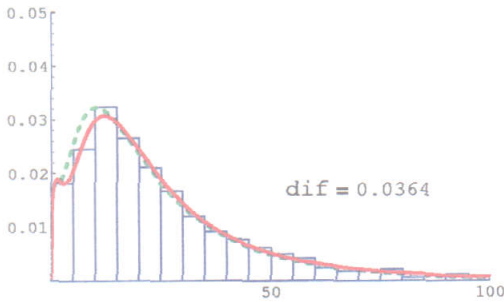
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
3.	0.7	2.	7.	1.	15.	0.021	0.437



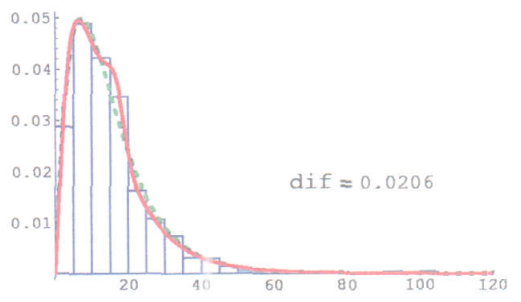
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
3.	0.7	2.	7.	1.	15.	0.498	0.427



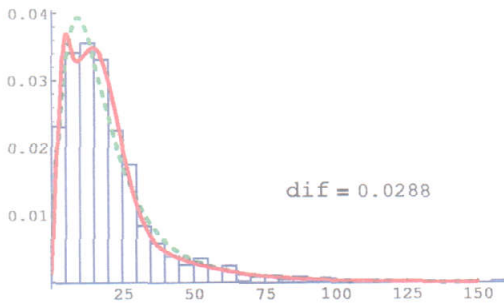
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
3.	0.7	2.	7.	1.	25.	0.497	0.071



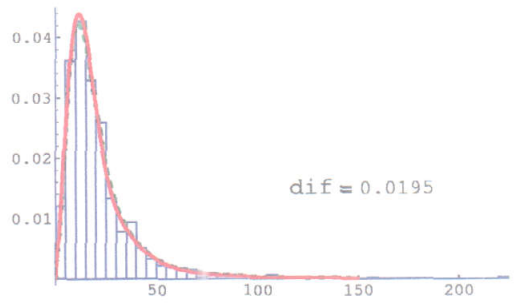
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
3.	0.7	2.	7.	1.	25.	0.072	0.896



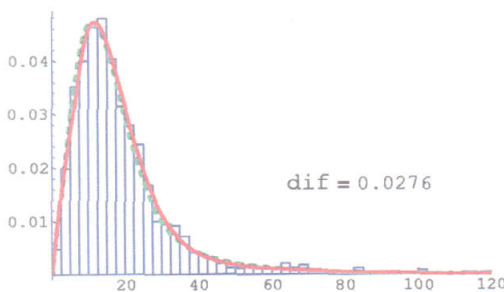
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
3.	0.7	2.	7.	1.	25.	0.395	0.419



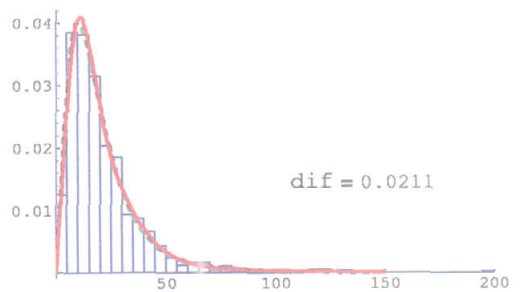
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
3.	0.7	2.	7.	2.	15.	0.646	0.101



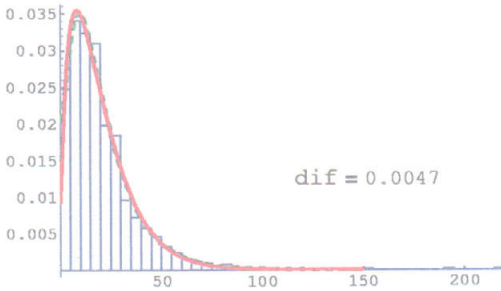
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
3.	0.7	2.	7.	2.	15.	0.429	0.049



