
MODELOS PREDICTIVOS BAYESIANOS:
ESTIMACIÓN DE PARÁMETROS EN
DISTRIBUCIONES DE TIPO MoTBF A PARTIR
DE GRANDES VOLÚMENES DE DATOS

TRABAJO FIN DE GRADO

Autora:

Marina Ortiz Montes

Tutor:

Antonio Salmerón Cerdán

GRADO EN MATEMÁTICAS



JULIO, 2021
Universidad de Almería

Índice general

1	Introducción	1
1.1.	Objetivos y Reseña histórica	1
1.2.	Preliminares	2
2	Distribuciones de tipo MoTBF	3
2.1.	Definiciones básicas	3
	Mixturas de Exponenciales Truncadas, 3.— Mixturas de Polinomios, 4.— Mixturas de Funciones Base Truncadas, 5.	
2.2.	Propiedades	6
	Distribución de tipo MTE univariante, 10.— Distribución de tipo MOP univariante, 14.	
2.3.	Ejemplos	17
	Distribución de tipo MTE univariante, 18.— Distribución de tipo MOP univariante, 19.	
3	Métodos de estimación de parámetros a partir de los datos	25
3.1.	Método de máxima verosimilitud	26
	Distribución de tipo MTE univariante, 27.— Distribución de tipo MOP univariante, 29.	
3.2.	Método de <i>Fisher Scoring</i>	30
	Conceptos y resultados previos, 30.— Algoritmo de <i>Fisher Scoring</i> , 33.— Distribución de tipo MOP univariante, 34.	
3.3.	Método de mínimos cuadrados	38
	Distribución de tipo MTE univariante, 40.— Distribución de tipo MOP univariante, 40.	
4	Análisis experimental	41
4.1.	Distribución de tipo MOP univariante	41
	Experimento 1. Algoritmo de <i>Fisher Scoring</i> , 42.— Experimento 2. Algoritmo de <i>Fisher Scoring</i> , 43.	
5	Conclusiones	47
	Bibliografía	49
A	Apéndice	51
A.1.	Implementación en R del método de <i>Fisher Scoring</i> para distribuciones univariantes de tipo MOP de grado 2	51
	Algoritmo de <i>Fisher Scoring</i> , 52.— Algoritmo de <i>Fisher Scoring</i> optimizado, 53.	

Abstract in English

In this work, a study is carried out on mixtures of truncated basis functions distributions (MoTBF). More specifically, we focus on two of their particular cases: mixtures of truncated exponentials (MTEs) and mixtures of polynomials (MOPs). The main goal of this work is, on the one hand, to study the basic properties of MTE and MOP distributions and, on the other hand, to develop methods that solve the problem of estimating parameters in these distributions from large volumes of data.

We begin by motivating the appearance of MoTBF models through a brief historical overview in chapter 1. In this chapter, we also introduce the notation required for the subsequent development of the work, as well as a general introduction to hybrid Bayesian networks, which constitute the main framework in which MoTBF models acquire a fundamental value.

Then, in chapter 2, the MoTBF model is already introduced; first, through its two particular cases, namely: the MTE model and the MOP model, and next, in a generic way. To do this, we present, in each case, the definition of both univariate and conditional distributions. Later, a study is carried out on the basic properties of univariate MTE and MOP distributions. We conclude this chapter by illustrating with examples the properties previously obtained.

In chapter 3, the second issue that is the purpose of this work is addressed, that is, the study of the parameter estimation problem in MoTBF distributions from large volumes of data. For it, we consider as possible solutions the (exact) maximum likelihood method, the Fisher Scoring method, which is a numerical method to roughly determine the maximum likelihood estimator and, finally, the least squares method. For each of these methods, we examine its application to univariate MTE and MOP distributions, except in the case of the Fisher Scoring method, for which we leave aside the study for MTEs.

Afterwards, in chapter 4, we carry out an experimental analysis of the previously developed methods, focusing mainly on the Fisher Scoring algorithm, in order to study its behavior regarding the parameter estimation in univariate MOP distributions from large volumes of data. For this analysis, the R software is used, through the implementation of the mentioned algorithm in this tool, which is included in appendix A.

To finish, in chapter 5 the conclusions of the present work are exposed.

Resumen en español

En este trabajo se realiza un estudio sobre las distribuciones de tipo mixtura de funciones base truncadas (MoTBF). Más concretamente, nos centramos en dos de sus casos particulares: las mixturas de exponenciales truncadas (MTEs) y las mixturas de polinomios (MOPs). El objetivo principal de este trabajo es, por un lado, estudiar las propiedades básicas de las distribuciones de tipo MTE y MOP y, por otro lado, desarrollar procedimientos que resuelvan el problema de la estimación de parámetros en dichas distribuciones a partir de grandes volúmenes de datos.

Comenzamos motivando la aparición de los modelos de tipo MoTBF a través de una breve reseña histórica en el capítulo 1. En este capítulo, también presentamos la notación necesaria para el posterior desarrollo del trabajo, así como una introducción general a las redes bayesianas híbridas, que constituyen el principal marco en el cual los modelos de tipo MoTBF adquieren una importancia fundamental.

A continuación, en el capítulo 2, se introduce ya el modelo MoTBF; primero, mediante sus dos casos particulares, a saber: el modelo MTE y el modelo MOP, y luego, de forma genérica. Para ello, presentamos, en cada caso, la definición tanto de las distribuciones univariantes como de las condicionadas. Posteriormente, se lleva a cabo un estudio acerca de las propiedades básicas de las distribuciones univariantes de tipo MTE y MOP. Finalizamos este capítulo ilustrando con ejemplos las propiedades obtenidas anteriormente.

En el capítulo 3, se aborda la segunda cuestión objeto de este trabajo, esto es, el estudio del problema de la estimación de parámetros en distribuciones de tipo MoTBF a partir de grandes volúmenes de datos. Para ello, consideramos como posibles soluciones el método de máxima verosimilitud (exacto), el método de *Fisher Scoring*, que es un método numérico para determinar de forma aproximada el estimador de máxima verosimilitud y, por último, el método de mínimos cuadrados. Para cada uno de dichos métodos, examinamos su aplicación a distribuciones de tipo MTE y MOP univariantes, exceptuando el caso del método de *Fisher Scoring* para el cual se prescinde del estudio para MTEs.

Después, en el capítulo 4, realizamos un análisis experimental de los métodos anteriormente desarrollados, centrándonos principalmente en el algoritmo de *Fisher Scoring*, a fin de estudiar su comportamiento en cuanto a la estimación de parámetros en distribuciones univariantes de tipo MOP a partir de grandes volúmenes de datos. Para dicho análisis, se hace uso del software R, a través de la implementación en esta herramienta del algoritmo mencionado, la cual incluimos en el apéndice A.

Finalmente, en el capítulo 5 se exponen las conclusiones del presente trabajo.

Introducción

1.1 Objetivos y Reseña histórica

En las últimas décadas, los modelos jerárquicos bayesianos, y en concreto las redes bayesianas [7], se han mostrado como una potente y extendida herramienta para el análisis de grandes volúmenes de datos. Dichas herramientas son capaces de gestionar de manera eficiente las relaciones de dependencia entre las variables aleatorias que componen el dominio del problema que queremos modelar, dado que la propia estructura de la red codifica las independencias.

Entre las principales virtudes de las redes bayesianas se encuentra la posibilidad de definir esquemas eficientes de cálculo de probabilidades que aprovechen la estructura del grafo subyacente para realizar dichos cálculos de forma local, evitando trabajar directamente con la distribución conjunta del modelo multivariante, usando en su lugar distribuciones condicionadas de menor dimensión que constituyen una factorización de la distribución conjunta. Un ejemplo destacado es el algoritmo de *propagación perezosa o lazy propagation* [5]. No obstante, este tipo de modelos también presenta algunos inconvenientes. Uno de ellos radica en la limitación de su alcance, en vista de que la mayor parte de los trabajos existentes en la literatura se centran en problemas donde todas las variables son cualitativas o discretas.

Una posible vía para incluir variables continuas consiste en discretizarlas. Sin embargo, este proceso conlleva una pérdida de información. Por este motivo, surge de forma natural el estudio de redes bayesianas donde puedan coexistir variables discretas y continuas, las cuales reciben el nombre de *redes bayesianas híbridas*.

La primera propuesta vino de la mano de S. L. Lauritzen, quien introdujo en 1992 el denominado modelo *Gaussiano condicional* [4], bajo el supuesto de que la distribución condicionada de las variables continuas dadas las discretas es una Normal multivariante, las discretas siguen una multinomial, y la distribución marginal sobre el vector de variables continuas es una mixtura de normales. No obstante, dicha solución presentaba una limitación, y es que impedía representar algunas relaciones, tales como las estructuras de variables discretas con padres continuos, debido a que las redes bayesianas que utilizaban el modelo Gaussiano condicional no admitían este tipo de estructuras.

Posteriormente, con el fin de resolver las restricciones estructurales del modelo Gaussiano condicional, se propuso el denominado modelo de *Mixturas de Exponenciales Truncadas (MTEs)* [6], que presentaba la ventaja de ser compatible con los esquemas de cálculo local desarrollados para variables discretas. A partir de ese momento, se desarrollaron múltiples trabajos en el campo de las redes bayesianas híbridas, entre los cuales destaca el modelo de *Mixturas de Polinomios (MOPs)* [12]. Además, recientemente, tanto las MTEs como las MOPs se han generalizado mediante el modelo de *Mixturas de Funciones Base Truncadas (MoTBFs)* [2].

En la actualidad, existen numerosos campos de aplicación de las redes bayesianas en los que conviven variables de tipo discreto y continuo. Sin embargo, no hace mucho, aún no se abordaban de forma totalmente satisfactoria, principalmente por dos motivos: en primer lugar, el escaso desarrollo metodológico de las redes bayesianas híbridas

y, en segundo lugar, la carencia de software que permitiese abordar tales aplicaciones por parte de usuarios de diferentes ámbitos. No obstante, I. Pérez-Bernabé en 2015 contribuyó a solucionar el primer problema [9] y, además, en el mismo texto, resolvió la segunda problemática expuesta, mediante la implementación del paquete MoTBFs en el software estadístico R [10].

Por este motivo, los objetivos principales de este trabajo son los siguientes:

1. Estudiar las propiedades básicas de las distribuciones de tipo mixtura de exponenciales truncadas y mixtura de polinomios.
2. Poner de manifiesto su utilidad en modelos jerárquicos bayesianos en entornos *big data*.
3. Estudiar el problema de la estimación de parámetros en dichas distribuciones, teniendo en cuenta la eficiencia en cuanto a capacidad de cálculo requerida.

1.2 Preliminares

Antes de comenzar es conveniente introducir algo de notación. Las variables aleatorias se denotarán con letras mayúsculas y, en el caso multidimensional, se utilizarán caracteres en negrita. El soporte de una variable aleatoria X se denota por Ω_X y el soporte de un vector aleatorio $\mathbf{X} = (X_1, \dots, X_n)$, por $\Omega_{\mathbf{X}} = \prod_{i=1}^n \Omega_{X_i}$. Con letras minúsculas, x (respectivamente, \mathbf{x}), denotaremos a los elementos de Ω_X (respectivamente, $\Omega_{\mathbf{X}}$).

Definición 1.1. Una *red bayesiana* sobre un vector aleatorio $\mathbf{X} = (X_1, \dots, X_n)$ es un grafo dirigido acíclico (DAG) donde cada nodo representa una variable aleatoria X_i , $i = 1, \dots, n$, y cada arco que une dos nodos indica la existencia de dependencia probabilística entre ambas variables.

Además, cada variable X_i , $i = 1, \dots, n$, tiene asociada una distribución de probabilidad condicionada dados sus padres en el grafo, denotados por $pa(X_i)$, de manera que la distribución de probabilidad conjunta sobre todas las variables de la red, $p(\mathbf{X})$, se factoriza como el producto de todas estas distribuciones condicionadas, esto es,

$$p(X_1, \dots, X_n) = \prod_{i=1}^n p(X_i | pa(X_i)). \quad (1.1)$$

Una de las ventajas más importantes de las redes bayesianas es que la topología del grafo asociado determina las relaciones de dependencia e independencia entre las variables, es decir, el flujo de información, según el criterio de d -separación introducido por J. Pearl [7]. De este modo, es posible conocer, sin necesidad de realizar cálculos numéricos, qué variables son relevantes o no para alguna otra variable de interés.

Las redes bayesianas se propusieron originalmente para el manejo de variables discretas, por lo que, hoy en día, podemos encontrar en la literatura una amplia y consolidada teoría al respecto. Sin embargo, en problemas reales, la presencia simultánea de dominios continuos y discretos es muy común.

Definición 1.2. Una *red bayesiana* se dice *híbrida* cuando variables aleatorias continuas y discretas coexisten simultáneamente en el modelo.

Distribuciones de tipo MoTBF

2.1 Definiciones básicas

En esta sección se introducen algunos marcos que permiten manejar simultáneamente variables de tipo continuo y discreto sin necesidad de recurrir a la discretización. De este modo, adoptaremos, en el presente texto, un enfoque que nos permitirá tratar con redes bayesianas híbridas. Dichos marcos son el denominado modelo de *Mixturas de Exponenciales Truncadas* (MTEs), el modelo de *Mixturas de Polinomios* (MOPs) y, en general, el modelo de *Mixturas de Funciones Base Truncadas* (MoTBFs). Estos modelos permiten representar en redes bayesianas híbridas ambos tipos de densidades presentes en la factorización (1.1), esto es, tanto densidades univariantes como densidades condicionadas.

Mixturas de Exponenciales Truncadas

Dado que la discretización se basa en aproximar cierta densidad objetivo mediante una *mixtura de uniformes*, es posible plantearse utilizar otro tipo de funciones en lugar de uniformes, con el fin de incrementar la precisión del modelo. Una alternativa favorable son las funciones exponenciales, debido a que presentan buenas propiedades como, por ejemplo, un alto poder de ajuste. Además, es sabido, gracias a [6], que la clase de los potenciales MTE es cerrada bajo restricción, marginalización y composición, operaciones que generalmente involucra la inferencia probabilística.

Ésta es la idea que subyace tras el llamado modelo de *Mixturas de Exponenciales Truncadas* (MTEs), introducido en 2001 por S. Moral, R. Rumí y A. Salmerón en [6].

Durante el proceso de inferencia probabilística, se suelen obtener funciones intermedias que no son necesariamente funciones de densidad. Por esta razón, es conveniente introducir el concepto de *potencial*, de la siguiente manera:

Caso univariante

Definición 2.1. Sea X una variable aleatoria continua con soporte Ω_X . Decimos que una función $f : \Omega_X \rightarrow \mathbb{R}_0^+$ es un **potencial de tipo *Mixtura de Exponenciales Truncadas*** (**potencial MTE**) si

$$f(x) = a_{0,j} + \sum_{i=1}^n a_{i,j} e^{b_{i,j}x}, \quad x \in A_j, \quad j = 1, \dots, l,$$

donde A_1, \dots, A_l es una partición de Ω_X por intervalos cuyos límites no dependen de x y $a_{i,j}, b_{k,j} \in \mathbb{R}$ para todo $i = 0, \dots, n$, $k = 1, \dots, n$ y $j = 1, \dots, l$.

Definición 2.2. Un potencial MTE f es una **función de densidad MTE** si

$$\int_{\Omega_X} f(x) dx = 1.$$

Se puede definir una *función de densidad MTE condicionada* para cierta variable aleatoria X dados sus padres mediante la partición del soporte de los padres de X (variables sobre las que se condiciona X) y la determinación de una densidad MTE para X para cada subconjunto de la partición de $\Omega_{pa(X)}$ dada.

Ejemplo 2.1. Sean X e Y dos variables aleatorias continuas con soporte Ω_X y Ω_Y , respectivamente. Supongamos que $\Omega_Y = [-2,9560, 2,8069]$. Entonces, una posible función de densidad MTE condicionada de X dado Y viene dada por

$$f(x|y) = \begin{cases} 0,1299 & \text{si } -2,9560 < y < -0,4240 \\ 0,4552 - 0,1090 e^{0,2x} - 0,1857 e^{-0,2x} & \text{si } -0,4240 < y < 2,8069 \end{cases} ,$$

para todo $x \in \Omega_X$.

Mixturas de Polinomios

Actualmente, las líneas de investigación referentes a las redes bayesianas híbridas se centran también en el denominado modelo de Mixturas de Polinomios (MOPs), el cual fue introducido en 2011 por P. P. Shenoy y J. C. West en [12] como una alternativa al modelo de Mixturas de Exponenciales Truncadas (MTEs). Mientras que en un potencial MTE la función base es una exponencial, el potencial MOP utiliza un polinomio como función base. Cabe señalar que las aproximaciones de tipo mixtura de polinomios tienen diversas ventajas sobre las aproximaciones de tipo mixtura de exponenciales truncadas, pues son más fáciles de encontrar.

Caso univariante

Definición 2.3. Sea X una variable aleatoria continua con soporte $\Omega_X \subset \mathbb{R}$. Una función unidimensional $f : \Omega_X \rightarrow \mathbb{R}$ se dice que es un **potencial de tipo Mixtura de Polinomios (potencial MOP)** si es una función definida a trozos de la forma:

$$f(x) = \sum_{i=0}^n a_{i,j} x^i, \quad x \in A_j, \quad j = 1, \dots, k,$$

donde A_1, \dots, A_k es una partición de Ω_X por intervalos cuyos límites no dependen de x y $a_{0,j}, \dots, a_{n,j}$ son constantes para todo $j = 1, \dots, k$. Diremos que f es una función MOP de k -trozos y grado n (suponiendo que $a_{n,j} \neq 0$ para algún j).

Definición 2.4. Un potencial MOP f es una **función de densidad MOP** si

$$\int_{\Omega_X} f(x) dx = 1.$$

De forma similar a la densidad MTE condicionada, es posible definir una *función de densidad MOP condicionada* para una variable aleatoria X dados sus padres a través de la partición del soporte de los padres de X y la definición de una densidad MOP para X para cada subconjunto de la partición de $\Omega_{pa(X)}$ realizada.

Ejemplo 2.2. Sean X e Y dos variables aleatorias continuas con soporte Ω_X y Ω_Y , respectivamente. Supongamos que $\Omega_Y = [-3,0411, 3,4207]$. Entonces, una posible función de densidad MOP condicionada de X dado Y viene dada por

$$f(x|y) = \begin{cases} 0,1774 - 0,0457x - 0,0099x^2 + 0,0026x^3 & \text{si } -3,0411 < y < -0,4203 \\ 0,1862 + 0,0054x - 0,0115x^2 - 0,0006x^3 & \text{si } -0,4203 < y < 0,4272, \\ 0,1861 + 0,0697x - 0,0117x^2 - 0,0048x^3 & \text{si } 0,4272 < y < 3,4207 \end{cases}$$

para todo $x \in \Omega_X$.

Mixturas de Funciones Base Truncadas

El marco MoTBF se introduce como una generalización de ambos marcos MTE y MOP. Esto es debido a que su definición generaliza el uso de funciones exponenciales y polinómicas en los dos marcos anteriores, respectivamente, mediante el uso de la notación abstracta de *funciones base*, $\psi(\cdot)$.

El éxito de las mixturas de funciones base truncadas se debe, en parte, al hecho de que permiten modelar redes bayesianas híbridas sin restricciones sobre la estructura de dependencias de las variables, en el sentido de que variables discretas pueden estar condicionadas a variables continuas. Asimismo, otro aspecto positivo es que, dado que la clase de las funciones MoTBF es cerrada bajo composición (producto), marginalización (integración o suma, según la variable sobre la que se marginaliza sea continua o discreta) y restricción (sustitución de una variable por un valor concreto), es posible utilizar esquemas eficientes y exactos de cálculo de probabilidades.