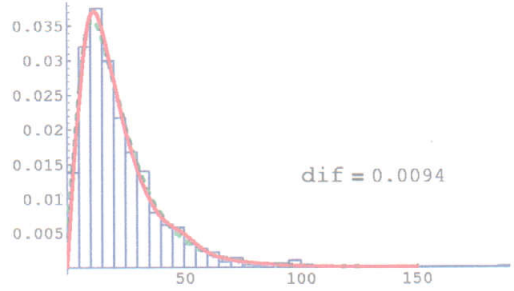
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
3.	0.7	2.	7.	2.	25.	0.557	0.405



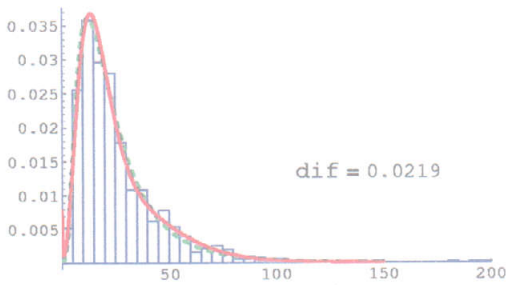
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
3.	0.7	2.	10.	1.	15.	0.376	0.317



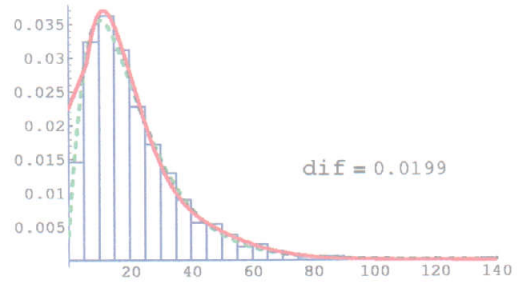
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
3.	0.7	2.	10.	1.	15.	0.483	0.421



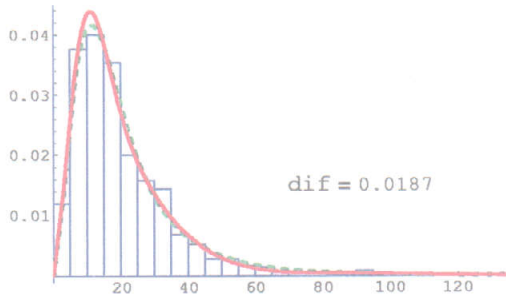
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
3.	0.7	2.	10.	1.	25.	0.982	0.01



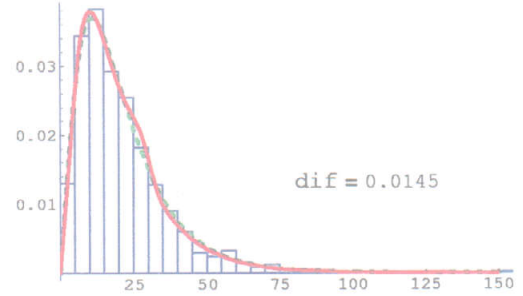
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
3.	0.7	2.	10.	1.	25.	0.185	0.724



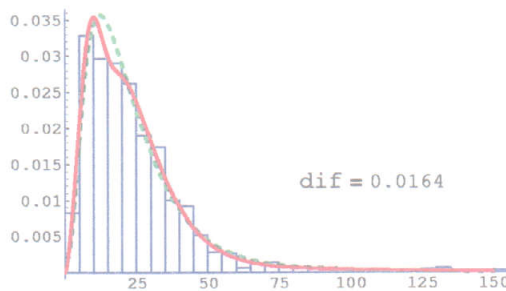
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
3.	0.7	2.	10.	2.	15.	0.546	0.193



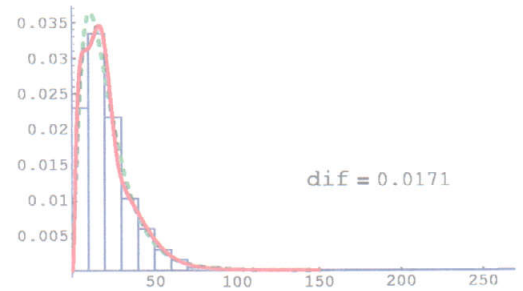
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
3.	0.7	2.	10.	2.	15.	0.287	0.686



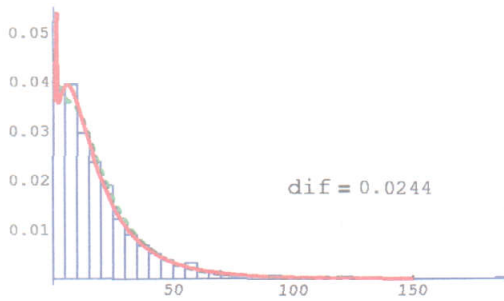
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
3.	0.7	2.	10.	2.	25.	0.726	0.172



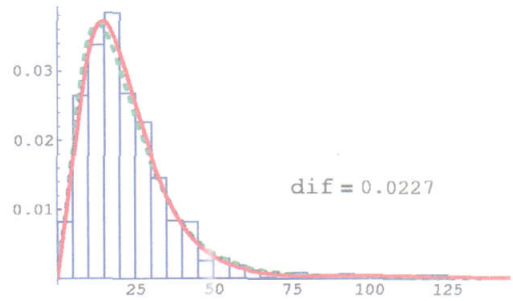
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
3.	0.7	2.	10.	2.	25.	0.29	0.707



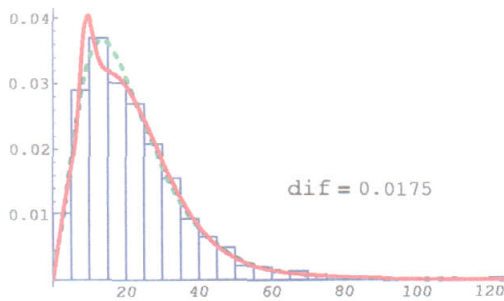
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
3.	0.7	3.	7.	1.	15.	0.32	0.062



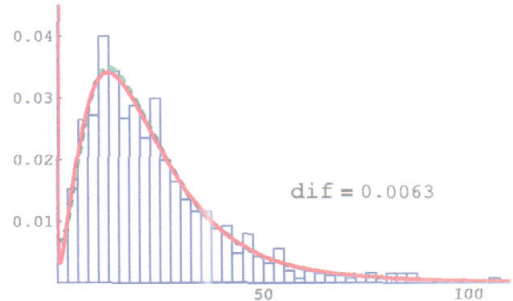
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
3.	0.7	3.	7.	1.	15.	0.385	0.543



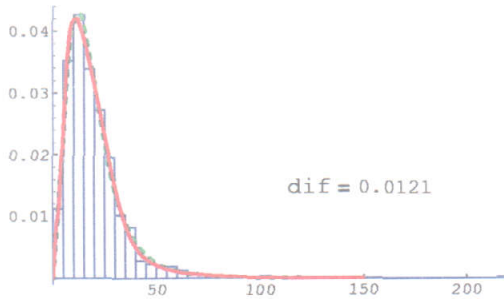
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
3.	0.7	3.	7.	1.	25.	0.18	0.723



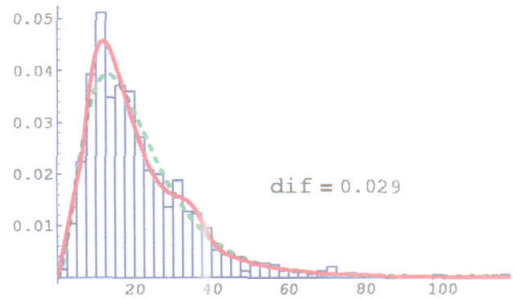
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
3.	0.7	3.	7.	1.	25.	0.469	0.36



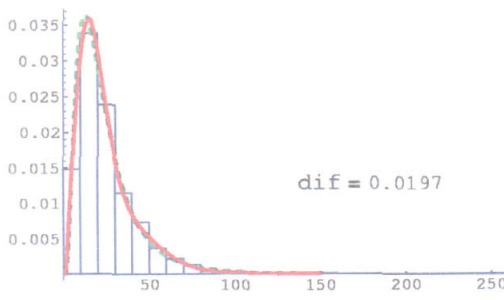
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
3.	0.7	3.	7.	2.	15.	0.192	0.514