Caso univariante

Definición 2.5. Sea X una variable aleatoria continua con soporte (truncado) $\Omega_X \subseteq \mathbb{R}$ y sea $\Psi = \{\psi_0, \dots, \psi_k\}$ un conjunto de funciones base reales tal que $\psi_i : \mathbb{R} \rightarrow \mathbb{R}$ para todo $i = 0, \dots, k$. Decimos que una función $g_k : \Omega_X \rightarrow \mathbb{R}_0^+$ es un **potencial de tipo Mixtura de Funciones Base Truncadas (potencial MoTBF)** de nivel k si g_k puede escribirse como

$$g_k(x) = \sum_{i=0}^k a_i \psi_i(x), \quad (2.1)$$

para todo $x \in \Omega_X$, donde $a_i \in \mathbb{R}$, $i = 0, \dots, k$.

Definición 2.6. Un potencial MoTBF g_k es una **función de densidad MoTBF** si

$$\int_{\Omega_X} g_k(x) dx = 1.$$

Es importante notar que, regularmente, una distribución de tipo MoTBF univariante para una variable aleatoria X no depende de una partición de Ω_X .

Observación 2.1. Los modelos de Mixturas de Exponenciales Truncadas (MTEs) y Mixturas de Polinomios (MOPs) son casos particulares del modelo de Mixturas de Funciones Base Truncadas (MoTBFs).

Sea X como en la definición 2.5. Entonces, efectivamente:

- i. Si las funciones base son funciones polinómicas, esto es, si $\psi_i(x) = x^i$ para todo $i = 0, \dots, k$, entonces el modelo MoTBF se reduce al modelo MOP para distribuciones univariantes.
- ii. Del mismo modo, si las funciones base $\psi_i : \mathbb{R} \rightarrow \mathbb{R}$ vienen definidas por $\psi_i(x) = e^{b_i x}$, donde $b_i \in \mathbb{Z}$ para todo $i = 0, \dots, k$, entonces el modelo MoTBF corresponde a un modelo MTE con la excepción de que los parámetros de las funciones exponenciales son fijos.

Observemos que, en ambos casos, el modelo no depende de una partición de Ω_X , a diferencia de lo expuesto en las definiciones estándar 2.1 y 2.3.

Por último, presentemos la definición de una *distribución de tipo MoTBF condicionada*, que generaliza las definiciones introducidas anteriormente para MTEs y MOPs. Así, la influencia que ejerce un conjunto de variables aleatorias continuas padres, \mathbf{Z} , sobre su variable aleatoria hija, X , se establece simplemente mediante la partición del soporte de \mathbf{Z} , $\Omega_{\mathbf{Z}}$, en hipercubos y la determinación de una densidad de la forma (2.1) para cada hipercubo. Es decir, dada una partición $\mathcal{P} = \{\Omega_{\mathbf{Z}}^1, \dots, \Omega_{\mathbf{Z}}^m\}$ de $\Omega_{\mathbf{Z}}$, la **función de densidad MoTBF condicionada** de X dado $\mathbf{Z} \in \Omega_{\mathbf{Z}}^j$ se define como

$$g_k^{(j)}(x | \mathbf{z} \in \Omega_{\mathbf{Z}}^j) = \sum_{i=0}^k a_i^{(j)} \psi_i^{(j)}(x), \quad j = 1, \dots, m,$$

para todo $x \in \Omega_X$, donde $a_i^{(j)} \in \mathbb{R}$, $i = 0, \dots, k$, $j = 1, \dots, m$.

2.2 Propiedades

En esta sección estudiaremos las propiedades básicas de las distribuciones de tipo mixtura de funciones base truncadas. En concreto, nos centraremos en las distribuciones de tipo mixtura de exponenciales truncadas y mixtura de polinomios, considerando solamente el caso de distribuciones univariantes. Esto es, si una variable aleatoria X sigue una distribución de tipo MTE (respectivamente, MOP), veremos qué forma adoptan los momentos centrados y no centrados (en particular, la esperanza y varianza), la función generatriz de momentos, la función de distribución y la mediana y moda, cuando sea posible, de la v.a. X .

A partir de este momento, para simplificar el análisis, nos vamos a restringir al caso en que la variable aleatoria toma valores entre 0 y 1. Los siguientes resultados afirman que dicha restricción no supone una pérdida de generalidad. Es decir, demuestran que trabajar en el intervalo $[0, 1]$ es equivalente a trabajar en cualquier intervalo $[\alpha, \beta]$ de extremos finitos de \mathbb{R} . Así pues, siempre podremos reescalar la variable aleatoria en cuestión a un intervalo cualquiera de este tipo.

Comencemos abordando el caso de distribuciones de tipo MTE.

Proposición 2.1. *Sea una MTE definida en un intervalo $[\alpha, \beta]$ cualquiera con $\alpha, \beta \in \mathbb{R}$. Entonces, dicha distribución se puede transformar en una MTE definida en el intervalo $[0, 1]$.*

Igualmente, se puede realizar la transformación en el otro sentido.

Demostración:

Sea X una variable aleatoria con función de densidad MTE.

$$f_X(x) = a_0 + \sum_{i=1}^n a_i e^{b_i x}, \quad \alpha \leq x \leq \beta.$$

Sea

$$Y = \frac{X - \alpha}{\beta - \alpha}.$$

Entonces, $0 \leq Y \leq 1$ y

$$F_Y(x) = P[Y \leq x] = P\left[\frac{X - \alpha}{\beta - \alpha} \leq x\right] = P[X \leq (\beta - \alpha)x + \alpha] = F_X((\beta - \alpha)x + \alpha), \quad (2.2)$$

para todo $x \in [0, 1]$.

Ahora bien,

$$\begin{aligned} F_X(x) &= P[X \leq x] = \int_{\alpha}^x f_X(u) du = \int_{\alpha}^x \left(a_0 + \sum_{i=1}^n a_i e^{b_i u} \right) du = a_0 \int_{\alpha}^x du \\ &+ \sum_{i=1}^n a_i \int_{\alpha}^x e^{b_i u} du = a_0 [u]_{u=\alpha}^{u=x} + \sum_{i=1}^n a_i \left[\frac{e^{b_i u}}{b_i} \right]_{u=\alpha}^{u=x} = a_0 (x - \alpha) + \sum_{i=1}^n \frac{a_i}{b_i} (e^{b_i x} - e^{b_i \alpha}), \end{aligned} \quad (2.3)$$

para todo $x \in [\alpha, \beta]$.

Por lo tanto, gracias a las igualdades (2.2) y (2.3), se sigue que

$$F_Y(x) = F_X((\beta - \alpha)x + \alpha) = a_0 (\beta - \alpha)x + \sum_{i=1}^n \frac{a_i}{b_i} (e^{b_i((\beta - \alpha)x + \alpha)} - e^{b_i \alpha}), \quad 0 \leq x \leq 1.$$

De esta forma, la función de densidad de Y es

$$\begin{aligned} f_Y(x) &= \frac{d}{dx} F_Y(x) = a_0 (\beta - \alpha) + \sum_{i=1}^n a_i (\beta - \alpha) e^{b_i((\beta - \alpha)x + \alpha)} = \\ &= a_0 (\beta - \alpha) + \sum_{i=1}^n [a_i (\beta - \alpha) e^{b_i \alpha}] e^{[b_i (\beta - \alpha)]x}, \quad 0 \leq x \leq 1. \end{aligned}$$

Luego, Y es una v.a. con densidad MTE definida en el intervalo $[0, 1]$, como queríamos probar.

Análogamente, se puede transformar una MTE definida en $[0, 1]$ en una MTE definida en un intervalo de extremos finitos $[\alpha, \beta]$, en este caso mediante el cambio de variable

$$Y = (\beta - \alpha)X + \alpha.$$

■

A continuación, procedemos a enunciar el resultado correspondiente para distribuciones de tipo MOP.

Proposición 2.2. *Sea una MOP definida en un intervalo $[\alpha, \beta]$ cualquiera con $\alpha, \beta \in \mathbb{R}$. Entonces, esta distribución se puede transformar en una MOP definida en el intervalo $[0, 1]$.*

Asimismo, se puede hacer la transformación en sentido opuesto.

Demostración:

Sea X una variable aleatoria con función de densidad MOP

$$f_X(x) = \sum_{i=0}^n a_i x^i, \quad \alpha \leq x \leq \beta.$$

Sea

$$Y = \frac{X - \alpha}{\beta - \alpha}.$$

Entonces, $0 \leq Y \leq 1$ y

$$F_Y(x) = P[Y \leq x] = P\left[\frac{X - \alpha}{\beta - \alpha} \leq x\right] = P[X \leq (\beta - \alpha)x + \alpha] = F_X((\beta - \alpha)x + \alpha), \quad (2.4)$$

para todo $x \in [0, 1]$.

Por otra parte,

$$\begin{aligned} F_X(x) &= P[X \leq x] = \int_{\alpha}^x f_X(u) du = \int_{\alpha}^x \left(\sum_{i=0}^n a_i u^i \right) du = \sum_{i=0}^n a_i \int_{\alpha}^x u^i du = \\ &= \sum_{i=0}^n a_i \left[\frac{u^{i+1}}{i+1} \right]_{u=\alpha}^{u=x} = \sum_{i=0}^n \frac{a_i}{i+1} (x^{i+1} - \alpha^{i+1}), \quad \alpha \leq x \leq \beta. \end{aligned} \quad (2.5)$$

Por lo tanto, aplicando las igualdades (2.4) y (2.5), se tiene que

$$F_Y(x) = F_X((\beta - \alpha)x + \alpha) = \sum_{i=0}^n \frac{a_i}{i+1} (((\beta - \alpha)x + \alpha)^{i+1} - \alpha^{i+1}), \quad 0 \leq x \leq 1.$$

Luego, la función de densidad de Y es

$$f_Y(x) = \frac{d}{dx} F_Y(x) = \sum_{i=0}^n a_i (\beta - \alpha) ((\beta - \alpha)x + \alpha)^i, \quad 0 \leq x \leq 1.$$

Por consiguiente, aplicando el teorema del binomio y reagrupando convenientemente los términos, se deduce que f_Y es una función de densidad de tipo MOP definida en el intervalo $[0, 1]$, como queríamos demostrar.

De forma análoga, se puede transformar una MOP definida en $[0, 1]$ en una MOP definida en un intervalo de extremos finitos $[\alpha, \beta]$ mediante el cambio de variable

$$Y = (\beta - \alpha)X + \alpha$$

y siguiendo un procedimiento análogo al anterior. ■

Finalmente, el caso general para distribuciones de tipo MoTBF cuyas funciones base admiten primitiva se puede demostrar siempre que se verifique una condición adicional, tal y como se muestra en la proposición 2.3.

Proposición 2.3. Sea una MoTBF definida en un intervalo $[\alpha, \beta]$ cualquiera con $\alpha, \beta \in \mathbb{R}$ y sea η_i la función primitiva de la función base ψ_i , $i = 0, \dots, k$. Supongamos que existe cierta función $h(\alpha, \beta)$, que depende del conjunto de funciones base $\Psi = \{\psi_0, \dots, \psi_k\}$, tal que

$$\psi_i((\beta - \alpha)x + \alpha) = h(\alpha, \beta) \psi_i(x), \quad (2.6)$$

para todo $i = 0, \dots, k$. Entonces, dicha distribución se puede transformar en una MoTBF definida en el intervalo $[0, 1]$.

De igual forma, se puede efectuar la transformación en sentido contrario.

Demostración:

Sea X una variable aleatoria con función de densidad MoTBF

$$g_k(x) = \sum_{i=0}^k a_i \psi_i(x), \quad \alpha \leq x \leq \beta.$$

Sea

$$Y = \frac{X - \alpha}{\beta - \alpha}.$$

Entonces, $0 \leq Y \leq 1$ y

$$F_Y(x) = P[Y \leq x] = P\left[\frac{X - \alpha}{\beta - \alpha} \leq x\right] = P[X \leq (\beta - \alpha)x + \alpha] = F_X((\beta - \alpha)x + \alpha), \quad (2.7)$$

para todo $x \in [0, 1]$.

Por otro lado,

$$\begin{aligned} F_X(x) &= P[X \leq x] = \int_{\alpha}^x g_k(u) du = \int_{\alpha}^x \left(\sum_{i=0}^k a_i \psi_i(u) \right) du = \sum_{i=0}^k a_i \int_{\alpha}^x \psi_i(u) du = \\ &= \sum_{i=0}^k a_i [\eta_i(u)]_{u=\alpha}^{u=x} = \sum_{i=0}^k a_i [\eta_i(x) - \eta_i(\alpha)], \quad \alpha \leq x \leq \beta, \end{aligned} \quad (2.8)$$

donde hemos utilizado que η_i es la función primitiva de la función base ψ_i , $i = 0, \dots, k$.

Por lo tanto, en virtud de las igualdades (2.7) y (2.8), tenemos que

$$F_Y(x) = F_X((\beta - \alpha)x + \alpha) = \sum_{i=0}^k a_i [\eta_i((\beta - \alpha)x + \alpha) - \eta_i(\alpha)], \quad 0 \leq x \leq 1.$$

Así, la función de densidad de Y es

$$\begin{aligned} f_Y(x) &= \frac{d}{dx} F_Y(x) = \sum_{i=0}^k a_i (\beta - \alpha) \eta_i'((\beta - \alpha)x + \alpha) = \\ &= \sum_{i=0}^k a_i (\beta - \alpha) \psi_i((\beta - \alpha)x + \alpha) = \sum_{i=0}^k a_i (\beta - \alpha) h(\alpha, \beta) \psi_i(x), \quad 0 \leq x \leq 1, \end{aligned}$$

donde en la última igualdad hemos hecho uso de la hipótesis (2.6).

Por tanto, Y es una v.a. con densidad MoTBF definida en el intervalo $[0, 1]$, como queríamos ver.

Análogamente, se puede transformar una MoTBF definida en $[0, 1]$ en una MoTBF definida en un intervalo de extremos finitos $[\alpha, \beta]$ mediante el cambio de variable

$$Y = (\beta - \alpha)X + \alpha$$

y siguiendo un procedimiento análogo al anterior. ■

Ahora sí, comencemos el estudio de las propiedades fundamentales de las distribuciones de tipo MoTBF.

Distribución de tipo MTE univariante

Esperanza y varianza

Proposición 2.4. *Sea X una variable aleatoria continua con función de densidad MTE $f_X : [0, 1] \rightarrow \mathbb{R}_0^+$ dada por*

$$f_X(x) = a_0 + \sum_{i=1}^n a_i e^{b_i x}, \quad 0 \leq x \leq 1,$$

donde $a_i, b_j \in \mathbb{R}$ para todo $i = 0, \dots, n, j = 1, \dots, n$. Entonces:

i. La esperanza de X viene dada por

$$\mathbb{E}[X] = \frac{a_0}{2} + \sum_{i=1}^n \frac{a_i}{b_i^2} [1 + e^{b_i} (b_i - 1)].$$

ii. El momento no centrado de orden 2 adopta la expresión

$$\mathbb{E}[X^2] = \frac{a_0}{3} + \sum_{i=1}^n \frac{a_i}{b_i^3} [-2 + e^{b_i} (b_i^2 - 2b_i + 2)].$$

Demostración:

i. La esperanza o momento no centrado de orden 1 tiene la expresión siguiente:

$$\begin{aligned} \mathbb{E}[X] &= \int_0^1 x dF_X(x) = \int_0^1 x f_X(x) dx = \int_0^1 x \left(a_0 + \sum_{i=1}^n a_i e^{b_i x} \right) dx = a_0 \int_0^1 x dx \\ &+ \sum_{i=1}^n a_i \int_0^1 x e^{b_i x} dx = a_0 \left[\frac{x^2}{2} \right]_{x=0}^{x=1} + \sum_{i=1}^n a_i \left[\frac{1 + e^{b_i} (b_i - 1)}{b_i^2} \right] = \frac{a_0}{2} + \sum_{i=1}^n \frac{a_i}{b_i^2} [1 + e^{b_i} (b_i - 1)], \end{aligned}$$

donde hemos utilizado que

$$\begin{aligned} \int_0^1 x e^{b_i x} dx &= \left[\begin{array}{l} u = x \Rightarrow du = dx \\ dv = e^{b_i x} dx \Rightarrow v = \frac{1}{b_i} e^{b_i x} \end{array} \right] = \left[\frac{1}{b_i} x e^{b_i x} \right]_{x=0}^{x=1} - \int_0^1 \frac{1}{b_i} e^{b_i x} dx = \\ &= \frac{1}{b_i} \left[x e^{b_i x} \right]_{x=0}^{x=1} - \frac{1}{b_i} \left[\frac{e^{b_i x}}{b_i} \right]_{x=0}^{x=1} = \frac{1}{b_i} e^{b_i} - \frac{1}{b_i^2} (e^{b_i} - 1) = \frac{b_i e^{b_i} - e^{b_i} + 1}{b_i^2} = \frac{1 + e^{b_i} (b_i - 1)}{b_i^2}. \end{aligned}$$

ii. Por otro lado, el momento no centrado de orden 2 viene dado por la siguiente expresión:

$$\begin{aligned} E[X^2] &= \int_0^1 x^2 f_X(x) dx = \int_0^1 x^2 \left(a_0 + \sum_{i=1}^n a_i e^{b_i x} \right) dx = a_0 \int_0^1 x^2 dx + \sum_{i=1}^n a_i \int_0^1 x^2 e^{b_i x} dx = \\ &= a_0 \left[\frac{x^3}{3} \right]_{x=0}^{x=1} + \sum_{i=1}^n a_i \left[\frac{-2 + e^{b_i} (b_i^2 - 2b_i + 2)}{b_i^3} \right] = \frac{a_0}{3} + \sum_{i=1}^n \frac{a_i}{b_i^3} [-2 + e^{b_i} (b_i^2 - 2b_i + 2)], \end{aligned}$$

donde hemos hecho uso de la siguiente igualdad:

$$\begin{aligned} \int_0^1 x^2 e^{b_i x} dx &= \left[\begin{array}{l} u = x^2 \Rightarrow du = 2x dx \\ dv = e^{b_i x} dx \Rightarrow v = \frac{1}{b_i} e^{b_i x} \end{array} \right] = \left[\frac{1}{b_i} x^2 e^{b_i x} \right]_{x=0}^{x=1} - \int_0^1 \frac{2}{b_i} x e^{b_i x} dx = \\ &= \frac{1}{b_i} e^{b_i} - \frac{2}{b_i} \int_0^1 x e^{b_i x} dx = \frac{1}{b_i} e^{b_i} - \frac{2}{b_i} \left[\frac{1 + e^{b_i} (b_i - 1)}{b_i^2} \right] = \frac{-2 + e^{b_i} (b_i^2 - 2b_i + 2)}{b_i^3}. \end{aligned}$$

■

Corolario 2.4.1. Sea X como en la proposición 2.4. Entonces, la varianza o momento centrado de orden 2 viene dado por

$$\text{Var}(X) = \frac{a_0}{3} + \sum_{i=1}^n \frac{a_i}{b_i^3} [-2 + e^{b_i} (b_i^2 - 2b_i + 2)] - \left[\frac{a_0}{2} + \sum_{i=1}^n \frac{a_i}{b_i^2} [1 + e^{b_i} (b_i - 1)] \right]^2.$$

Demostración:

La varianza o momento centrado de orden 2 se define como

$$\text{Var}(X) = E[(X - E[X])^2] = E[X^2] - E[X]^2.$$

Por lo tanto, basta sustituir $E[X^2]$ y $E[X]$ por sus respectivas expresiones en la proposición 2.4 para obtener lo deseado.