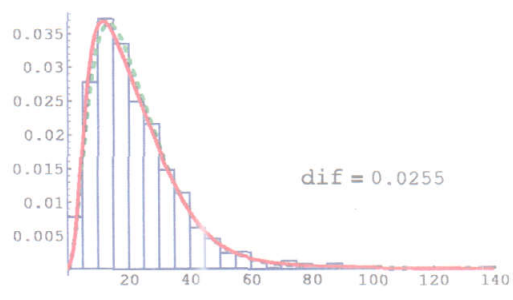
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
3.	0.7	3.	7.	2.	15.	0.229	0.687



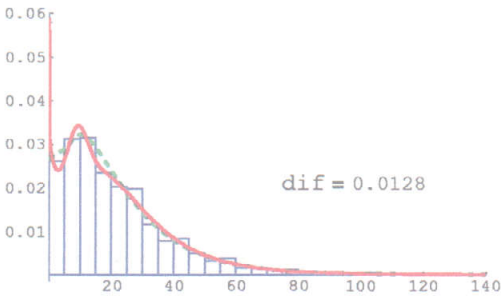
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
3.	0.7	3.	7.	2.	25.	0.865	0.124



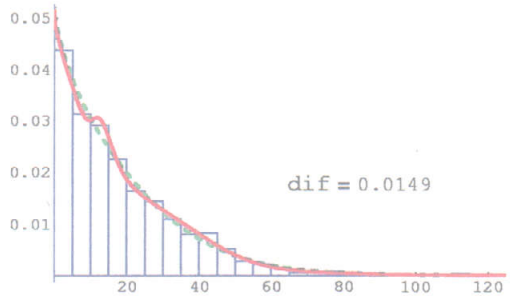
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
3.	0.7	3.	7.	2.	25.	0.343	0.466



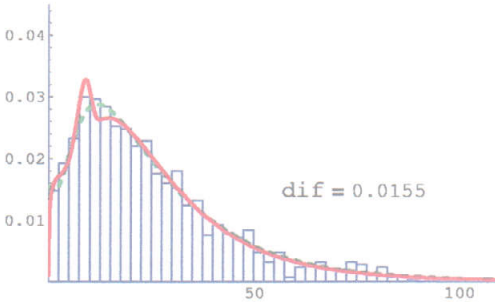
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
3.	0.7	3.	10.	1.	15.	0.455	0.143



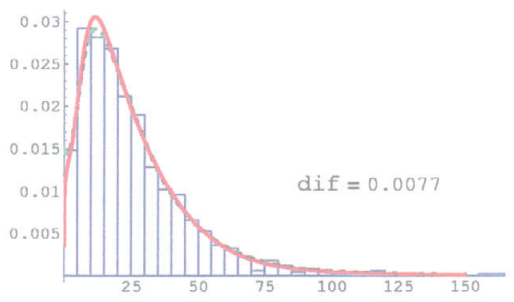
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
3.	0.7	3.	10.	1.	15.	0.003	0.239



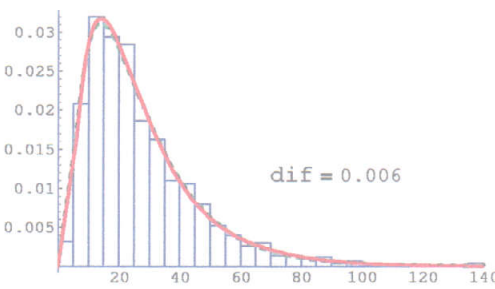
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
3.	0.7	3.	10.	1.	25.	0.406	0.209