■

Función generatriz de momentos

Proposición 2.5. *Sea X como en la proposición 2.4. Entonces, no existe la función generatriz de momentos de la v.a. X .*

Demostración:

Sea $c \in \mathbb{R}^+ \cup \{+\infty\}$ y sea $t \in (-c, c)$. Entonces, la función generatriz de momentos se calcula como

$$M_X(t) = E[e^{tX}] = \int_0^1 e^{tx} f_X(x) dx = \int_0^1 e^{tx} \left(a_0 + \sum_{i=1}^n a_i e^{b_i x} \right) dx = a_0 \int_0^1 e^{tx} dx + \sum_{i=1}^n a_i \int_0^1 e^{(t+b_i)x} dx = a_0 \left[\frac{e^{tx}}{t} \right]_{x=0}^{x=1} + \sum_{i=1}^n a_i \left[\frac{e^{(t+b_i)x}}{t+b_i} \right]_{x=0}^{x=1} = a_0 \frac{e^t - 1}{t} + \sum_{i=1}^n a_i \frac{e^{t+b_i} - 1}{t+b_i}.$$

Por consiguiente, como la variable t constituye el denominador del primer miembro, se sigue que la función generatriz de momentos de X no existe en un entorno $(-c, c)$ de 0, puesto que no es posible dividir entre 0. ■

Observación 2.2. *No obstante, se puede demostrar que la función generatriz de momentos existe si consideramos una distribución de tipo MTE sin término independiente.*

En efecto, sea X una variable aleatoria continua con función de densidad MTE $f_X : [0, 1] \rightarrow \mathbb{R}_0^+$ dada por

$$f_X(x) = \sum_{i=1}^n a_i e^{b_i x}, \quad 0 \leq x \leq 1,$$

donde $a_i, b_j \in \mathbb{R}$ para todo $i, j = 1, \dots, n$. Entonces, la función generatriz de momentos de la v.a. X viene dada por

$$M_X(t) = \sum_{i=1}^n a_i \frac{e^{t+b_i} - 1}{t+b_i},$$

para todo $t \in (-c, c)$ con $c \in \mathbb{R}^+ \cup \{+\infty\}$.

Función de distribución

Proposición 2.6. *Sea X como en la proposición 2.4. Entonces, la función de distribución de la v.a. X es una función $F_X : \mathbb{R} \rightarrow [0, 1]$ definida como*

$$F_X(x) = \begin{cases} 0, & x < 0 \\ a_0 x + \sum_{i=1}^n \frac{a_i}{b_i} (e^{b_i x} - 1), & 0 \leq x < 1 \\ 1, & x \geq 1 \end{cases}$$

Demostración:

Sea $x \in [0, 1]$. Entonces, la expresión de la función de distribución viene dada por

$$F_X(x) = P[X \leq x] = \int_0^x f_X(u) du = \int_0^x \left(a_0 + \sum_{i=1}^n a_i e^{b_i u} \right) du =$$

$$\begin{aligned}
&= a_0 \int_0^x du + \sum_{i=1}^n a_i \int_0^x e^{b_i u} du = a_0 [u]_{u=0}^{u=x} + \sum_{i=1}^n a_i \frac{1}{b_i} [e^{b_i u}]_{u=0}^{u=x} = \\
&= a_0 x + \sum_{i=1}^n \frac{a_i}{b_i} (e^{b_i x} - 1).
\end{aligned}$$

■

Mediana y moda

Proposición 2.7. Sea X como en la proposición 2.4. Entonces, ni la mediana ni la moda de la v.a. X se pueden calcular de forma cerrada.

Demostración:

- i. Por una parte, la mediana se define como el valor $m \in [0, 1]$ que verifica la igualdad $F_X(m) = \frac{1}{2}$.

Luego, por la proposición 2.6, la mediana m es la solución en $[0, 1]$ de la ecuación

$$a_0 x + \sum_{i=1}^n \frac{a_i}{b_i} (e^{b_i x} - 1) = \frac{1}{2}.$$

Como podemos observar, el hecho de que la incógnita aparezca en el exponente de los términos exponenciales impide que podamos despejarla en la ecuación. Por lo tanto, no es posible calcular la mediana de forma cerrada, como queríamos demostrar.

- ii. Por otra parte, la moda es el valor más probable, es decir, $Mo = \hat{x}$ tal que

$$f_X(\hat{x}) = \max\{f_X(x) : 0 \leq x \leq 1\} = \max\{f_X(0), f_X(\tilde{x}), f_X(1)\},$$

donde \tilde{x} es un máximo relativo de f_X , esto es, $\tilde{x} \in]0, 1[$ verificando $f_X'(\tilde{x}) = 0$ y $f_X''(\tilde{x}) < 0$.

En el proceso de obtención de la moda son necesarios los siguientes cálculos:

$$\begin{aligned}
f_X'(x) &= \sum_{i=1}^n a_i b_i e^{b_i x}, & f_X(0) &= a_0 + \sum_{i=1}^n a_i e^{b_i \cdot 0} = \sum_{i=0}^n a_i, \\
f_X''(x) &= \sum_{i=1}^n a_i b_i^2 e^{b_i x}, & f_X(1) &= a_0 + \sum_{i=1}^n a_i e^{b_i}.
\end{aligned}$$

De este modo, en el cálculo de la moda se añade la dificultad de que se pueda alcanzar en el 0, en el 1 o en un punto interior del intervalo $[0, 1]$. En este último caso, siguiendo el procedimiento usual de cálculo de un máximo relativo, nos encontramos con el mismo problema que presentaba el cálculo de la mediana: no es posible despejar x de la ecuación $f_X'(x) = 0$. Por consiguiente, la moda tampoco se puede calcular de forma cerrada.

■

Por lo tanto, la mediana y la moda de una distribución de tipo MTE se podrán obtener solamente mediante métodos numéricos o mediante simulación.

Distribución de tipo MOP univariante

Esperanza y varianza

Proposición 2.8. Sea X una variable aleatoria continua con función de densidad MOP $f_X : [0, 1] \rightarrow \mathbb{R}_0^+$ dada por

$$f_X(x) = \sum_{i=0}^n a_i x^i, \quad 0 \leq x \leq 1$$

donde $a_i \in \mathbb{R}$ para todo $i = 0, \dots, n$. Entonces, el momento no centrado de orden k viene dado por

$$\mathbb{E}[X^k] = \sum_{i=0}^n \frac{a_i}{i+k+1},$$

para todo $k \in \mathbb{N}$.

En particular:

i. La esperanza (momento no centrado de orden 1) de X viene dada por

$$\mathbb{E}[X] = \sum_{i=0}^n \frac{a_i}{i+2}.$$

ii. El momento no centrado de orden 2 es

$$\mathbb{E}[X^2] = \sum_{i=0}^n \frac{a_i}{i+3}.$$

Demostración:

Sea $k \in \mathbb{N}$. Entonces, el momento no centrado de orden k viene dado por la siguiente expresión:

$$\begin{aligned} \mathbb{E}[X^k] &= \int_0^1 x^k f_X(x) dx = \int_0^1 x^k \left(\sum_{i=0}^n a_i x^i \right) dx = \int_0^1 \sum_{i=0}^n a_i x^{i+k} dx = \\ &= \sum_{i=0}^n a_i \int_0^1 x^{i+k} dx = \sum_{i=0}^n a_i \left[\frac{x^{i+k+1}}{i+k+1} \right]_{x=0}^{x=1} = \sum_{i=0}^n \frac{a_i}{i+k+1} (1-0) = \sum_{i=0}^n \frac{a_i}{i+k+1}. \end{aligned}$$

En particular, evaluando la expresión anterior en $k = 1$ y $k = 2$ se obtienen las expresiones del momento no centrado de orden 1 y 2, respectivamente. ■

Corolario 2.8.1. Sea X como en la proposición 2.8. Entonces, la varianza o momento centrado de orden 2 tiene la siguiente expresión:

$$\text{Var}(X) = \sum_{i=0}^n \frac{a_i}{i+3} - \left(\sum_{i=0}^n \frac{a_i}{i+2} \right)^2.$$

Demostración:

Es conocido que la varianza o momento centrado de orden 2 viene dada por:

$$\text{Var}(X) = E[(X - E[X])^2] = E[X^2] - E[X]^2.$$

Por tanto, basta sustituir $E[X^2]$ y $E[X]$ por sus respectivas expresiones, obtenidas en la proposición 2.8, para obtener lo que queremos. ■

Función generatriz de momentos

Proposición 2.9. *Sea X como en la proposición 2.8. Entonces, la función generatriz de momentos de la v.a. X viene dada por*

$$M_X(t) = \sum_{k=0}^{\infty} \left(\frac{t^k}{k!} \sum_{i=0}^n \frac{a_i}{i+k+1} \right),$$

para todo $t \in (-c, c)$ con $c \in \mathbb{R}^+ \cup \{+\infty\}$.

Demostración:

La función generatriz de momentos se puede expresar a partir de los momentos, valga la redundancia, como la suma de una serie, de la forma

$$M_X(t) = 1 + t m_1 + \frac{t^2}{2!} m_2 + \dots = \sum_{k=0}^{\infty} \frac{t^k}{k!} m_k, \quad t \in (-c, c) \text{ con } c \in \mathbb{R}^+ \cup \{+\infty\},$$

donde m_k es el momento no centrado de orden k , esto es, $m_k = E[X^k]$. Por la proposición 2.8, se tiene que

$$m_k = E[X^k] = \sum_{i=0}^n \frac{a_i}{i+k+1}$$

para todo $k \in \mathbb{N}$. Por consiguiente, dados $c \in \mathbb{R}^+ \cup \{+\infty\}$ y $t \in (-c, c)$, se sigue que

$$M_X(t) = 1 + t \sum_{i=0}^n \frac{a_i}{i+2} + \frac{t^2}{2!} \sum_{i=0}^n \frac{a_i}{i+3} + \dots = \sum_{k=0}^{\infty} \left(\frac{t^k}{k!} \sum_{i=0}^n \frac{a_i}{i+k+1} \right).$$

La expresión de la función de generatriz dada por la proposición 2.9 permite obtener una aproximación hasta cierto término N , pese a no poderse calcular la suma de la serie.

Por otro lado, cabe señalar la siguiente observación:

Observación 2.3. *Se puede comprobar que $M_X^{(n)}(0) = E[X^n]$.*

En particular, para $n = 1$, se tiene que $M_X'(0) = E[X]$. En efecto,

$$M_X(t) = \sum_{k=0}^{\infty} \left(\frac{t^k}{k!} \sum_{i=0}^n \frac{a_i}{i+k+1} \right) = \sum_{i=0}^n \frac{a_i}{i+1} + \sum_{k=1}^{\infty} \left(\frac{t^k}{k!} \sum_{i=0}^n \frac{a_i}{i+k+1} \right)$$

$$\begin{aligned} \Rightarrow M'_X(t) &= \sum_{k=1}^{\infty} \left(\frac{k t^{k-1}}{k!} \sum_{i=0}^n \frac{a_i}{i+k+1} \right) = \sum_{i=0}^n \frac{a_i}{i+2} + \sum_{k=2}^{\infty} \left(\frac{k t^{k-1}}{k!} \sum_{i=0}^n \frac{a_i}{i+k+1} \right) \\ \Rightarrow M'_X(0) &= \sum_{i=0}^n \frac{a_i}{i+2} = E[X]. \end{aligned}$$

Función de distribución

Proposición 2.10. Sea X como en la proposición 2.8. Entonces, la función de distribución de la v.a. X es una función $F_X : \mathbb{R} \rightarrow [0, 1]$ definida como

$$F_X(x) = \begin{cases} 0, & x < 0 \\ \sum_{i=0}^n \frac{a_i}{i+1} x^{i+1}, & 0 \leq x < 1. \\ 1, & x \geq 1 \end{cases}$$

Demostración:

Sea $x \in [0, 1]$. Entonces, la función de distribución viene dada por la expresión

$$\begin{aligned} F_X(x) &= P[X \leq x] = \int_0^x f_X(u) du = \int_0^x \left(\sum_{i=0}^n a_i u^i \right) du = \sum_{i=0}^n a_i \int_0^x u^i du = \\ &= \sum_{i=0}^n a_i \left[\frac{u^{i+1}}{i+1} \right]_{u=0}^{u=x} = \sum_{i=0}^n \frac{a_i}{i+1} (x^{i+1} - 0) = \sum_{i=0}^n \frac{a_i}{i+1} x^{i+1}. \end{aligned}$$

■

Mediana

Proposición 2.11. Sea X como en la proposición 2.8. Entonces, la mediana de la v.a. X es la solución en $[0, 1]$ de la ecuación

$$\sum_{i=0}^n \frac{a_i}{i+1} x^{i+1} = \frac{1}{2}$$

y se puede calcular de forma cerrada.

Demostración:

La mediana se define como el valor $m \in [0, 1]$ que verifica la igualdad $F_X(m) = \frac{1}{2}$.

Luego, por la proposición 2.10, la mediana m es la solución en $[0, 1]$ de la ecuación

$$\sum_{i=0}^n \frac{a_i}{i+1} x^{i+1} = \frac{1}{2}. \quad (2.9)$$

De esta forma, se obtiene una ecuación polinómica en la variable x de grado $n+1$. Por el teorema fundamental del álgebra, la ecuación (2.9) tiene exactamente $n+1$ soluciones complejas, teniendo en cuenta multiplicidades. Entre ellas, la mediana será aquella solución que pertenezca al conjunto de los números reales y, en particular, al intervalo $[0, 1]$.

■

Moda

Proposición 2.12. *Sea X como en la proposición 2.8. Entonces, la moda de la v.a. X es la solución de la ecuación*

$$f_X(x) = \text{máx}\{f_X(0), f_X(\tilde{x}), f_X(1)\},$$

donde \tilde{x} es la solución en $]0, 1[$ del sistema de ecuaciones

$$\begin{cases} \sum_{i=1}^n a_i i x^{i-1} = 0 \\ \sum_{i=2}^n a_i i (i-1) x^{i-2} < 0 \end{cases}$$

y se puede obtener de forma cerrada.

Demostración:

La moda se define como el valor donde se alcanza el máximo de f_X , es decir, $Mo = \hat{x}$ donde

$$f_X(\hat{x}) = \text{máx}\{f_X(x) : 0 \leq x \leq 1\} = \text{máx}\{f_X(0), f_X(\tilde{x}), f_X(1)\},$$

siendo \tilde{x} un máximo relativo de f_X , esto es, $\tilde{x} \in]0, 1[$ tal que $f'_X(\tilde{x}) = 0$ y $f''_X(\tilde{x}) < 0$.

Durante el procedimiento de obtención de la moda son necesarios los siguientes cálculos:

$$\begin{aligned} f'_X(x) &= \sum_{i=1}^n a_i i x^{i-1}, & f_X(0) &= a_0, \\ f''_X(x) &= \sum_{i=2}^n a_i i (i-1) x^{i-2}, & f_X(1) &= \sum_{i=0}^n a_i. \end{aligned}$$

De nuevo, al igual que ocurría con la distribución de tipo MTE, en el cálculo de la moda se añade la dificultad de que se pueda alcanzar en los extremos del intervalo $[0, 1]$ o en un punto interior de dicho intervalo, esto es, en algún punto del intervalo $]0, 1[$. Sin embargo, en el caso de las mixturas de polinomios, sí es posible obtener de forma cerrada los máximos relativos de f_X , debido a que de la igualdad $f'_X(x) = 0$ se obtiene una ecuación polinómica en x de grado $n - 1$, la cual tiene, por el teorema fundamental del álgebra, $n - 1$ soluciones. Entre ellas, la moda será aquella que pertenezca al intervalo $[0, 1]$ y que verifique la desigualdad $f''_X(\tilde{x}) < 0$. ■

2.3 Ejemplos

En la presente sección haremos uso del paquete MoTBFs [10] implementado en el software estadístico R para generar funciones de densidad de tipo MoTBF, tanto de tipo MTE como de tipo MOP, e ilustrar así, con un ejemplo, las propiedades básicas que presentan este tipo de distribuciones. Para ello, haremos uso de las fórmulas genéricas obtenidas en la sección 2.2.

Distribución de tipo MTE univariante

Sea X una variable aleatoria continua con función de densidad MTE $f_X : [0, 1] \rightarrow \mathbb{R}_0^+$ dada por

$$f_X(x) = 34,7148483 - 14,1688395 e^{2x} + 34,6492469 e^{-2x} + 1,81756650 e^{4x} - 228,642716 e^{-4x} - 0,0574034011 e^{6x} + 281,635541 e^{-6x} - 0,00243540255 e^{8x} - 109,944808 e^{-8x},$$

para todo $x \in [0, 1]$. La representación gráfica de esta función de densidad se puede observar en la Figura 2.1.

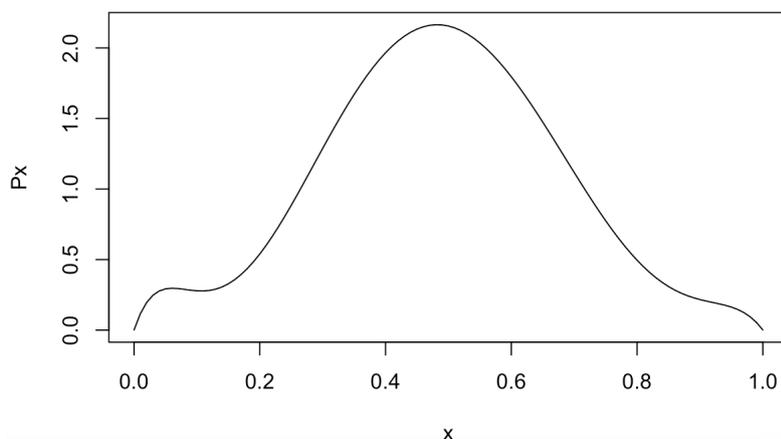


Figura 2.1: Función de densidad de tipo MTE

Cabe mencionar que tomaremos, en la presente subsección, solamente 9 cifras significativas para todos los coeficientes y valores numéricos que aparezcan a continuación.

Esperanza y varianza

Sustituyendo los valores de los respectivos parámetros en la fórmula de la esperanza obtenida en la proposición 2.4, se deduce que la esperanza de X es

$$E[X] = \frac{a_0}{2} + \sum_{i=1}^8 \frac{a_i}{b_i^2} [1 + e^{b_i} (b_i - 1)] \approx 0,488466468.$$

De la misma forma, aplicando la proposición 2.4, obtenemos que el momento no centrado de orden 2 es

$$E[X^2] = \frac{a_0}{3} + \sum_{i=1}^8 \frac{a_i}{b_i^3} [-2 + e^{b_i} (b_i^2 - 2b_i + 2)] \approx 0,272528405.$$

Así, por el corolario 2.4.1, la varianza o momento centrado de orden 2 será

$$Var(X) = E[X^2] - E[X]^2 \approx 0,272528405 - (0,488466468)^2 \approx 0,0339289142.$$

Función generatriz de momentos

Según la proposición 2.5, la función generatriz de momentos de la v.a. X viene dada por

$$\begin{aligned} M_X(t) &= a_0 \frac{e^t - 1}{t} + \sum_{i=1}^8 a_i \frac{e^{t+b_i} - 1}{t + b_i} = \\ &= 34,7148483 \frac{e^t - 1}{t} - 14,1688395 \frac{e^{t+2} - 1}{t+2} + 34,6492469 \frac{e^{t-2} - 1}{t-2} \\ &+ 1,81756650 \frac{e^{t+4} - 1}{t+4} - 228,642716 \frac{e^{t-4} - 1}{t-4} - 0,0574034011 \frac{e^{t+6} - 1}{t+6} \\ &+ 281,635541 \frac{e^{t-6} - 1}{t-6} - 0,00243540255 \frac{e^{t+8} - 1}{t+8} - 109,944808 \frac{e^{t-8} - 1}{t-8}, \end{aligned}$$

para todo $t \in (-c, c)$ con $c \in \mathbb{R}^+ \cup \{+\infty\}$.

Función de distribución

La función de distribución $F_X : \mathbb{R} \rightarrow [0, 1]$ de la v.a. X se define a trozos como sigue:

$$F_X(x) = \begin{cases} 0, & x < 0 \\ \hat{F}(x), & 0 \leq x < 1, \\ 1, & x \geq 1 \end{cases}$$

en virtud de la proposición 2.6, donde

$$\begin{aligned} \hat{F}(x) &:= a_0 x + \sum_{i=1}^8 \frac{a_i}{b_i} (e^{b_i x} - 1) = 34,7148483 x \\ &- 7,08441975 e^{2x} - 17,3246235 e^{-2x} + 0,454391625 e^{4x} \\ &+ 57,1606790 e^{-4x} - 0,00956723352 e^{6x} - 46,9392568 e^{-6x} \\ &- 0,000304425319 e^{8x} + 13,7431010 e^{-8x} + 6,71687506 \cdot 10^{-8}. \end{aligned}$$

Su representación gráfica viene dada por la Figura 2.2.