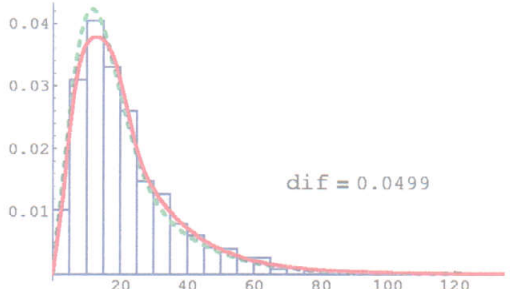
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
3.	0.7	3.	10.	1.	25.	0.437	0.192



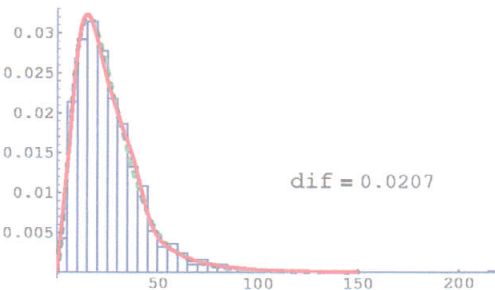
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
3.	0.7	3.	10.	2.	15.	0.604	0.394



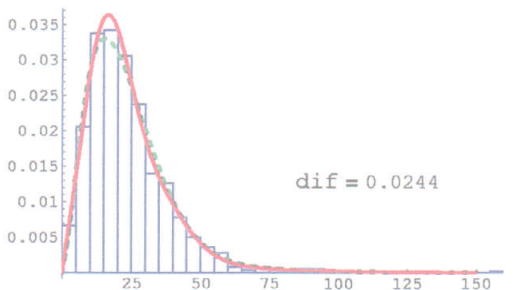
$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
3.	0.7	3.	10.	2.	15.	0.145	0.322



$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
3.	0.7	3.	10.	2.	25.	0.382	0.253



$\mu$	$\sigma$	a	b	c	d	$\pi_1$	$\pi_2$
3.	0.7	3.	10.	2.	25.	0.355	0.118



# Bibliografía

- [1] K.E. Ahmad. Identifiability of finite mixtures using a new transform. *Ann. Inst. Statist. Math.*, 40:261–265, 1988.
- [2] Y. Ahn, D.W. Nicholson, M.C. Wang, and P. Ni. Inverse method for identifying the underlying crack distribution in plates with random strengths. *Acta Mechanica*, 144:137–154, 2000.
- [3] F. A. Al-Baidhani and C. D. Sinclair. Comparison of methods of estimation of parameters of the Weibull distribution. *Commun. Statist. Simula.*, 16:373–384, 1987.
- [4] R. B. Ash. *Real Analysis and Probability*. Academic Press, New York, 1972.
- [5] L. J. Bain and C. E. Antle. Estimation of parameters in the Weibull distribution. *Technometrics*, 9:621–627, 1967.
- [6] O. Barndorff-Nielsen. Identifiability of mixtures of exponential families. *J.Math.Anal.Appl.*, 12:115–121, 1965.
- [7] J. N. Berrettoni. Practical applications of the Weibull distribution. *Industrial Quality Control*, 21:71–79, 1964.
- [8] I. J. Bienaymé. Mémoire sur la probabilité des résultats moyens des observations; démonstration directe de la règle de Laplace. *Mémoires de l'Académie de Sciences de l'Institut de France, Paris, Series étrangers*, 5:513–558, 1838.
- [9] S. Chanda. A note on the consistency and maxima of the roots of the likelihood equations. *Biometrika*, 41:56–61, 1954.
- [10] S. Chandra. On the mixtures of probability distributions. *Scand. J. Statist.*, 4:105–112, 1977.
- [11] A. C. Cohen. Maximum likelihood estimation in the Weibull distribution based on complete and on censored samples. *Technometrics*, 7:579–588, 1965.