Distribución de tipo MOP univariante

Sea X una variable aleatoria continua con función de densidad MOP $f_X : [0, 1] \rightarrow \mathbb{R}_0^+$ dada por

$$\begin{aligned} f_X(x) &= 0,1035292111 + 4,016135217 x - 55,21646272 x^2 + 389,0507868 x^3 \\ &- 927,4755004 x^4 + 894,9692824 x^5 - 305,4467705 x^6, \end{aligned}$$

para todo $x \in [0, 1]$. La representación gráfica de dicha función de densidad se puede ver en la Figura 2.3.

Es importante resaltar que consideraremos, en este caso, únicamente 10 cifras significativas para todos los valores numéricos y coeficientes que aparezcan a partir de este momento en la presente subsección.

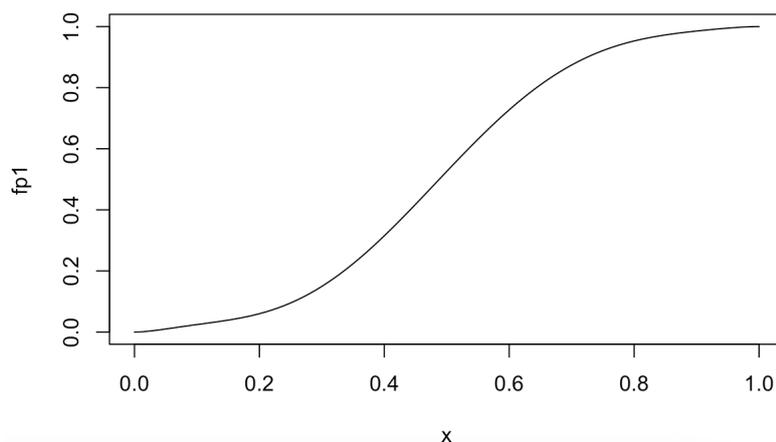


Figura 2.2: Función de distribución de tipo MTE

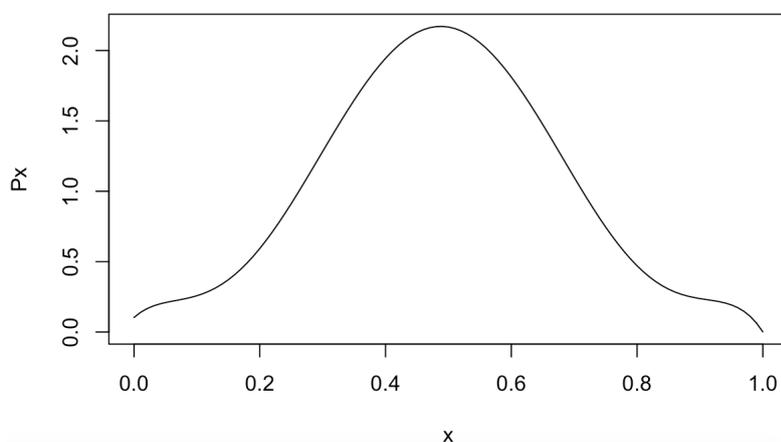


Figura 2.3: Función de densidad de tipo MOP

Esperanza y varianza

Sustituyendo los valores de los parámetros en la fórmula calculada en la proposición 2.8, se deduce que el momento no centrado de orden k es

$$E[X^k] = \sum_{i=0}^6 \frac{a_i}{i+k+1} = \frac{0,1035292111}{k+1} + \frac{4,016135217}{k+2} - \frac{55,21646272}{k+3} + \frac{389,0507868}{k+4} - \frac{927,4755004}{k+5} + \frac{894,9692824}{k+6} - \frac{305,4467705}{k+7},$$

para todo $k = 1, 2, \dots$

En particular, el momento no centrado de orden 1, es decir, la esperanza es

$$E[X] = \sum_{i=0}^6 \frac{a_i}{i+2} \approx 0,4891762740.$$

Asimismo, también gracias a la proposición 2.8, el momento no centrado de orden 2 es

$$E[X^2] = \sum_{i=0}^6 \frac{a_i}{i+3} \approx 0,2731789846.$$

Luego, por el corolario 2.8.1, la varianza o momento centrado de orden 2 será

$$\begin{aligned} Var(X) &= E[(X - E[X])^2] = E[X^2] - E[X]^2 \approx \\ &\approx 0,2731789846 - (0,4891762740)^2 \approx 0,03388555759. \end{aligned}$$

Función generatriz de momentos

En virtud de la proposición 2.9, se sigue que la función generatriz de momentos de la v.a. X viene dada por

$$\begin{aligned} M_X(t) &= 1 + 0,4891762740t + 0,2731789846 \frac{t^2}{2!} + \dots = \sum_{k=0}^{\infty} \left(\frac{t^k}{k!} \sum_{i=0}^6 \frac{a_i}{i+k+1} \right) = \\ &= \sum_{k=0}^{\infty} \frac{t^k}{k!} \left(\frac{0,1035292111}{k+1} + \frac{4,016135217}{k+2} - \frac{55,21646272}{k+3} + \frac{389,0507868}{k+4} \right. \\ &\quad \left. - \frac{927,4755004}{k+5} + \frac{894,9692824}{k+6} - \frac{305,4467705}{k+7} \right), \quad t \in (-c, c) \text{ con } c \in \mathbb{R}^+ \cup \{+\infty\}. \end{aligned}$$

Función de distribución

Ahora, por la proposición 2.10, la función de distribución $F_X : \mathbb{R} \rightarrow [0, 1]$ de la v.a. X se define a trozos de la forma

$$F_X(x) = \begin{cases} 0, & x < 0 \\ \hat{F}(x), & 0 \leq x < 1, \\ 1, & x \geq 1 \end{cases}$$

donde

$$\begin{aligned} \hat{F}(x) &:= \sum_{i=0}^6 \frac{a_i}{i+1} x^{i+1} = 0,1035292111x + 2,008067609x^2 - 18,40548757x^3 \\ &\quad + 97,26269670x^4 - 185,4951001x^5 + 149,1615471x^6 - 43,63525293x^7. \end{aligned}$$

Su representación gráfica viene dada por la Figura 2.4.

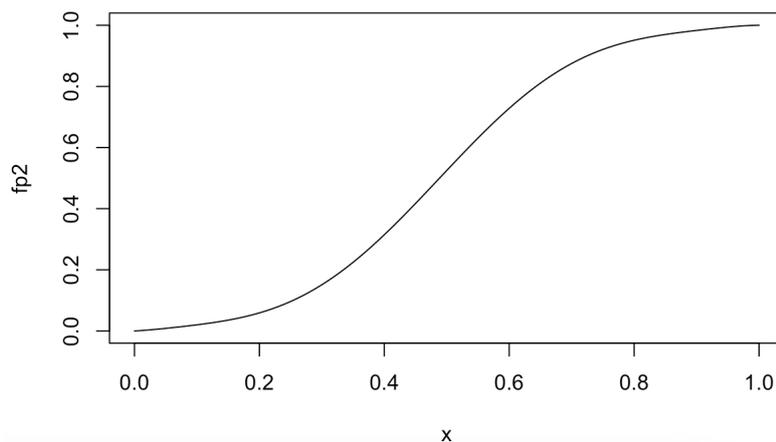


Figura 2.4: Función de distribución de tipo MOP

Mediana

Aplicando la proposición 2.11, se deduce que la mediana de la v.a. X es la solución en $[0, 1]$ de la ecuación

$$\sum_{i=0}^6 \frac{a_i}{i+1} x^{i+1} = \frac{1}{2}.$$

Para obtenerla, comenzamos resolviendo dicha ecuación en \mathbb{C} . Como

$$\sum_{i=0}^6 \frac{a_i}{i+1} x^{i+1} = \frac{1}{2} \Leftrightarrow 0,1035292111 x + 2,008067609 x^2 - 18,40548757 x^3$$

$$+ 97,26269670 x^4 - 185,4951001 x^5 + 149,1615471 x^6 - 43,63525293 x^7 = \frac{1}{2},$$

se obtiene una ecuación polinómica en la variable x de grado 7. Por el teorema fundamental del álgebra, esta ecuación tiene exactamente 7 soluciones complejas, dadas por

$$x_1 = -0,2082219919,$$

$$x_2 = 0,01081584061 - 0,3044549667i, \quad x_3 = 0,01081584061 + 0,3044549667i,$$

$$x_4 = 0,4884589283,$$

$$x_5 = 0,9652284343 - 0,3030189672i, \quad x_6 = 0,9652284343 + 0,3030189672i,$$

$$x_7 = 1,186046978.$$

Observemos que el hecho de que cada raíz compleja vaya acompañada de su conjugada justifica que el polinomio anterior tenga coeficientes reales.

Finalmente, entre dichas soluciones, la mediana será aquella que pertenezca al conjunto de los números reales y, en particular, al intervalo $[0, 1]$. Por lo tanto, se sigue que la mediana es $x_4 = 0,4884589283$.

Moda

Por último, gracias a la proposición 2.12, se tiene que la moda de la v.a. X es la solución de la ecuación

$$f_X(x) = \max\{f_X(0), f_X(\tilde{x}), f_X(1)\},$$

donde \tilde{x} es la solución en $]0, 1[$ del sistema de ecuaciones

$$\begin{cases} \sum_{i=1}^n a_i i x^{i-1} = 0 \\ \sum_{i=2}^n a_i i(i-1)x^{i-2} < 0 \end{cases}$$

Para su cálculo debemos obtener, en primer lugar, los extremos relativos de f_X . Para ello, calcularemos su derivada y la igualaremos a cero. Como, por la proposición 2.12, se tiene que

$$f'_X(x) = \sum_{i=1}^6 a_i i x^{i-1} = 4,016135217 - 110,4329254x + 1167,152360x^2 - 3709,902002x^3 + 4474,846412x^4 - 1832,680623x^5,$$

de $f'_X(x) = 0$ se obtiene una ecuación polinómica en x de grado 5, la cual tendrá, de nuevo, por el teorema fundamental del álgebra, 5 soluciones complejas, que son las siguientes:

$$x_1 = 0,06238409809 - 0,03835946549i, \quad x_2 = 0,06238409809 + 0,03835946549i,$$

$$x_3 = 0,4878511579,$$

$$x_4 = 0,9145376315 - 0,03414808009i, \quad x_5 = 0,9145376315 + 0,03414808009i.$$

Entre ellas, nos quedaremos únicamente con los máximos relativos, esto es, con aquellas soluciones que pertenezcan al intervalo $]0, 1[$ (en este caso, solamente x_3) y verifiquen, además, la desigualdad $f''_X(x) < 0$. Como, aplicando la proposición 2.12, tenemos que

$$f''_X(x) = \sum_{i=2}^6 a_i i(i-1)x^{i-2} = -110,4329254 + 2334,304721x - 11129,70600x^2 + 17899,38565x^3 - 9163,403115x^4,$$

se sigue que

$$f''_X(x_3) = -61,28089460 < 0,$$

con lo cual $x_3 = 0,4878511579$ verifica ambas condiciones y, por ello, es un máximo relativo de f_X .

Finalmente, para obtener el máximo absoluto de f_X , es decir, la moda de X , resolvemos la ecuación

$$f_X(x) = \max\{f_X(0), f_X(x_3), f_X(1)\}.$$

Como, de nuevo haciendo uso de la proposición 2.12, se tiene que

$$f_X(0) = a_0 = 0,1035292111,$$

$$f_X(x_3) = 2,171313958,$$

y

$$f_X(1) = \sum_{i=0}^6 a_i = 0,001000008100,$$

se deduce que la moda de la variable aleatoria X es $x_3 = 0,4878511579$. Este hecho se puede apreciar de forma gráfica en la Figura 2.3.

Métodos de estimación de parámetros a partir de los datos

En este capítulo nos centraremos en la estimación de densidades univariantes de tipo mixtura de funciones base truncadas, las cuales corresponden a variables asociadas a nodos raíz en una red bayesiana.

El problema de estimar una densidad univariante de tipo MoTBF a partir de datos se puede dividir en tres tareas:

1. Dividir el soporte de la variable en intervalos disjuntos.
2. Determinar el número de funciones base para cada intervalo.
3. Estimar los parámetros para un intervalo dado y un número fijo de términos.

En este trabajo nos centraremos en la estimación de los parámetros, suponiendo que el soporte de la variable es siempre el intervalo $[0, 1]$ (sin particiones) y que el número de términos exponenciales (en el caso MTE) o términos del polinomio (en el caso MOP) es fijo.

Cabe mencionar que la forma más usual de determinar el número de términos es mediante un proceso iterativo que busca optimizar la verosimilitud. Este procedimiento consiste en ir probando con distintos números de términos intentando maximizar la verosimilitud. Se empieza con cierto número de términos y se van añadiendo términos y reestimando siempre que se incremente la verosimilitud. No obstante, no entraremos en mayor detalle con respecto a esta cuestión en la presente memoria.

Para abordar el problema de la estimación de parámetros en distribuciones de tipo MoTBF a partir de grandes volúmenes de datos recurriremos no solo a métodos populares, como el método de máxima verosimilitud, sino también a métodos «no estándar», como mínimos cuadrados. Este hecho viene motivado por lo siguiente.

Teorema 3.1. *Ninguna distribución de tipo MoTBF pertenece a la familia exponencial.*¹

Como consecuencia del teorema 3.1, se deduce que no es posible obtener estadísticos suficientes de dimensión fija menor que la muestra (para cualquier tamaño muestral) para los parámetros de las distribuciones de tipo MoTBF, lo que obliga a realizar nuevamente todos los cálculos si se incorporan nuevos datos a la muestra. De esta forma, el problema de estimación de parámetros se hace intratable para modelos de gran tamaño. Por este motivo, en el contexto de grandes volúmenes de datos, necesitamos

¹Una familia de f.d.p. o de f.m.p. $\mathcal{F}_\theta = \{f(x; \theta) | \theta \in \Theta \subseteq \mathbb{R}^p\}$ pertenece a la familia exponencial p -paramétrica si

$$f(x; \theta) = H(x) C(\theta) \exp \left\{ \sum_{i=1}^p T_i(x) Q_i(\theta) \right\},$$

donde $H(x)$ y $T_i(x)$, $i = 1, \dots, p$, son funciones reales de la observación x , $C(\theta)$ y $Q_i(\theta)$, $i = 1, \dots, p$, son funciones reales de los parámetros y el soporte de la distribución no depende de ningún parámetro desconocido.

un método de estimación que sea lo más eficiente posible. De ahí, el interés por considerar métodos de estimación de parámetros «no estándar», como el método de mínimos cuadrados, sobre todo teniendo en cuenta que, como veremos más adelante, no se pueden obtener de forma cerrada las soluciones de las ecuaciones de verosimilitud. No obstante, el método de mínimos cuadrados también presenta problemas de eficiencia, principalmente debido a las restricciones para garantizar que la función ajustada es realmente una función de densidad o de distribución.

3.1 Método de máxima verosimilitud

Nos planteamos, en primer lugar, la utilización de estimadores de máxima verosimilitud para estimar los parámetros desconocidos en distribuciones de tipo MoTBF. Este método consiste en elegir como estimación el valor del parámetro que maximiza la función de verosimilitud. Intuitivamente, se basa en suponer que ocurre lo que es «más verosímil» que ocurra.

Sea X_1, \dots, X_n una muestra aleatoria de tamaño n cuya distribución viene dada por una familia paramétrica $\mathcal{F}_\theta = \{f(x; \theta) | \theta \in \Theta\}$. A partir de este momento, denotaremos por \mathbf{X} a una muestra aleatoria de tamaño n .

Puesto que la muestra es aleatoria (variables aleatorias independientes e idénticamente distribuidas), la función de densidad o masa de probabilidad adopta la siguiente expresión:

$$f(\mathbf{x}; \theta) = \prod_{i=1}^n f(x_i; \theta).$$

Recordemos, en primer lugar, la definición de la función de verosimilitud.

Definición 3.1. Sea $\mathbf{X} \sim f(\mathbf{x}; \theta)$ una muestra aleatoria de tamaño n , con $\theta \in \Theta$. Dada una observación muestral \mathbf{x} , se denomina **función de verosimilitud**, y se denota por $L(\theta | \mathbf{x})$, a la función de densidad (o función de masa de probabilidad) conjunta de la muestra considerada como función del parámetro, es decir,

$$L(\theta | \mathbf{x}) = f(\mathbf{x}; \theta) = \prod_{i=1}^n f(x_i; \theta), \quad \theta \in \Theta.$$

Definición 3.2. Sea $\mathbf{X} \sim f(\mathbf{x}; \theta)$ una muestra aleatoria, con $\theta \in \Theta$. Se define el **estimador de máxima verosimilitud (EMV)** para el parámetro θ como el valor $\hat{\theta}$ que maximiza la función de verosimilitud, esto es,

$$L(\hat{\theta} | \mathbf{x}) = \sup_{\theta \in \Theta} L(\theta | \mathbf{x}).$$

Normalmente, si la función de verosimilitud es derivable, se utiliza el procedimiento general de maximización de una función para la obtención del EMV. De esta forma, podemos resumir dicho proceso en los siguientes pasos:

1. Obtener la función de verosimilitud $L(\theta | \mathbf{x})$ de θ .
2. Calcular su primera derivada con respecto a θ .

3. Igualarla a cero y resolver la ecuación resultante.
4. Comprobar que la solución de dicha ecuación es un máximo, es decir, calcular la segunda derivada de la función con respecto a θ y verificar que, evaluada en el valor hallado, es menor que cero.

Habitualmente los cálculos se simplifican si, en lugar de maximizar la función de verosimilitud de θ , consideramos su logaritmo, ya que calcular la derivada de un producto puede llegar a ser un procedimiento tedioso. Observemos que ambas funciones $L(\theta|\mathbf{x})$ y $\log L(\theta|\mathbf{x})$ alcanzan su máximo en el mismo punto, puesto que la función logarítmica es monótona creciente. En este caso, el proceso de obtención del EMV sería análogo al descrito anteriormente, añadiendo entre los pasos 1 y 2 el cálculo del logaritmo de la función de verosimilitud, $\log L(\theta|\mathbf{x})$.

En ocasiones, dicho estimador se puede obtener explícitamente como una función de los datos observados: x_1, \dots, x_n . Sin embargo, a menudo hay que recurrir a métodos de optimización numérica para encontrar el máximo de la función de verosimilitud. Además, el estimador de máxima verosimilitud puede no existir o no ser único.

Por otra parte, cabe mencionar que θ puede ser, o bien un único parámetro, o bien un vector de parámetros (como es el caso de las distribuciones de tipo MoTBF que estudiaremos a continuación). En este último caso, se utiliza la notación $\boldsymbol{\theta} = (\theta_1, \dots, \theta_p)$ y se verifica el siguiente resultado.

Proposición 3.1. *Sea $X \sim f(\mathbf{x}; \boldsymbol{\theta})$ una muestra aleatoria, con $\boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^p$ y supongamos que $L(\boldsymbol{\theta}|\mathbf{x})$ es una función diferenciable con respecto a $\boldsymbol{\theta} = (\theta_1, \dots, \theta_p)$. Entonces, si existe el estimador de máxima verosimilitud, $\hat{\boldsymbol{\theta}}$, este satisface las denominadas **ecuaciones de verosimilitud**, dadas por*

$$\frac{\partial}{\partial \theta_i} \log L(\boldsymbol{\theta}|\mathbf{x}) = 0, \quad i = 1, \dots, p.$$

Distribución de tipo MTE univariante

Consideraremos, por simplicidad, una mezcla de exponenciales truncadas con un término independiente y un único término exponencial. Sea \mathbf{X} una muestra aleatoria de tamaño n extraída de una población X modelizada por la función de densidad MTE

$$f(x; \boldsymbol{\theta}) = a_0 + a_1 e^{b_1 x}, \quad x \in [0, 1], \quad (3.1)$$

donde $a_0, a_1, b_1 \in \mathbb{R}$. La función (3.1) tiene 3 parámetros, a saber $\boldsymbol{\theta} = (a_0, a_1, b_1)$.