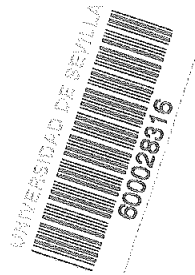
- [12] P.C. Cooley, L.E. Myers, and D.N. Hamill. A meta-analysis of estimates of the aids incubation distribution. *European Journal of Epidemiology*, 12(3):229–235, 1996.
- [13] A.A. Costa, C.J. de Oliveira, J.C.P. de Oliveira, and A.J.D. Sampaio. Microphysical observations of warm cumulus clouds in ceara, brazil. *Atmospheric Research*, 54:167–199, 2000.
- [14] H. Cramér. *Mathematical Methods of Statistics*. Princeton, New Jersey:Princeton University Press, 1946.
- [15] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm (with discussion). *J. Roy. Statist. Soc. Ser. B*, 39:1–38, 1977.
- [16] F. Galton. The geometric mean in vital and social statistics. *Proc. Royal Soc. London*, 29:365–367, 1879.
- [17] J. A. Greenwood and D. Durand. Aids for fitting the Gamma distribution by maximum likelihood. *Technometrics*, 2:55–56, 1960.
- [18] E. J. Gumbel. *Statistics of Extremes*. Columbia University Press, New York, 1958.
- [19] H. L. Harter and A. H. Moore. Maximum likelihood estimation of the parameters Gamma and Weibull populations from complete and from censored samples. *Technometrics*, 7:639–643, 1965.
- [20] J. Henna. Examples of identifiable mixture. *J. Japan Statist. Soc.*, 24:193–200, 1994.
- [21] D. W. Hosmer. A comparison of iterative maximum likelihood estimates of the parameters of a mixture of two normal distributions under three different types of samples. *Biometrics*, 29:761–770, 1973.
- [22] M. Jamshidian and R. I. Jennrich. Conjugate gradient acceleration of the em algorithm. *Journal of the American Statistical Association*, 88:221–228, 1993.
- [23] M. Jamshidian and R. I. Jennrich. Acceleration of the EM algorithm by using quasi-newton methods. *J. Roy. Statist. Soc. Ser. B*, 59:569–587, 1997.
- [24] L. W. Johnson and R. D. Riess. *Numerical Analysis*. MA Addison-Wesley, 1982.
- [25] N. L. Johnson and S. Kotz. *Continuous Univariate Distributions I*. John Wiley & Sons, Inc., 1970.
- [26] J. H. K. Kao. A graphical estimation of mixed Weibull parameters in life-testing electron tubes. *Technometrics*, 1:389–407, 1959.

- [27] D. Karlis and E. Xekalaki. Choosing initial values for the EM algorithm for finite mixtures. *Computational Statistics & Data Analysis*, 2002. Preprint.
- [28] J. Kiefer and J. Wolfowitz. Consistency of the maximum-likelihood estimation in the presence of infinitely many incidental parameters. *Ann. Math. Statist.*, 27:887–906, 1956.
- [29] A.N. Kolmogorov. Sulla determinazione empirica di una legge di distribuzione. *Giornale dell’Istituto Italiano degli Attuari*, 4:83–91, 1933.
- [30] G.M. Kondolf and A. Adhikari. Weibull vs. Lognormal distributions for fluvial gravels. *Journal of Sedimentary Research*, 70:456–460, 2000.
- [31] K. Lange. *Numerical Analysis for Statisticians*. Springer, New York, 1998.
- [32] D. Lansky and G. Casella. Improving the EM algorithm. *Computing Science and Statistics: Proc. 22nd Symp. Interface*, pages 420–424, 1977. Springer, New York.
- [33] P. S. Laplace. *Théorie Analytique des Probabilités*. (Supplement to the third edition), 1836.
- [34] A.H. Lee, A.S.K. NG, and K.K.W. Yau. Determinants of maternity length of stay: A Gamma mixture risk-adjusted model. *Health Care Management Science*, 4:249–255, 2001.
- [35] F.C. Leone, Y.H. Rutenberg, and C.W. Topp. Order statistics and estimators for the Weibull distributions. *Case Statistical Laboratory*, pages 1026+, 1960.
- [36] M.J. Lindstrom and D.M. Bates. Newton-raphson and EM algorithms for linear mixed effects models for repeated measures data. *J. Amer. Statist. Assoc.*, 83:1014–1022, 1988.
- [37] A. Marazzi, F. Paccaud, C. Ruffieux, and C. Beguin. Fitting the distributions of length of stay by parametric models. *Medical-Care*, 36(6):915–927, 1998.
- [38] D.E.F. McKeegan. Spontaneous and odour evoked activity in single avian olfactory bulb neurones. *Brain Research*, 929(1):48–58, 2002.
- [39] G. J. McLachlan and T. Krishnan. *The EM algorithm and extensions*. Wiley & Sons, New York, 1997.
- [40] G. J. McLachlan and D. Peel. *Finite mixture models*. Wiley & Sons, New York, 2001.
- [41] I. Meilijson. A fast improvement to the EM algorithm on its own terms. *J. R. Statist. Soc. B*, 51:127–138, 1989.