En primer lugar, calculamos la función de verosimilitud de $\boldsymbol{\theta}$:

$$L(\boldsymbol{\theta}|\mathbf{x}) = \prod_{i=1}^n f(x_i; \boldsymbol{\theta}) = \prod_{i=1}^n (a_0 + a_1 e^{b_1 x_i}). \quad (3.2)$$

Para encontrar una solución de forma cerrada para los estimadores de máxima verosimilitud, necesitamos derivar la función (3.2) respecto de los diferentes parámetros

e igualar los respectivos resultados a cero. Para ilustrarlo, derivamos, por ejemplo, con respecto a a_0 y obtenemos:

$$\begin{aligned} \frac{\partial}{\partial a_0} L(\boldsymbol{\theta}|\mathbf{x}) &= \frac{\partial}{\partial a_0} \prod_{i=1}^n f(x_i; \boldsymbol{\theta}) = \sum_{i=1}^n \left\{ \left(\frac{\partial}{\partial a_0} f(x_i; \boldsymbol{\theta}) \right) \prod_{j \in \{1, \dots, n\}, j \neq i} f(x_j; \boldsymbol{\theta}) \right\} = \\ &= \sum_{i=1}^n \left\{ \prod_{j \in \{1, \dots, n\}, j \neq i} (a_0 + a_1 e^{b_1 x_j}) \right\}. \end{aligned}$$

Por lo tanto, debemos resolver la siguiente ecuación:

$$\sum_{i=1}^n \left\{ \prod_{j \in \{1, \dots, n\}, j \neq i} (a_0 + a_1 e^{b_1 x_j}) \right\} = 0. \quad (3.3)$$

Sin embargo, la ecuación (3.3) es no lineal en los parámetros desconocidos $\boldsymbol{\theta} = (a_0, a_1, b_1)$. Luego, alternativamente, se podría considerar maximizar el logaritmo de la función de verosimilitud. Para ello, calculamos primero dicho logaritmo:

$$\log L(\boldsymbol{\theta}|\mathbf{x}) = \sum_{i=1}^n \log (a_0 + a_1 e^{b_1 x_i}). \quad (3.4)$$

A continuación, derivamos la función (3.4) con respecto a a_0 :

$$\frac{\partial}{\partial a_0} \log L(\boldsymbol{\theta}|\mathbf{x}) = \sum_{i=1}^n \frac{1}{a_0 + a_1 e^{b_1 x_i}}$$

y, por último, igualamos a cero, obteniendo la siguiente ecuación:

$$\sum_{i=1}^n \frac{1}{a_0 + a_1 e^{b_1 x_i}} = 0.$$

De aquí, deducimos que no se puede obtener el EMV de a_0 de forma cerrada.

Análogamente, derivando con respecto a a_1 y b_1 en lugar de a_0 , obtenemos las ecuaciones

$$\frac{\partial}{\partial a_1} \log L(\boldsymbol{\theta}|\mathbf{x}) = \sum_{i=1}^n \frac{e^{b_1 x_i}}{a_0 + a_1 e^{b_1 x_i}} = 0$$

y

$$\frac{\partial}{\partial b_1} \log L(\boldsymbol{\theta}|\mathbf{x}) = \sum_{i=1}^n \frac{a_1 x_i e^{b_1 x_i}}{a_0 + a_1 e^{b_1 x_i}} = 0,$$

respectivamente; de las cuales se extrae la misma conclusión anterior, a saber: no es posible obtener el EMV de a_1 ni de b_1 de forma cerrada. En definitiva, no se puede calcular el estimador de máxima verosimilitud de $\boldsymbol{\theta} = (a_0, a_1, b_1)$ de forma exacta.

Distribución de tipo MOP univariante

En este caso, consideraremos una distribución de tipo MOP de grado 2. Sea \mathbf{X} una muestra aleatoria de tamaño n extraída de una población X cuyo modelo probabilístico viene dado por la siguiente función de densidad

$$f(x; \boldsymbol{\theta}) = a_0 + a_1 x + a_2 x^2, \quad x \in [0, 1], \quad (3.5)$$

donde $a_0, a_1, a_2 \in \mathbb{R}$. El objetivo es estimar el vector de parámetros, $\boldsymbol{\theta} = (a_0, a_1, a_2)$, de la función (3.5).

Comenzamos calculando la función de verosimilitud de $\boldsymbol{\theta}$:

$$L(\boldsymbol{\theta}|\mathbf{x}) = \prod_{i=1}^n f(x_i; \boldsymbol{\theta}) = \prod_{i=1}^n (a_0 + a_1 x_i + a_2 x_i^2). \quad (3.6)$$

Al igual que en el apartado anterior, debemos derivar la función (3.6) respecto de los diferentes parámetros e igualar los resultados a cero, respectivamente. Por ejemplo, si derivamos dicha función con respecto a a_2 , obtenemos:

$$\begin{aligned} \frac{\partial}{\partial a_2} L(\boldsymbol{\theta}|\mathbf{x}) &= \frac{\partial}{\partial a_2} \prod_{i=1}^n f(x_i; \boldsymbol{\theta}) = \sum_{i=1}^n \left\{ \left(\frac{\partial}{\partial a_2} f(x_i; \boldsymbol{\theta}) \right) \prod_{j \in \{1, \dots, n\}, j \neq i} f(x_j; \boldsymbol{\theta}) \right\} = \\ &= \sum_{i=1}^n \left\{ x_i^2 \prod_{j \in \{1, \dots, n\}, j \neq i} (a_0 + a_1 x_j + a_2 x_j^2) \right\}. \end{aligned}$$

Así, la ecuación que debemos resolver es

$$\sum_{i=1}^n \left\{ x_i^2 \prod_{j \in \{1, \dots, n\}, j \neq i} (a_0 + a_1 x_j + a_2 x_j^2) \right\} = 0. \quad (3.7)$$

No obstante, la ecuación (3.7) es, de nuevo, no lineal en los parámetros desconocidos $\boldsymbol{\theta} = (a_0, a_1, a_2)$. Por tanto, de forma alternativa, procedemos a maximizar el logaritmo de la función de verosimilitud. Calculamos, para empezar, dicho logaritmo:

$$\log L(\boldsymbol{\theta}|\mathbf{x}) = \sum_{i=1}^n \log (a_0 + a_1 x_i + a_2 x_i^2). \quad (3.8)$$

Ahora, derivamos la función (3.8) con respecto a a_2 :

$$\frac{\partial}{\partial a_2} \log L(\boldsymbol{\theta}|\mathbf{x}) = \sum_{i=1}^n \frac{x_i^2}{a_0 + a_1 x_i + a_2 x_i^2}$$

y la igualamos a cero, obteniendo la siguiente ecuación:

$$\sum_{i=1}^n \frac{x_i^2}{a_0 + a_1 x_i + a_2 x_i^2} = 0. \quad (3.9)$$

Finalmente, de la ecuación (3.9) se deduce que no se puede obtener el EMV de a_2 de forma exacta.

Semejantemente, si derivamos con respecto a a_0 y a_1 , en lugar de a_2 , se obtienen las ecuaciones

$$\frac{\partial}{\partial a_0} \log L(\boldsymbol{\theta}|\mathbf{x}) = \sum_{i=1}^n \frac{1}{a_0 + a_1 x_i + a_2 x_i^2} = 0$$

y

$$\frac{\partial}{\partial a_1} \log L(\boldsymbol{\theta}|\mathbf{x}) = \sum_{i=1}^n \frac{x_i}{a_0 + a_1 x_i + a_2 x_i^2} = 0,$$

respectivamente. De ellas deducimos que tampoco es posible calcular el estimador de máxima verosimilitud de a_0 o a_1 de forma cerrada ni, en conclusión, el de $\boldsymbol{\theta} = (a_0, a_1, a_2)$.

3.2 Método de Fisher Scoring

En la sección 3.1 se concluye que no es posible resolver las ecuaciones de máxima verosimilitud de forma cerrada. Por este motivo, en esta nueva sección, estudiaremos un método numérico para optimizar los parámetros de forma iterativa.

El algoritmo de *Fisher Scoring* es una variante del método de Newton utilizada en estadística para resolver numéricamente ecuaciones de máxima verosimilitud. Esto es, es un método iterativo para determinar de forma numérica el estimador de máxima verosimilitud. Cabe mencionar que dicho algoritmo fue denominado así en honor al estadístico y biólogo Ronald Fisher.

Conceptos y resultados previos

Definición 3.3. Sea $\mathbf{X} \sim f(\mathbf{x};\boldsymbol{\theta})$ una muestra aleatoria de tamaño n , con $\boldsymbol{\theta} \in \Theta$. Entonces, dada una realización de la muestra \mathbf{x} , se define la **función de log-verosimilitud**, denotada por $\ell(\boldsymbol{\theta}|\mathbf{x})$, como el logaritmo de la función de verosimilitud, esto es,

$$\ell(\boldsymbol{\theta}|\mathbf{x}) = \log L(\boldsymbol{\theta}|\mathbf{x}).$$

Puesto que la muestra es aleatoria (variables independientes e idénticamente distribuidas), la función de log-verosimilitud adopta la siguiente expresión:

$$\ell(\boldsymbol{\theta}|\mathbf{x}) = \sum_{i=1}^n \log f(x_i; \boldsymbol{\theta}).$$

Definición 3.4. Sea $\mathbf{X} \sim f(\mathbf{x};\boldsymbol{\theta})$ una muestra aleatoria, con $\boldsymbol{\theta} \in \Theta$. Definimos el **score** de la muestra como la cantidad

$$S(\mathbf{X};\boldsymbol{\theta}) = \frac{\partial}{\partial \boldsymbol{\theta}} \log f(\mathbf{X};\boldsymbol{\theta}).$$

En el caso p -paramétrico, esto es, si $\mathbf{X} \sim f(\mathbf{x}; \theta)$ es una muestra aleatoria, con $\theta \in \Theta \subseteq \mathbb{R}^p$, se define el **score** de la muestra como el gradiente de la función de log-verosimilitud con respecto al vector de parámetros θ , es decir,

$$S(\mathbf{X}; \theta) = \nabla \log f(\mathbf{X}; \theta) = \left(\frac{\partial}{\partial \theta_1} \log f(\mathbf{X}; \theta), \dots, \frac{\partial}{\partial \theta_p} \log f(\mathbf{X}; \theta) \right).$$

Definición 3.5. Sea $\mathbf{X} \sim f(\mathbf{x}; \theta)$ una muestra aleatoria, con $\theta \in \Theta$. Se denomina **función de información de Fisher** a

$$\mathcal{I}_{\mathbf{X}}(\theta) = \mathbb{E}[S(\mathbf{X}; \theta)^2] = \mathbb{E}\left[\left(\frac{\partial}{\partial \theta} \log f(\mathbf{X}; \theta)\right)^2\right].$$

Lema 3.1. Bajo las condiciones de regularidad de Cramér-Rao ² se verifica que la esperanza del score es nula para cualquier valor de $\theta \in \Theta$, es decir,

$$\mathbb{E}[S(\mathbf{X}; \theta)] = 0, \quad \forall \theta \in \Theta.$$

Demostración:

Sea $\theta \in \Theta$. Entonces,

$$\begin{aligned} \mathbb{E}[S(\mathbf{X}; \theta)] &= \int_{\mathbb{R}^n} \left(\frac{\partial}{\partial \theta} \log f(\mathbf{x}; \theta) \right) f(\mathbf{x}; \theta) d\mathbf{x} = \int_{\mathbb{R}^n} \frac{\frac{\partial}{\partial \theta} f(\mathbf{x}; \theta)}{f(\mathbf{x}; \theta)} f(\mathbf{x}; \theta) d\mathbf{x} = \\ &= \int_{\mathbb{R}^n} \frac{\partial}{\partial \theta} f(\mathbf{x}; \theta) d\mathbf{x} = \frac{\partial}{\partial \theta} \int_{\mathbb{R}^n} f(\mathbf{x}; \theta) d\mathbf{x} = 0, \end{aligned}$$

donde en la última igualdad hemos utilizado que

$$\int_{\mathbb{R}^n} f(\mathbf{x}; \theta) d\mathbf{x} = 1, \quad \forall \theta \in \Theta.$$

■

De este lema y de las propiedades de la varianza se puede deducir fácilmente que, bajo las condiciones de regularidad de Cramér-Rao, la función de información de Fisher es realmente la varianza de la función score, esto es,

$$\mathcal{I}_{\mathbf{X}}(\theta) = \text{Var}[S(\mathbf{X}; \theta)].$$

²Una familia de f.d.p. o de f.m.p. $\mathcal{F}_{\theta} = \{f(\mathbf{x}; \theta) | \theta \in \Theta\}$ es **regular en el sentido de Cramér-Rao** si verifica las siguientes propiedades:

1. Su soporte, $\{x \in \mathcal{X} | f(x; \theta) > 0\}$, no depende de θ .
2. Θ es \mathbb{R} o un intervalo de \mathbb{R} .
3. $\frac{\partial^i}{\partial \theta^i} f(\mathbf{x}; \theta)$, $i = 1, 2$, existe y es finita para todo $\theta \in \Theta$.
4. $\int_{\mathbb{R}} \left| \frac{\partial^i}{\partial \theta^i} f(\mathbf{x}; \theta) \right| d\mathbf{x}$, $i = 1, 2$, existe y es finita para todo $\theta \in \Theta$.
5. $\mathbb{E}\left[\left(\frac{\partial}{\partial \theta} \log f(\mathbf{X}; \theta)\right)^2\right]$ existe y es finita para todo $\theta \in \Theta$.

Lema 3.2. *Bajo las condiciones de regularidad de Cramér-Rao se verifica que*

$$\mathbb{E}\left[\left(\frac{\partial}{\partial\theta}\log f(\mathbf{X};\theta)\right)^2\right] = -\mathbb{E}\left[\frac{\partial^2}{\partial\theta^2}\log f(\mathbf{X};\theta)\right].$$

Corolario 3.2.1. *Bajo las condiciones de regularidad de Cramér-Rao, la función de información de Fisher también puede escribirse como*

$$\mathcal{I}_{\mathbf{X}}(\theta) = -\mathbb{E}\left[\frac{\partial^2}{\partial\theta^2}\log f(\mathbf{X};\theta)\right].$$

Lema 3.3. *Sea \mathbf{X} un vector formado por n variables aleatorias independientes regulares Cramér-Rao. Entonces,*

$$\mathcal{I}_{\mathbf{X}}(\theta) = \sum_{i=1}^n \mathcal{I}_{X_i}(\theta).$$

Demostración:

Sea $\theta \in \Theta$. Entonces,

$$\begin{aligned} \mathcal{I}_{\mathbf{X}}(\theta) &= \mathbb{E}[S(\mathbf{X};\theta)^2] = \mathbb{E}\left[\left(\frac{\partial}{\partial\theta}\log f(\mathbf{X};\theta)\right)^2\right] = \mathbb{E}\left[\left(\frac{\partial}{\partial\theta}\log \prod_{i=1}^n f(X_i;\theta)\right)^2\right] = \\ &= \mathbb{E}\left[\left(\frac{\partial}{\partial\theta}\sum_{i=1}^n \log f(X_i;\theta)\right)^2\right] = \mathbb{E}\left[\left(\sum_{i=1}^n \frac{\partial}{\partial\theta}\log f(X_i;\theta)\right)^2\right] = \\ &= \mathbb{E}\left[\sum_{i=1}^n \left(\frac{\partial}{\partial\theta}\log f(X_i;\theta)\right)^2 + 2\sum_{i<j}^n \left(\frac{\partial}{\partial\theta}\log f(X_i;\theta)\right)\left(\frac{\partial}{\partial\theta}\log f(X_j;\theta)\right)\right] = \\ &= \sum_{i=1}^n \mathbb{E}\left[\left(\frac{\partial}{\partial\theta}\log f(X_i;\theta)\right)^2\right] + 2\sum_{i<j}^n \mathbb{E}\left[\left(\frac{\partial}{\partial\theta}\log f(X_i;\theta)\right)\left(\frac{\partial}{\partial\theta}\log f(X_j;\theta)\right)\right]. \end{aligned}$$

Por hipótesis, las v.a. $X_i, i = 1, \dots, n$ son independientes, con lo cual

$$\mathbb{E}\left[\left(\frac{\partial}{\partial\theta}\log f(X_i;\theta)\right)\left(\frac{\partial}{\partial\theta}\log f(X_j;\theta)\right)\right] = \mathbb{E}\left[\frac{\partial}{\partial\theta}\log f(X_i;\theta)\right]\mathbb{E}\left[\frac{\partial}{\partial\theta}\log f(X_j;\theta)\right], \quad (3.10)$$

para todo $i, j \in \{1, \dots, n\}$ distintos.

Además, por el lema 3.1, se tiene que

$$\mathbb{E}\left[\frac{\partial}{\partial\theta}\log f(X_i;\theta)\right] = \mathbb{E}[S(X_i;\theta)] = 0, \quad \forall i = 1, \dots, n. \quad (3.11)$$

Luego, de (3.10) y (3.11), se sigue que

$$\mathcal{I}_{\mathbf{X}}(\theta) = \sum_{i=1}^n \mathbb{E}\left[\left(\frac{\partial}{\partial\theta}\log f(X_i;\theta)\right)^2\right] = \sum_{i=1}^n \mathbb{E}[S(X_i;\theta)^2] = \sum_{i=1}^n \mathcal{I}_{X_i}(\theta).$$

■

Corolario 3.3.1. Sea \mathbf{X} un vector formado por n variables aleatorias i.i.d. como X regular Cramér-Rao. Entonces,

$$\mathcal{I}_{\mathbf{X}}(\theta) = n\mathcal{I}_X(\theta).$$

Definición 3.6. Sea $\mathbf{X} \sim f(\mathbf{x}; \theta)$ una muestra aleatoria, con $\theta \in \Theta \subseteq \mathbb{R}^p$. Se denomina **matriz de información observada** en θ^* al opuesto de la matriz Hessiana de la log-verosimilitud, es decir,

$$\mathcal{J}_{\mathbf{X}}(\theta^*) = -\nabla\nabla^\top \log f(\mathbf{X}; \theta) \Big|_{\theta=\theta^*}$$

o, lo que es lo mismo,

$$[\mathcal{J}_{\mathbf{X}}(\theta^*)]_{i,j} = -\left[\frac{\partial^2}{\partial\theta_i \partial\theta_j} \log f(\mathbf{X}; \theta) \Big|_{\theta=\theta^*} \right].$$

En este caso, $\mathcal{J}_{\mathbf{X}}(\theta^*)$ es una matriz cuadrada de orden p .

Definición 3.7. Sea $\mathbf{X} \sim f(\mathbf{x}; \theta)$ una muestra aleatoria, con $\theta \in \Theta \subseteq \mathbb{R}^p$. Se denomina **matriz de información de Fisher** a

$$[\mathcal{I}_{\mathbf{X}}(\theta)]_{i,j} = \mathbb{E} \left[\left(\frac{\partial}{\partial\theta_i} \log f(\mathbf{X}; \theta) \right) \left(\frac{\partial}{\partial\theta_j} \log f(\mathbf{X}; \theta) \right) \right].$$

Bajo las condiciones de regularidad de Cramér-Rao, la matriz de información de Fisher también puede escribirse como

$$[\mathcal{I}_{\mathbf{X}}(\theta)]_{i,j} = -\mathbb{E} \left[\frac{\partial^2}{\partial\theta_i \partial\theta_j} \log f(\mathbf{X}; \theta) \right].$$

En este caso, $\mathcal{I}_{\mathbf{X}}(\theta)$ es una matriz de orden p .

De las definiciones 3.6 y 3.7 se deduce el siguiente resultado.

Proposición 3.2. Sea $\mathbf{X} \sim f(\mathbf{x}; \theta)$ una muestra aleatoria, con $\theta \in \Theta \subseteq \mathbb{R}^p$. La matriz de información de Fisher es el valor esperado de la información observada, es decir,

$$\mathcal{I}_{\mathbf{X}}(\theta) = \mathbb{E}[\mathcal{J}_{\mathbf{X}}(\theta)].$$

Dicho de otra forma, la información observada se trata de una versión de la información de Fisher basada en la muestra.

Algoritmo de Fisher Scoring

Veamos ya en qué consiste el método de *Fisher Scoring*, sin entrar en los detalles de la demostración de su convergencia.

Sean X_1, \dots, X_n variables aleatorias independientes e idénticamente distribuidas según una función de densidad $f(x; \theta)$ dos veces diferenciable, con $\theta \in \Theta \subseteq \mathbb{R}^p$ desconocido. Nuestro objetivo es aproximar el estimador de máxima verosimilitud (EMV) $\hat{\theta}$ de θ .

Para comenzar, sea θ_0 el valor inicial del vector de parámetros, es decir, un punto de partida para nuestro algoritmo, y consideremos el desarrollo en serie de Taylor de la función score, $S(\mathbf{X};\theta)$, alrededor de θ_0 :

$$S(\mathbf{X};\theta) \approx S(\mathbf{X};\theta_0) - \mathcal{J}_{\mathbf{X}}(\theta_0)(\theta - \theta_0),$$

donde $\mathcal{J}_{\mathbf{X}}(\theta_0)$ es la matriz de información observada en θ_0 . Entonces, estableciendo $\theta = \hat{\theta}$, usando que $S(\mathbf{X};\hat{\theta}) = 0$ (por la proposición 3.1) y reorganizando términos, se puede demostrar que

$$\hat{\theta} \approx \theta_0 + \mathcal{J}_{\mathbf{X}}^{-1}(\theta_0) S(\mathbf{X};\theta_0).$$

Por lo tanto, podemos utilizar el algoritmo

$$\theta_{m+1} = \theta_m + \mathcal{J}_{\mathbf{X}}^{-1}(\theta_m) S(\mathbf{X};\theta_m),$$

donde

$$S(\mathbf{X};\theta_m) \approx S(\mathbf{X};\theta_0) - \mathcal{J}_{\mathbf{X}}(\theta_0)(\theta_m - \theta_0),$$

para aproximarnos de forma iterativa al EMV, puesto que, bajo las condiciones de regularidad de Cramér-Rao, se puede probar que $\theta_m \rightarrow \hat{\theta}$.

En la práctica, $\mathcal{J}_{\mathbf{X}}(\theta)$ es reemplazado generalmente por $\mathcal{I}_{\mathbf{X}}(\theta) = E[\mathcal{J}_{\mathbf{X}}(\theta)]$, la matriz de información de Fisher, dando lugar al algoritmo de *Fisher Scoring*:

$$\theta_{m+1} = \theta_m + \mathcal{I}_{\mathbf{X}}^{-1}(\theta_m) S(\mathbf{X};\theta_m),$$

donde

$$S(\mathbf{X};\theta_m) \approx S(\mathbf{X};\theta_0) - \mathcal{I}_{\mathbf{X}}(\theta_0)(\theta_m - \theta_0).$$

A continuación, estudiaremos la aplicación del algoritmo de *Fisher Scoring* para la estimación de parámetros en distribuciones de tipo MoTBF, en concreto, en distribuciones univariantes de tipo mixtura de polinomios.

Distribución de tipo MOP univariante

En esta ocasión, al igual que en el método de máxima verosimilitud, consideraremos una distribución de tipo MOP de grado 2, es decir, con tres términos. Sea \mathbf{X} una muestra aleatoria de tamaño n extraída de una población X cuyo modelo probabilístico viene dado por la función de densidad MOP

$$f(x; a_0, a_1, a_2) = a_0 + a_1 x + a_2 x^2, \quad x \in [0, 1], \quad (3.12)$$

donde $a_0, a_1, a_2 \in \mathbb{R}$. Nuestro objetivo es calcular de forma numérica el estimador de máxima verosimilitud, $\hat{\theta}$, del vector de parámetros, $\theta = (a_0, a_1, a_2)$, de la función (3.12) mediante el algoritmo iterativo de *Fisher Scoring*.

En este caso, el vector de parámetros que queremos estimar tiene dimensión 3. Luego, el algoritmo de *Fisher Scoring* aplicado a dicho vector θ quedaría de la siguiente

forma:

$$\begin{aligned} (a_0, a_1, a_2)_{m+1} &= (a_0, a_1, a_2)_m + \mathcal{I}_X^{-1}(a_{0_m}, a_{1_m}, a_{2_m}) S(\mathbf{X}; a_{0_m}, a_{1_m}, a_{2_m}) \Leftrightarrow \\ &\Leftrightarrow \begin{pmatrix} a_{0_{m+1}} \\ a_{1_{m+1}} \\ a_{2_{m+1}} \end{pmatrix} = \begin{pmatrix} a_{0_m} \\ a_{1_m} \\ a_{2_m} \end{pmatrix} + \frac{1}{n} \mathcal{I}_X^{-1}(a_{0_m}, a_{1_m}, a_{2_m}) \begin{pmatrix} \frac{\partial}{\partial a_0} \log f(\mathbf{X}; a_0, a_1, a_2) \\ \frac{\partial}{\partial a_1} \log f(\mathbf{X}; a_0, a_1, a_2) \\ \frac{\partial}{\partial a_2} \log f(\mathbf{X}; a_0, a_1, a_2) \end{pmatrix} \Bigg|_{\substack{a_0=a_{0_m} \\ a_1=a_{1_m} \\ a_2=a_{2_m}}} \end{pmatrix}, \quad (3.13) \end{aligned}$$

donde

$$\mathcal{I}_X(a_{0_m}, a_{1_m}, a_{2_m}) = \begin{pmatrix} -\mathbb{E}\left[\frac{\partial^2}{\partial a_0^2} \log f(\mathbf{X}; \boldsymbol{\theta})\right] \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_m} & -\mathbb{E}\left[\frac{\partial^2}{\partial a_0 \partial a_1} \log f(\mathbf{X}; \boldsymbol{\theta})\right] \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_m} & -\mathbb{E}\left[\frac{\partial^2}{\partial a_0 \partial a_2} \log f(\mathbf{X}; \boldsymbol{\theta})\right] \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_m} \\ -\mathbb{E}\left[\frac{\partial^2}{\partial a_1 \partial a_0} \log f(\mathbf{X}; \boldsymbol{\theta})\right] \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_m} & -\mathbb{E}\left[\frac{\partial^2}{\partial a_1^2} \log f(\mathbf{X}; \boldsymbol{\theta})\right] \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_m} & -\mathbb{E}\left[\frac{\partial^2}{\partial a_1 \partial a_2} \log f(\mathbf{X}; \boldsymbol{\theta})\right] \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_m} \\ -\mathbb{E}\left[\frac{\partial^2}{\partial a_2 \partial a_0} \log f(\mathbf{X}; \boldsymbol{\theta})\right] \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_m} & -\mathbb{E}\left[\frac{\partial^2}{\partial a_2 \partial a_1} \log f(\mathbf{X}; \boldsymbol{\theta})\right] \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_m} & -\mathbb{E}\left[\frac{\partial^2}{\partial a_2^2} \log f(\mathbf{X}; \boldsymbol{\theta})\right] \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_m} \end{pmatrix}$$

y, además, hemos hecho uso del corolario 3.3.1 para afirmar que $\mathcal{I}_X(a_{0_m}, a_{1_m}, a_{2_m}) = n \mathcal{I}_X(a_{0_m}, a_{1_m}, a_{2_m})$.

Sin embargo, calcular la matriz inversa presente en la igualdad (3.13) resulta bastante engorroso, debido a la complejidad en la expresión de la esperanza que conforma cada celda de la matriz de información de Fisher $\mathcal{I}_X(a_{0_m}, a_{1_m}, a_{2_m})$.

Por tal motivo, para simplificar el procedimiento, lo que haremos será aplicar el método de *Fisher Scoring* por coordenadas, haciendo uso de la técnica conocida usualmente como *coordinate ascent optimization*. Así, dado $i \in \{0, 1, 2\}$, el algoritmo adopta la siguiente expresión para el parámetro a_i :

$$\begin{aligned} a_{i_{m+1}} &= a_{i_m} + \mathcal{I}_X^{-1}(a_{i_m}) S(\mathbf{X}; a_{i_m}) \Leftrightarrow a_{i_{m+1}} = a_{i_m} + \frac{1}{\mathcal{I}_X(a_{i_m})} S(\mathbf{X}; a_{i_m}) \Leftrightarrow \\ &\Leftrightarrow a_{i_{m+1}} = a_{i_m} + \frac{1}{\mathbb{E}\left[\left(\frac{\partial}{\partial a_i} \log f(\mathbf{X}; a_i)\right)^2\right] \Big|_{a_i=a_{i_m}}} \left(\frac{\partial}{\partial a_i} \log f(\mathbf{X}; a_i) \Big|_{a_i=a_{i_m}} \right). \quad (3.14) \end{aligned}$$

De esta forma, el método se simplifica considerablemente, puesto que la matriz de información de Fisher (matriz cuadrada de orden 3) es reemplazada por la función de información de Fisher, esto es, una matriz de orden 1, cuya inversa se puede obtener fácilmente, tal y como podemos observar en la igualdad (3.14).

No obstante, la función de densidad conjunta de la muestra, presente en el segundo miembro de la ecuación (3.14), no depende únicamente del parámetro a_i como se indica, sino también de los parámetros a_j y a_k , donde $i, j, k \in \{0, 1, 2\}$ tales que $i \neq j \neq k$. Esto es, el algoritmo que aplicaremos tiene, en realidad, la forma

$$a_{i_{m+1}} = a_{i_m} + \frac{1}{\mathbb{E}\left[\left(\frac{\partial}{\partial a_i} \log f(\mathbf{X}; a_0, a_1, a_2)\right)^2\right] \Big|_{\substack{a_0=a_{0_m} \\ a_1=a_{1_m} \\ a_2=a_{2_m}}}} \left(\frac{\partial}{\partial a_i} \log f(\mathbf{X}; a_0, a_1, a_2) \Big|_{\substack{a_0=a_{0_m} \\ a_1=a_{1_m} \\ a_2=a_{2_m}}} \right), \quad (3.15)$$

para todo $i \in \{0, 1, 2\}$.

Se puede observar que en cada iteración se utilizan los valores de los parámetros a_0 , a_1 y a_2 obtenidos en la iteración anterior para actualizar el valor de cada parámetro a_i , $i = 0, 1, 2$, mediante la fórmula (3.15).

Finalmente, cabe resaltar la importancia de normalizar en cada iteración, tal y como se indica en la observación 3.1.

Observación 3.1. *Es esencial tener en cuenta que para que $f(x; a_0, a_1, a_2)$ sea una función de densidad debe cumplir*

$$\int_0^1 f(x; a_0, a_1, a_2) dx = 1. \quad (3.16)$$

Por consiguiente, debemos asegurarnos de que los parámetros a_0 , a_1 y a_2 obtenidos en cada iteración verifiquen la condición (3.16). Para ello, tras cada iteración, debemos sustituir los valores obtenidos mediante el algoritmo: a_0 , a_1 y a_2 , por $\frac{a_0}{K}$, $\frac{a_1}{K}$ y $\frac{a_2}{K}$, donde

$$K := \int_0^1 f(x; a_0, a_1, a_2) dx = \int_0^1 (a_0 + a_1 x + a_2 x^2) dx.$$

Demostración:

Sean a_0 , a_1 y a_2 los parámetros obtenidos por el algoritmo de Fisher Scoring en cierta iteración m y sea

$$K = \int_0^1 (a_0 + a_1 x + a_2 x^2) dx = \left[a_0 x + a_1 \frac{x^2}{2} + a_2 \frac{x^3}{3} \right]_{x=0}^{x=1} = a_0 + \frac{a_1}{2} + \frac{a_2}{3}.$$

Entonces, se tiene lo siguiente:

$$\begin{aligned} 1 &= \frac{1}{K} K = \frac{1}{K} \int_0^1 f(x; a_0, a_1, a_2) dx = \frac{1}{K} \int_0^1 (a_0 + a_1 x + a_2 x^2) dx = \\ &= \int_0^1 \frac{a_0 + a_1 x + a_2 x^2}{K} dx = \int_0^1 \left(\frac{a_0}{K} + \frac{a_1}{K} x + \frac{a_2}{K} x^2 \right) dx = \int_0^1 f\left(x; \frac{a_0}{K}, \frac{a_1}{K}, \frac{a_2}{K}\right) dx. \end{aligned}$$

Por tanto, los parámetros $\frac{a_0}{K}$, $\frac{a_1}{K}$ y $\frac{a_2}{K}$ verifican la igualdad (3.16). ■

Eficiencia del algoritmo

Ahora bien, podemos mejorar la eficiencia del algoritmo implementado anteriormente a través de la sustitución del score de la muestra

$$S(\mathbf{X}; a_{i_m}) = \frac{\partial}{\partial a_i} \log f(\mathbf{X}; a_0, a_1, a_2) \Bigg|_{\substack{a_0=a_{0_m} \\ a_1=a_{1_m} \\ a_2=a_{2_m}}}$$

cuyo valor debe ser nuevamente calculado en cada iteración m , por su desarrollo en serie de Taylor alrededor de a_{i_0} , esto es,

$$\begin{aligned} S(\mathbf{X}; a_{i_0}) - \mathcal{I}_{\mathbf{X}}(a_{i_0})(a_{i_m} - a_{i_0}) = \\ = \left(\frac{\partial}{\partial a_i} \log f(\mathbf{X}; a_0, a_1, a_2) \Bigg|_{\substack{a_0=a_{0_0} \\ a_1=a_{1_0} \\ a_2=a_{2_0}}} \right) - \mathbb{E} \left[\left(\frac{\partial}{\partial a_i} \log f(\mathbf{X}; a_0, a_1, a_2) \right)^2 \Bigg|_{\substack{a_0=a_{0_0} \\ a_1=a_{1_0} \\ a_2=a_{2_0}}} \right] (a_{i_m} - a_{i_0}). \end{aligned}$$

El incremento de la eficiencia en cuanto a número de operaciones requeridas se debe a que los valores de $S(\mathbf{X}, a_{i_0})$ e $\mathcal{I}_{\mathbf{X}}(a_{i_0})$ no dependen de m , es decir, son constantes para todas las iteraciones, ya que dependen exclusivamente del tamaño de la muestra y de los valores iniciales de cada parámetro a_i , $i = 0, 1, 2$.

Cálculos necesarios para la implementación en R

Por último, realizaremos algunos cálculos de cara a la implementación de ambos algoritmos en el software estadístico R³. Dada una observación muestral \mathbf{x} , la función de verosimilitud de $\boldsymbol{\theta} = (a_0, a_1, a_2)$ viene dada por

$$L(a_0, a_1, a_2 | \mathbf{x}) = f(\mathbf{x}; a_0, a_1, a_2) = \prod_{i=1}^n f(x_i; a_0, a_1, a_2) = \prod_{i=1}^n (a_0 + a_1 x_i + a_2 x_i^2).$$

De esta forma, la función de log-verosimilitud puede escribirse como

$$\log L(a_0, a_1, a_2 | \mathbf{x}) = \sum_{i=1}^n \log(a_0 + a_1 x_i + a_2 x_i^2). \quad (3.17)$$

Ahora, derivamos la función (3.17) con respecto a a_i , $i = 0, 1, 2$, para obtener el *score* de la muestra en cada caso:

$$S(\mathbf{x}; a_0) = \frac{\partial}{\partial a_0} \log L(a_0, a_1, a_2 | \mathbf{x}) = \sum_{i=1}^n \frac{1}{a_0 + a_1 x_i + a_2 x_i^2}, \quad (3.18)$$

$$S(\mathbf{x}; a_1) = \frac{\partial}{\partial a_1} \log L(a_0, a_1, a_2 | \mathbf{x}) = \sum_{i=1}^n \frac{x_i}{a_0 + a_1 x_i + a_2 x_i^2}, \quad (3.19)$$

$$S(\mathbf{x}; a_2) = \frac{\partial}{\partial a_2} \log L(a_0, a_1, a_2 | \mathbf{x}) = \sum_{i=1}^n \frac{x_i^2}{a_0 + a_1 x_i + a_2 x_i^2}. \quad (3.20)$$

Es importante destacar que las expresiones (3.18), (3.19) y (3.20) dependen solo de los parámetros a_0 , a_1 y a_2 , tras la sustitución de los x_i por la realización muestral dada.

Por otro lado, de nuevo, en virtud del corolario 3.3.1, se sigue que la función de información de Fisher de a_0 puede expresarse como

$$\begin{aligned} \mathcal{I}_{\mathbf{x}}(a_0) &= n \mathcal{I}_x(a_0) = n \mathbb{E} \left[\left(\frac{\partial}{\partial a_0} \log f(x; a_0, a_1, a_2) \right)^2 \right] = n \mathbb{E} \left[\left(\frac{\partial}{\partial a_0} \log(a_0 + a_1 x + a_2 x^2) \right)^2 \right] = \\ &= n \mathbb{E} \left[\left(\frac{1}{a_0 + a_1 x + a_2 x^2} \right)^2 \right] = n \int_0^1 \left(\frac{1}{a_0 + a_1 x + a_2 x^2} \right)^2 (a_0 + a_1 x + a_2 x^2) dx = \\ &= n \int_0^1 \frac{1}{a_0 + a_1 x + a_2 x^2} dx = n \left[\frac{2 \arctan \frac{a_1 + 2a_2 x}{\sqrt{-a_1^2 + 4a_0 a_2}}}{\sqrt{-a_1^2 + 4a_0 a_2}} \right]_{x=0}^{x=1}. \end{aligned} \quad (3.21)$$

³La implementación en R del algoritmo de *Fisher Scoring* dado por la ecuación (3.15), así como de su versión optimizada, se puede consultar en el apéndice A.

De igual forma, se pueden obtener la función de información de Fisher de a_1 y a_2 , cuyas expresiones vienen dadas por

$$\begin{aligned} \mathcal{I}_x(a_1) &= n \mathbb{E} \left[\left(\frac{\partial}{\partial a_1} \log(a_0 + a_1 x + a_2 x^2) \right)^2 \right] = n \mathbb{E} \left[\left(\frac{x}{a_0 + a_1 x + a_2 x^2} \right)^2 \right] = \\ &= n \int_0^1 \left(\frac{x}{a_0 + a_1 x + a_2 x^2} \right)^2 (a_0 + a_1 x + a_2 x^2) dx = n \int_0^1 \frac{x^2}{a_0 + a_1 x + a_2 x^2} dx = \\ &= n \left[\frac{x}{a_2} + \frac{(a_1^2 - 2a_0 a_2) \arctan \frac{a_1 + 2a_2 x}{\sqrt{-a_1^2 + 4a_0 a_2}}}{a_2^2 \sqrt{-a_1^2 + 4a_0 a_2}} - \frac{a_1 \log(a_0 + a_1 x + a_2 x^2)}{2a_2^2} \right]_{x=0}^{x=1} \end{aligned} \quad (3.22)$$

y

$$\begin{aligned} \mathcal{I}_x(a_2) &= n \mathbb{E} \left[\left(\frac{\partial}{\partial a_2} \log(a_0 + a_1 x + a_2 x^2) \right)^2 \right] = n \mathbb{E} \left[\left(\frac{x^2}{a_0 + a_1 x + a_2 x^2} \right)^2 \right] = \\ &= n \int_0^1 \left(\frac{x^2}{a_0 + a_1 x + a_2 x^2} \right)^2 (a_0 + a_1 x + a_2 x^2) dx = n \int_0^1 \frac{x^4}{a_0 + a_1 x + a_2 x^2} dx = \\ &= n \left[\frac{1}{6a_2^4} \left(a_2 x (6a_1^2 - 6a_0 a_2 - 3a_1 a_2 x + 2a_2^2 x^2) + \frac{6(a_1^4 - 4a_0 a_1^2 a_2 + 2a_0^2 a_2^2) \arctan \frac{a_1 + 2a_2 x}{\sqrt{-a_1^2 + 4a_0 a_2}}}{\sqrt{-a_1^2 + 4a_0 a_2}} \right. \right. \\ &\quad \left. \left. - 3(a_1^3 - 2a_0 a_1 a_2) \log(a_0 + a_1 x + a_2 x^2) \right) \right]_{x=0}^{x=1}, \end{aligned} \quad (3.23)$$

respectivamente.