- [42] Z. L. Meng and D. B. Rubin. Using EM to obtain asymptotic variance-covariance matrices: the SEM algorithm. *J. Amer. Statist. Assoc.*, 86:899–909, 1991.
- [43] I. Miller and J.E. Freund. *Probability and Statistics for Engineers*. Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1965.
- [44] I.A. Parvez and A. Ram. Probabilistic assessment of earthquake hazards in the indian subcontinent. *Pure And Applied Geophysics*, 154:23–40, 1999.
- [45] B.C. Peters and H.F. Walker. An iterative procedure for obtaining maximum-likelihood estimates of the parameters for a mixture of normal distributions. *SIAM J. Appl. Math.*, 35:362–378, 1978.
- [46] R. A. Redner and H. F. Walker. Mixture densities, maximum likelihood and the EM algorithm. *SIAM Review*, 26:195–239, 1984.
- [47] P. R. Rider. Estimating the parameters of mixed poisson, binomial, and Weibull distributions by the method of moments. *Bulletin de l'Institut International de Statistique*, 39:225–232, 1961.
- [48] A. E. Sarhan and B.G. Greenberg. *Contributions to Order Statistics*. John Wiley and Sons, New York, 1962.
- [49] R.S. Singer, N.J. MacLachlan, and T.E. Carpenter. Maximal predicted duration of viremia in bluetongue virus-infected cattle. *Journal of Veterinary Diagnostic Investigation*, 13:43–49, 2001.
- [50] D.B. Stephenson, K.R. Kumar, F.J. Doblas-Reyes, J.F. Royer, E. Chauvin, and S. Pezzulli. Extreme daily rainfall events and their impact on ensemble forecasts of the indian monsoon. *Monthly Weather Review*, 70:1954–1966, 1999.
- [51] H. Teicher. Identifiability of finite mixtures. *Ann. Math. Statist.*, 34:1265–1269, 1963.
- [52] D. M. Titterington, A. F. M. Smith, and U. E. Makov. *Statistical analysis of finite mixture distributions*. John Wiley & Sons, New York, 1985.
- [53] R.C. Whiting and M.H. Golden. Variation among escherichia coli o157 : H7 strains relative to their growth, survival, thermal inactivation, and toxin production in broth. *International Journal of Food Microbiology*, 75:127–133, 2002.
- [54] S.J. Yakowitz and J.D. Spragins. On the identifiability of finite mixtures. *Ann. Math. Statist.*, 39:209–214, 1968.

- [55] K.K.W. Yau, A.S.K. NG, and A.H. Lee. Finite mixture regression model with random effects: application to neonatal hospital length of stay. *Computational Statistics and Data Analysis*, 41:359–366, 2003.
- [56] V.M. Zolotarev. Probability metrics. *Theory Probab. Appl.*, 28:278–302, 1983.

# UNIVERSIDAD DE SEVILLA



Reunido el Tribunal integrado por los abajo firmantes,  
en el día de la fecha, para juzgar la Tesis Doctoral de  
D. NIEVES ALLENCA MARTINEZ,  
titulada MIXTURA DE DISTRIBUCIONES: MODELIZACION  
DE EXPERIENCIAS COMO ASIMETRIA EN LOS  
DATOS  
acorde otorgarle la calificación de SOBRESALIENTE CUM LAUDE  
POR UNANIMIDAD

Sevilla, 24 de Abril 2003

El Vocal,

  
EL PRESIDENTE

El Vocal

  
El Secretario

El Vocal

  
El Doctorado