Cabe resaltar que cada una de las funciones de información de Fisher (3.21), (3.22) y (3.23) no depende de x , es decir, de una muestra concreta, sino solamente del tamaño de la muestra, n , y de los parámetros a_0 , a_1 y a_2 .

3.3 Método de mínimos cuadrados

El único software disponible hasta ahora para estimar MoTBFs a partir de datos [1], el paquete MoTBFs de R [10], se basa en la técnica de mínimos cuadrados. El procedimiento de estimación asociado a este método consiste en considerar la función de distribución empírica como representación de los datos.

Definición 3.8. Sea X una muestra aleatoria de tamaño n extraída de una población X con función de distribución $F_X(x)$. Dada una observación muestral x , se define la **función de distribución empírica** asociada a la muestra como

$$G_n(x) = \frac{1}{n} \sum_{l=1}^n \mathbf{1}\{x_l \leq x\}, \quad x \in \mathbb{R},$$

donde $\mathbf{1}\{\cdot\}$ es la función indicadora.

Las funciones de densidad y distribución de una mezcla de funciones base truncadas están estrechamente vinculadas. Por ende, si somos capaces de encontrar una aproximación precisa de la función de distribución, podremos también obtener una estimación precisa de la densidad. El siguiente resultado establece la relación entre los parámetros de ambas funciones.

Lema 3.4. Sea X una variable aleatoria con función de distribución de tipo MoTBF asociada

$$F_X(x) = \sum_{i=0}^k a_i \eta_i(x).$$

Sea $\eta_0 = 1$ y sea η_i la función primitiva de la función base ψ_{i-1} , $i = 1, \dots, k$. Entonces, la función de densidad de X es también de tipo MoTBF.

Demostración:

La función de densidad de X es sencillamente la derivada de la función de distribución F_X . Esto es,

$$f_X(x) = \frac{d}{dx} F_X(x) = \sum_{i=0}^k a_i \eta_i'(x) = \sum_{i=1}^k a_i \psi_{i-1}(x) = \sum_{i=0}^{k-1} a_{i+1} \psi_i(x), \quad (3.24)$$

que es una función de tipo MoTBF. ■

El algoritmo desarrollado en [1] se basa en ajustar una MoTBF, cuya derivada es nuevamente una función de tipo MoTBF, a la función de distribución empírica mediante mínimos cuadrados. La optimización de mínimos cuadrados se utiliza para minimizar el error cuadrático medio entre la distribución empírica y la estimada.

Teniendo en cuenta el lema 3.4, podemos formular el siguiente problema de optimización para estimar los parámetros de una función de distribución correspondiente a una mezcla de funciones base truncadas definida en un intervalo $\Omega_X = [\alpha, \beta]$, a partir de una realización muestral \mathbf{x} :

$$\begin{aligned} & \text{minimize} && \sum_{l=1}^n \left(G_n(x_l) - \sum_{i=0}^k a_i \eta_i(x_l) \right)^2 \\ & \text{subject to} && \sum_{i=0}^{k-1} a_{i+1} \psi_i(x) \geq 0, \quad \forall x \in \Omega_X, \\ & && \sum_{i=0}^k a_i \eta_i(\alpha) = 0, \quad \sum_{i=0}^k a_i \eta_i(\beta) = 1, \end{aligned} \quad (3.25)$$

donde la primera restricción asegura que la función de densidad asociada a la función de distribución estimada sea no negativa o, equivalentemente, que dicha función de distribución sea no decreciente; mientras que las restricciones segunda y tercera juntas aseguran que esta densidad integre a uno. Esto último equivale a afirmar que la función de distribución debe ser igual a 0 en α e igual a 1 en β .

Es importante tener en cuenta que la función estimada a través de la resolución del problema (3.25) no es una función de densidad, sino de distribución. Sin embargo, basta aplicar el lema 3.4 para obtener la función de densidad MoTBF asociada, simplemente derivando dicha función de distribución como ilustra la ecuación (3.24).

La proposición 2 en [9, pág. 47] establece la consistencia de los estimadores de mínimos cuadrados bajo el supuesto de que la densidad objetivo es de tipo MoTBF. En la práctica, esta suposición no es demasiado restrictiva, ya que se sabe, gracias a [3] y [2], que las MoTBFs son capaces de representar con precisión cualquier distribución de probabilidad univariante.

En particular, el método de mínimos cuadrados para la estimación de parámetros en distribuciones univariantes de tipo mixtura de exponenciales truncadas y mixtura de polinomios adopta las expresiones (3.26) y (3.27), respectivamente.

Distribución de tipo MTE univariante

Sea X una variable aleatoria con función de distribución de tipo MTE asociada $F_X(x) = a_0 + \sum_{i=1}^k a_i e^{b_i x}$. Entonces, $H = \{1, e^{b_1 x}, \dots, e^{b_k x}\}$ y, de aquí, $\Psi = \{b_1 e^{b_1 x}, \dots, b_k e^{b_k x}\}$. Así, el problema de optimización para distribuciones univariantes de tipo MTE quedaría como:

$$\begin{aligned} & \text{minimize} && \sum_{l=1}^n \left(G_n(x_l) - \left(a_0 + \sum_{i=1}^k a_i e^{b_i x_l} \right) \right)^2 \\ & \text{subject to} && \sum_{i=1}^k a_i b_i e^{b_i x} \geq 0, \quad \forall x \in \Omega_X, \\ & && a_0 + \sum_{i=1}^k a_i e^{b_i \alpha} = 0, \quad a_0 + \sum_{i=1}^k a_i e^{b_i \beta} = 1. \end{aligned} \tag{3.26}$$

Distribución de tipo MOP univariante

Sea, ahora, una v.a. X con función de distribución de tipo MOP asociada

$$F_X(x) = \sum_{i=0}^k a_i x^i.$$

Entonces, $H = \{1, x, \dots, x^k\}$ y, de aquí, $\Psi = \{1, \dots, k x^{k-1}\}$. De esta forma, se obtiene el siguiente problema de optimización para distribuciones univariantes de tipo MOP:

$$\begin{aligned} & \text{minimize} && \sum_{l=1}^n \left(G_n(x_l) - \sum_{i=0}^k a_i x_l^i \right)^2 \\ & \text{subject to} && \sum_{i=1}^k i a_i x^{i-1} \geq 0, \quad \forall x \in \Omega_X, \\ & && \sum_{i=0}^k a_i \alpha^i = 0, \quad \sum_{i=0}^k a_i \beta^i = 1. \end{aligned} \tag{3.27}$$

Análisis experimental

Veremos, en este capítulo, el comportamiento de los métodos descritos en el capítulo 3 para estimar parámetros en distribuciones de tipo MoTBF a partir de datos, centrándonos principalmente en el algoritmo de *Fisher Scoring*.

Desde un punto de vista estadístico, el método de mínimos cuadrados no hace uso de la función de verosimilitud, que es la que reúne toda la información contenida en la muestra acerca de los parámetros desconocidos. De hecho, es un método numérico general de optimización. Además, es numéricamente inestable y poco eficiente cuando se parte de muestras grandes y muchos parámetros. De ahí, surge la necesidad de encontrar un método que a priori sea más eficiente y que sí aproveche la información en la muestra para estimar los parámetros. Un método que reúne estas características es el método de máxima verosimilitud. Luego, lo deseable sería obtener una estimación basada en la verosimilitud. Por este motivo, como no es posible resolver las ecuaciones de máxima verosimilitud en distribuciones de tipo MoTBF, el objetivo del presente capítulo consiste en la experimentación del método de *Fisher Scoring*.

No obstante, el método de mínimos cuadrados para distribuciones de tipo MoTBF está implementado en el paquete MoTBFs de R y, además, funciona muy bien en este caso, sobre todo para muestras pequeñas. Así pues, será nuestro marco de referencia, a fin de comparar los resultados obtenidos mediante el algoritmo de *Fisher Scoring*, cuyo comportamiento para densidades MoTBF desconocemos.

El procedimiento a seguir se divide en los siguientes pasos:

1. Primero, generar una muestra de valores entre 0 y 1, por ejemplo, una muestra aleatoria extraída de una población $X \sim \text{Beta}(5, 1)$.
2. Después, ajustar a la muestra generada una distribución de tipo MOP (respectivamente, MTE) de solo tres parámetros. Para ello, utilizaremos la función `uniMOTBF` del paquete MoTBFs de R, que sirve para ajustar una mixtura univariante de funciones base truncadas.
3. Por último, suponer los tres parámetros de la distribución obtenida en el paso anterior desconocidos y estimarlos por el método de *Fisher Scoring* presentado en la sección 3.2.

La idea consiste en comparar los valores de los parámetros obtenidos a través del algoritmo de *Fisher Scoring* con los calculados a través del ajuste por mínimos cuadrados de la función `uniMOTBF`. Es decir, comprobaremos qué tan similares son los ajustes proporcionados por cada uno de estos métodos de estimación, dada una muestra concreta.

4.1 Distribución de tipo MOP univariante

Para comenzar, sea X una muestra aleatoria de tamaño 1000 de una población $X \sim \text{Beta}(5, 1)$ y sea x una realización de la muestra. Entonces, le ajustamos a x una mixtura

de polinomios univariante mediante la función univMOTBF, obteniendo, así, la función de densidad MOP

$$f(x) = 0,54427424035255 - 4,00203375992288x + 7,37022791882667x^2 \quad (4.1)$$

para todo $x \in [0, 1]$. Esto es, el vector de parámetros estimado a partir de x mediante el método de mínimos cuadrados es

$$\begin{pmatrix} a_0 \\ a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} 0,54427424035255 \\ -4,00203375992288 \\ 7,37022791882667 \end{pmatrix}.$$

La representación gráfica de f sobre el histograma de la muestra viene dada por la Figura 4.1.

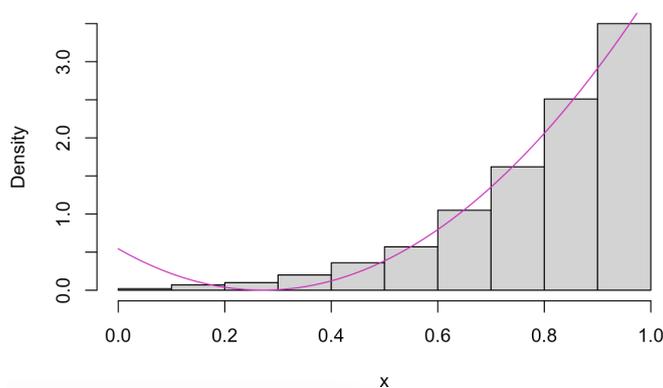


Figura 4.1: Ajuste de tipo MOP obtenido por mínimos cuadrados

A partir de aquí, consideraremos la función (4.1) como la función de densidad verdadera de la v.a. X .

Finalmente, procedemos a estimar los parámetros a_0 , a_1 y a_2 de la función

$$f(x; a_0, a_1, a_2) = a_0 + a_1 x + a_2 x^2, \quad x \in [0, 1],$$

a través del algoritmo iterativo de *Fisher Scoring*. Para ello, presentaremos varios experimentos, según el vector de parámetros de partida y el algoritmo utilizado.

Experimento 1. Algoritmo de Fisher Scoring

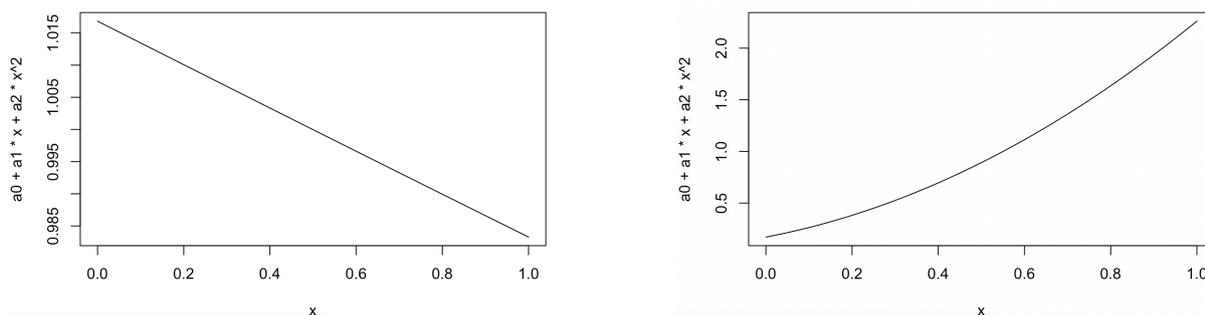
Sean $a_{0_0} = 3$, $a_{1_0} = -0,1$ y $a_{2_0} = 0,001$ los valores iniciales del vector de parámetros (a_0, a_1, a_2) . Entonces, la función de densidad MOP de partida (teniendo en cuenta la observación 3.1) se representa en la Figura 4.2a.

Ahora, tras realizar 100 iteraciones del algoritmo de *Fisher Scoring* descrito por la ecuación (3.15), obtenemos los parámetros

$$a_0 = 0,1711151, \quad a_1 = 0,792796, \quad a_2 = 1,297461,$$

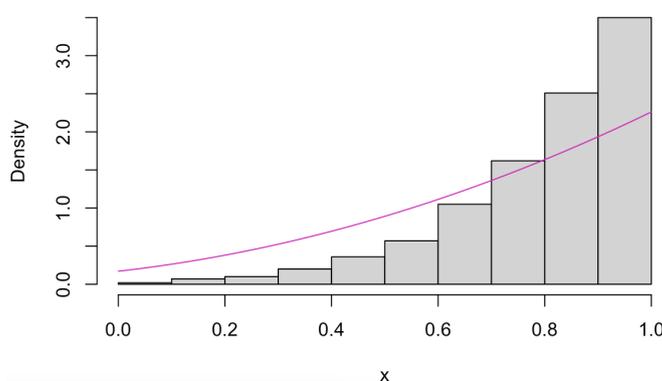
cuya función de densidad asociada se representa en la Figura 4.2b.

Observemos que, en este caso, a partir de una función de densidad decreciente (Figura 4.2a) se obtiene una buena aproximación de la densidad verdadera, tal y como se evidencia en la Figura 4.3.



(a) Función de densidad MOP inicial

(b) Función de densidad MOP estimada

Figura 4.2: Estimación de la densidad MOP mediante el método de *Fisher Scoring*Figura 4.3: Ajuste de tipo MOP obtenido mediante el método de *Fisher Scoring*

Algoritmo de *Fisher Scoring* optimizado

Ahora bien, si, partiendo del mismo vector de parámetros inicial, en lugar de aplicar el algoritmo (3.15), empleamos su versión optimizada en cuanto a número de operaciones, se obtiene lo siguiente.

Claramente, se parte de la misma función de densidad representada en la Figura 4.2a y, tras 100 iteraciones del algoritmo de *Fisher Scoring* optimizado descrito en 3.2, se obtienen los parámetros

$$a_0 = 0,3588236, \quad a_1 = 0,6228108, \quad a_2 = 0,9893131,$$

cuya función de densidad MOP asociada se representa en la Figura 4.4. De nuevo, se puede analizar la bondad del ajuste mediante la observación de la Figura 4.5.

Experimento 2. Algoritmo de *Fisher Scoring*

Partamos, ahora, de un vector de parámetros formado por valores aleatorios entre 0 y 1, según la función runif de R. Por ejemplo, sean $a_{0_0} = 0,2572191$, $a_{1_0} = 0,2222187$ y $a_{2_0} = 0,4238321$. De este modo, la función de densidad MOP inicial (teniendo en cuenta la observación 3.1) viene dada por la Figura 4.6.

Luego, tras realizar 100 iteraciones del algoritmo de *Fisher Scoring* descrito por la ecuación (3.15), se obtienen los parámetros

$$a_0 = 0,1711151, \quad a_1 = 0,792796, \quad a_2 = 1,297461,$$

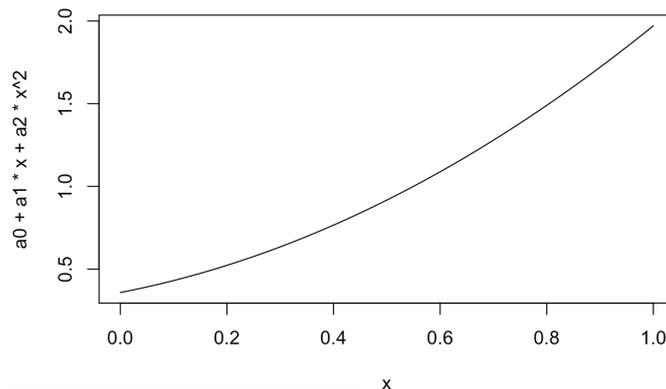


Figura 4.4: Función de densidad MOP obtenida mediante el método de *Fisher Scoring* optimizado

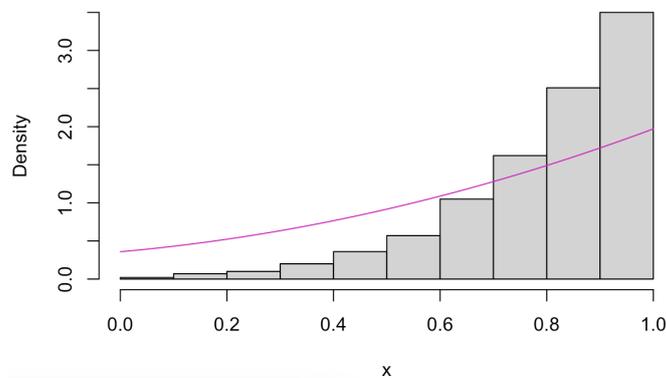


Figura 4.5: Ajuste de tipo MOP mediante el método de *Fisher Scoring* optimizado

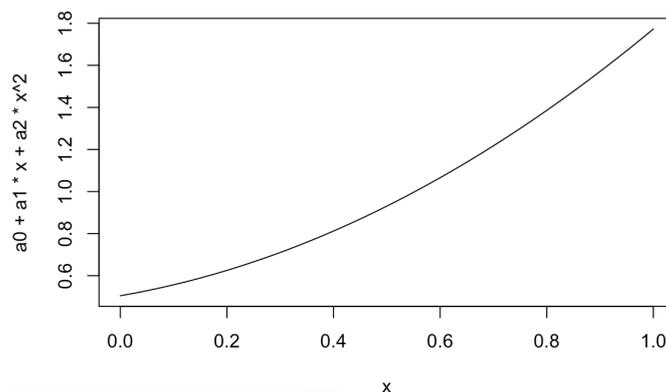


Figura 4.6: Función de densidad MOP inicial para el método de *Fisher Scoring*

que coinciden con los obtenidos por este mismo algoritmo en el Experimento 1.

Es decir, partiendo de valores iniciales distintos, el algoritmo de *Fisher Scoring* definido por la ecuación (3.15) obtiene los mismos valores para los parámetros a_0 , a_1 y a_2 , dada la muestra x .

Sin embargo, en este caso, el cambio en cuanto al valor de los parámetros no es tan brusco como en el Experimento 1, debido a que la función de densidad MOP inicial

de la que partía el algoritmo (Figura 4.6) ya se aproximaba bastante a la función de densidad obtenida por el método (Figura 4.2b).

Algoritmo de Fisher Scoring optimizado

De nuevo, estudiaremos el comportamiento de la versión optimizada del algoritmo (3.15), en este caso, partiendo de un vector de parámetros formado por valores aleatorios entre 0 y 1. Sean, por ejemplo, $a_{0_0} = 0,79425$, $a_{1_0} = 0,3768935$ y $a_{2_0} = 0,8716043$. Entonces, la función de densidad MOP inicial (teniendo en cuenta la observación 3.1) se representa en la Figura 4.7a.

Posteriormente, tras realizar 100 iteraciones del algoritmo de *Fisher Scoring* optimizado descrito en 3.2, obtenemos los parámetros

$$a_0 = 0,2875939, \quad a_1 = 0,6692, \quad a_2 = 1,133418,$$

cuya función de densidad MOP asociada se representa en la Figura 4.7b.

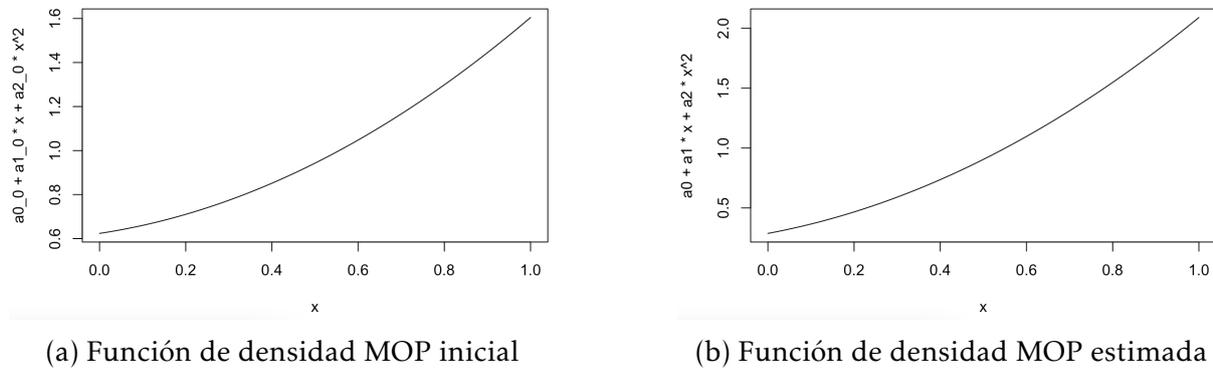


Figura 4.7: Estimación de la función de densidad MOP mediante el método de *Fisher Scoring* optimizado

La Figura 4.8 evidencia la bondad del ajuste realizado por este método.

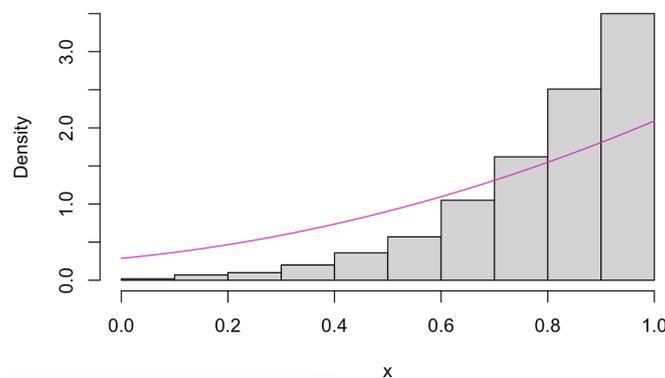


Figura 4.8: Ajuste de tipo MOP mediante el método de *Fisher Scoring* optimizado

No obstante, al igual que en el caso anterior, la diferencia entre la densidad de partida (Figura 4.7a) y la obtenida a través del método (Figura 4.7b) es mínima, con lo cual el cambio no es tan sorprendente como en el Experimento 1.

Conclusiones

Este Trabajo Fin de Grado ha tenido como objetivo principal el estudio de las distribuciones de tipo MoTBF (en particular, MTEs y MOPs), comprendiendo aspectos que van desde sus propiedades fundamentales, hasta el problema de estimación de parámetros en dichas distribuciones que da nombre al trabajo, pasando por su utilidad en modelos jerárquicos bayesianos en entornos *big data*, dado que no se imponen restricciones en cuanto a la estructura del modelo jerárquico (a diferencia del modelo Gaussiano condicional) y se pueden usar los algoritmos basados en cálculos locales desarrollados para el caso de inferencia con variables discretas.

Para empezar, con respecto a la primera cuestión, esto es, el estudio de las propiedades básicas de las distribuciones de tipo MTE y MOP, se han obtenido resultados innovadores, a saber: las MTEs no tienen función generatriz de momentos, a diferencia de las MOPs; así como otros resultados relevantes que han contribuido a justificar las ventajas de las MOPs sobre las MTEs. Entre ellos, destacamos dos: por un lado, la expresión del momento no centrado de orden k se puede obtener de forma sencilla para distribuciones de tipo MOP, mientras que para MTEs se complica el cálculo integral conforme aumenta el orden; y, por otro lado, ni la mediana ni la moda de una distribución de tipo MTE se pueden calcular de forma cerrada, al contrario que para distribuciones de tipo MOP. En definitiva, concluimos de dicho estudio que el hecho de utilizar polinomios en lugar de funciones exponenciales simplifica los cálculos y supone ciertas ventajas, como la existencia de la función generatriz de momentos.

La segunda parte del trabajo ha consistido en el desarrollo y análisis de procedimientos de estimación de parámetros en distribuciones de tipo MoTBF a partir de grandes volúmenes de datos, teniendo en cuenta la eficiencia en cuanto a número de operaciones requeridas. De entre varios enfoques posibles, en este proyecto se ha decidido abordar el estudio la estimación de parámetros en distribuciones univariantes de tipo MTE y MOP, a través de los métodos de máxima verosimilitud (tanto en su versión exacta como en su versión aproximada: método de *Fisher Scoring*) y mínimos cuadrados. Nuestro objetivo era encontrar un método de estimación eficiente que hiciese uso de la función de verosimilitud, es decir, que aprovechara la información contenida en la muestra, para estimar los parámetros. Por ello, como en la sección 3.1 se concluye que no es posible resolver las ecuaciones de máxima verosimilitud de forma exacta en este tipo de distribuciones, nos planteamos estudiar en la sección 3.2 un método aproximado, como es el método de *Fisher Scoring*. En el capítulo 4, se analizó de forma experimental la aplicación de este método a la estimación de parámetros en distribuciones univariantes de tipo MOP. De ahí se concluye que, pese a no cumplir con los objetivos planteados, este método apunta a ser una idea prometedora de cara a futuros proyectos de investigación.

De esta forma, el presente trabajo constituye una primera aproximación en la línea de investigación referente al desarrollo de métodos de estimación de parámetros en distribuciones de tipo MoTBF a partir de grandes volúmenes de datos. Así, en futuros proyectos de investigación, se podrían utilizar los desarrollos de este trabajo para, por ejemplo, ampliar el estudio a distribuciones condicionadas (muy necesarias en el contexto de modelos jerárquicos con muchas variables) y distribuciones multivariantes. Una ventaja que presentan los modelos de tipo MoTBF es que la estimación de

distribuciones condicionadas se reduce, en realidad, a la estimación de distribuciones univariantes, debido a la naturaleza de su definición (véase la sección 2.1). Asimismo, en futuros trabajos, también se podría estudiar la posibilidad de adoptar un enfoque Bayesiano, para lo cual se considerarían tipos particulares de distribuciones dentro de la familia MoTBF en las que los parámetros se pudiesen expresar como variables aleatorias y fuese posible obtener su distribución a posteriori.

Por otro lado, cabe mencionar que no ha sido posible incluir en esta memoria todos los aspectos que se han estudiado, debido a la limitación de cincuenta páginas. En este sentido, también se ha examinado la aplicación del método de *Fisher Scoring* a la estimación de parámetros en distribuciones de tipo MTE univariantes. No obstante, la implementación de dicho método en R presentó algunos inconvenientes, provocando que el análisis experimental pertinente no fuese fructífero. Por ese motivo, se ha decidido incluir el estudio correspondiente a distribuciones de tipo MOP en lugar del mencionado. Además, se ha sacrificado gran parte del material que constituía la sección de ejemplos del capítulo 2, también por motivos de espacio.

Para la elaboración del presente trabajo, ha resultado imprescindible la utilización de conceptos y resultados estudiados en diversas materias del Grado, destacando principalmente las asignaturas *Introducción a la Probabilidad y a la Estadística*, *Probabilidad*, *Estadística*, *Métodos Estadísticos para Big Data* y *Análisis de Datos*.

Para terminar, me gustaría destacar que el proceso de elaboración de este trabajo ha contribuido de forma muy positiva a mi desarrollo académico en numerosos aspectos. En cuanto a informática, el trabajo realizado en este proyecto se ha apoyado en el uso del software estadístico R, lo cual me ha permitido ampliar mi formación con respecto a esta herramienta software y adquirir competencias en el ámbito del cálculo estadístico en entornos con gran capacidad de proceso. Además, la redacción del presente texto ha contribuido a desarrollar considerablemente mis destrezas con \LaTeX . Por otra parte, la realización de este trabajo ha favorecido también la mejora de mi nivel de inglés, especialmente en cuanto a vocabulario matemático, gracias a que toda la bibliografía consultada para este fin estaba en dicho idioma. Por último, me ha permitido introducirme en el mundo de la investigación matemática, junto con la ayuda de mi tutor.

Bibliografía

- [1] H. Langseth, T.D. Nielsen, I. Pérez-Bernabé, A. Salmerón, *Learning mixtures of truncated basis functions from data*, International Journal of Approximate Reasoning, **55** (2014), 940–956.
- [2] H. Langseth, T.D. Nielsen, R. Rumí, A. Salmerón, *Mixtures of truncated basis functions*, International Journal of Approximate Reasoning, **53** (2012), 212–227.
- [3] H. Langseth, T.D. Nielsen, R. Rumí, A. Salmerón, *Parameter estimation and model selection for mixtures of truncated exponentials*, International Journal of Approximate Reasoning, **51** (2010), 485–498.
- [4] S. L. Lauritzen, *Propagation of probabilities, means, and variances in mixed graphical association models*, Journal of the American Statistical Association, **87** (1992), 1098–1108.
- [5] A. L. Madsen, F. V. Jensen, *Lazy propagation: a junction tree inference algorithm based on lazy evaluation*, Artificial Intelligence, **113** (1999), 203–245.
- [6] S. Moral, R. Rumí, A. Salmerón, *Mixtures of truncated exponentials in hybrid Bayesian networks*, ECSQARU'01. Lecture Notes in Artificial Intelligence, **2143** (2001), 156–167.
- [7] J. Pearl, *Probabilistic Reasoning in Intelligent Systems Morgan Kaufmann Publishers*, San Mateo, California, 1988.
- [8] I. Pérez-Bernabé, A. Fernández, R. Rumí, A. Salmerón, *Parameter learning in hybrid Bayesian networks using prior knowledge*, Data mining and knowledge discovery, **30** (2016), 576–604.
- [9] I. Pérez-Bernabé, *Learning hybrid bayesian networks using mixtures of truncated basis functions. Aprendizaje de redes bayesianas híbridas con mixturas de funciones base truncadas*, Tesis Doctoral, Universidad de Almería, 2015.
- [10] I. Pérez-Bernabé, A.D. Maldonado, T.D. Nielsen, A. Salmerón, *MoTBFs: An R package for learning hybrid Bayesian networks using mixtures of truncated basis functions*, R Journal, **12** (2020), 342–358.
- [11] R. Rumí, A. Salmerón, S. Moral, *Estimating mixtures of truncated exponentials in hybrid Bayesian networks*, Test, **15** (2006), 397–421.
- [12] P. P. Shenoy, J. C. West, *Inference in hybrid Bayesian networks using mixtures of polynomials*, International Journal of Approximate Reasoning, **52** (2011), 641–657.
- [13] Página web de la matriz de información de Fisher: https://en.wikipedia.org/wiki/Fisher_information.
- [14] Página web de la matriz de información observada: https://en.wikipedia.org/wiki/Observed_information.
- [15] Página web del algoritmo de Fisher Scoring: https://en.wikipedia.org/wiki/Scoring_algorithm.

Apéndice

En este apéndice se incluye el código de R utilizado para la realización del análisis experimental correspondiente al capítulo 4.

A.1 Implementación en R del método de Fisher Scoring para distribuciones univariantes de tipo MOP de grado 2

Generamos una muestra de valores entre 0 y 1:

```
x <- rbeta(1000, 5, 1)
hist(x, prob = TRUE, main = "")
```

Le ajustamos una función de densidad MOP de grado 2 haciendo uso de la función `univMOTBF` del paquete `MoTBFS` de R:

```
library(MoTBFS)
f <- univMOTBF(x, POTENTIAL_TYPE = "MOP", nparam = 3); f
plot(f, xlim = range(x), col = 6, add=TRUE)
```

Procedemos a estimar los parámetros a_0 , a_1 y a_2 de una distribución MOP univariante de grado 2 mediante el método de *Fisher Scoring* presentado en la sección 3.2:

```
# Funciones necesarias para implementar el algoritmo:

score_a0 <- function(x, a0, a1, a2) {
  sum(1/(a0 + a1*x + a2*x^2))
}

score_a1 <- function(x, a0, a1, a2) {
  sum(x/(a0 + a1*x + a2*x^2))
}

score_a2 <- function(x, a0, a1, a2) {
  sum(x^2/(a0 + a1*x + a2*x^2))
}

integral_a0 <- function(a0, a1, a2, x) {
  2*atan((a1 + 2*a2*x)/sqrt(-a1^2 + 4*a0*a2))/sqrt(-a1^2 + 4*a0*a2)
}

infFisher_a0 <- function(a0, a1, a2, n) {
  n*(integral_a0(a0, a1, a2, 1) - integral_a0(a0, a1, a2, 0))
}
```

```

integral_a1 <- function(a0, a1, a2, x) {
  x/a2 + (a1^2 - 2*a0*a2)*atan((a1 + 2*a2*x)/sqrt(-a1^2 + 4*a0*a2))/
  (a2^2*sqrt(-a1^2 + 4*a0*a2)) - a1*log(a0 + a1*x + a2*x^2)/(2*a2^2)
}

infFisher_a1 <- function(a0, a1, a2, n) {
  n*(integral_a1(a0, a1, a2, 1) - integral_a1(a0, a1, a2, 0))
}

integral_a2 <- function(a0, a1, a2, x) {
  1/(6*a2^4)*(a2*x*(6*a1^2 - 6*a0*a2 - 3*a1*a2*x + 2*a2^2*x^2) +
  6*(a1^4 - 4*a0*a1^2*a2 + 2*a0^2*a2^2)*atan((a1 + 2*a2*x)/
  sqrt(-a1^2 + 4*a0*a2))/sqrt(-a1^2 + 4*a0*a2) - 3*(a1^3
  - 2*a0*a1*a2)*log(a0 + x*(a1 + a2*x)))
}

infFisher_a2 <- function(a0, a1, a2, n) {
  n*(integral_a2(a0, a1, a2, 1) - integral_a2(a0, a1, a2, 0))
}

```

Algoritmo de Fisher Scoring

```

# Experimento 1
a0 = 3
a1 = -0.1
a2 = 0.001

# Experimento 2 (valores iniciales aleatorios)
a0 = runif(1, 0, 1); a0
a1 = runif(1, 0, 1); a1
a2 = runif(1, 0, 1); a2

# Algoritmo de Fisher Scoring:

t <- proc.time() # Inicia el cronómetro

K = a0 + a1/2 + a2/3
a0 = a0/K
a1 = a1/K
a2 = a2/K
curve(a0 + a1*x + a2*x^2, 0, 1)

for (m in 0:100) {
  nuevo_a0 = a0 + 1/infFisher_a0(a0, a1, a2, length(x))*

```

A.1. Implementación en R del método de *Fisher Scoring* para distribuciones univariantes de tipo MOP de grado 2

```
score_a0(x, a0, a1, a2)
nuevo_a1 = a1 + 1/infFisher_a1(a0, a1, a2, length(x))*
score_a1(x, a0, a1, a2)
nuevo_a2 = a2 + 1/infFisher_a2(a0, a1, a2, length(x))*
score_a2(x, a0, a1, a2)
K = nuevo_a0 + nuevo_a1/2 + nuevo_a2/3
a0 = nuevo_a0/K
a1 = nuevo_a1/K
a2 = nuevo_a2/K
cat('a0 = ', a0, ' a1 = ', a1, ' a2 = ', a2, '\n')
}
proc.time()-t      # Detiene el cronómetro

# Representación gráfica de la función de densidad estimada
curve(a0 + a1*x + a2*x^2, 0, 1)

# Representación de la densidad estimada sobre el histograma de la muestra:
hist(x, prob = TRUE, main = "")
curve(a0 + a1*x + a2*x^2, 0, 1, col = 6, add = TRUE)
```

Algoritmo de Fisher Scoring optimizado

```
# Experimento 1
a0_0 = 3
a1_0 = -0.1
a2_0 = 0.001

# Experimento 2 (valores iniciales aleatorios)
a0_0 = runif(1, 0, 1); a0_0
a1_0 = runif(1, 0, 1); a1_0
a2_0 = runif(1, 0, 1); a2_0

# Algoritmo de Fisher Scoring optimizado:
t <- proc.time()      # Inicia el cronómetro

K = a0_0 + a1_0/2 + a2_0/3
a0_0 = a0_0/K
a1_0 = a1_0/K
a2_0 = a2_0/K
curve(a0_0 + a1_0*x + a2_0*x^2, 0, 1)

s_a0_0 = score_a0(x, a0_0, a1_0, a2_0)
i_a0_0 = infFisher_a0(a0_0, a1_0, a2_0, length(x))
```

```

s_a1_0 = score_a1(x, a0_0, a1_0, a2_0)
i_a1_0 = infFisher_a1(a0_0, a1_0, a2_0, length(x))
s_a2_0 = score_a2(x, a0_0, a1_0, a2_0)
i_a2_0 = infFisher_a2(a0_0, a1_0, a2_0, length(x))

nuevo_a0 = a0_0 + 1/infFisher_a0(a0_0, a1_0, a2_0, length(x))*s_a0_0
nuevo_a1 = a1_0 + 1/infFisher_a1(a0_0, a1_0, a2_0, length(x))*s_a1_0
nuevo_a2 = a2_0 + 1/infFisher_a2(a0_0, a1_0, a2_0, length(x))*s_a2_0
K = nuevo_a0 + nuevo_a1/2 + nuevo_a2/3
a0 = nuevo_a0/K
a1 = nuevo_a1/K
a2 = nuevo_a2/K
curve(a0 + a1*x + a2*x^2, 0, 1)

for (m in 0:100) {
  nuevo_a0 = a0 + 1/infFisher_a0(a0, a1, a2, length(x))*
    (s_a0_0 - i_a0_0*(a0 - a0_0))
  nuevo_a1 = a1 + 1/infFisher_a1(a0, a1, a2, length(x))*
    (s_a1_0 - i_a1_0*(a1 - a1_0))
  nuevo_a2 = a2 + 1/infFisher_a2(a0, a1, a2, length(x))*
    (s_a2_0 - i_a2_0*(a2 - a2_0))
  K = nuevo_a0 + nuevo_a1/2 + nuevo_a2/3
  a0 = nuevo_a0/K
  a1 = nuevo_a1/K
  a2 = nuevo_a2/K
  cat('a0 = ', a0, ' a1 = ', a1, ' a2 = ', a2, '\n')
}
proc.time()-t # Detiene el cronómetro

# Representación gráfica de la función de densidad estimada
curve(a0 + a1*x + a2*x^2, 0, 1)

# Representación de la densidad estimada sobre el histograma de la muestra
hist(x, prob = TRUE, main = "")
curve(a0 + a1*x + a2*x^2, 0, 1, col = 6, add = TRUE)

```