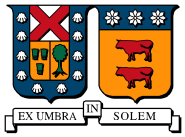
2020-09

# ESTIMACIÓN EMPÍRICA BAYESIANA UTILIZANDO SUMA DE DISTRIBUCIONES GAUSSIANAS

ORELLANA PRATO, RAFAEL ANGEL




UNIVERSIDAD TÉCNICA FEDERICO SANTA MARÍA
Departamento de Ingeniería Electrónica

# Estimación empírica bayesiana utilizando suma de distribuciones gaussianas

Tesis de grado presentada por
**Rafael Angel Orellana Prato**

Como requisito parcial para optar al grado de
**Magíster en Ciencias de la Ingeniería Electrónica**

Supervisor
Juan C. Agüero, Ph.D.

Co-supervisor
Rodrigo Carvajal., Ph.D.

Valparaíso, 15 de septiembre de 2020

UNIVERSIDAD TÉCNICA FEDERICO SANTA MARÍA
Departamento de Ingeniería Electrónica

# Estimación empírica bayesiana utilizando suma de distribuciones gaussianas

Tesis de grado presentada por
**Rafael Angel Orellana Prato**

Como requisito parcial para optar al grado de
**Magíster en Ciencias de la Ingeniería Electrónica**

Supervisor
Juan C. Agüero, Ph.D.

Co-supervisor
Rodrigo Carvajal., Ph.D.

Valparaíso, 15 de septiembre de 2020

Tesis:

**Estimación empírica bayesiana utilizando suma de distribuciones gaussianas**

Autor:

**Rafael Angel Orellana Prato**

TRABAJO DE TESIS, presentado en cumplimiento parcial de los requisitos para el grado de Magister en Ciencias de la Ingeniería Electrónica de la Universidad Técnica Federico Santa María, Valparaíso, Chile.

Supervisor,
Juan C. Agüero, Ph.D. _____

Co-Supervisor,
Rodrigo Carvajal, Ph.D. _____

Revisor Interno,
Francisco Vargas, Ph.D. _____

Revisor Externo,
Ramon Delgado, Ph.D. _____

Título:

**Estimación empírica bayesiana utilizando suma de distribuciones gaussianas**


RESUMEN

El problema clásico de estimación bayesiana ha sido estudiado en diferentes áreas de la ciencia y tecnología. La idea principal se fundamenta en el teorema de Bayes, con el fin de obtener una estimación de un proceso estocástico basado en observaciones relacionadas a la variable aleatoria y el conocimiento previo de la función de densidad de probabilidad del proceso. Sin embargo, el correcto desempeño del algoritmo de estimación está sujeto a las suposiciones iniciales acerca de la distribución de probabilidad a *priori* de la variable aleatoria. En este trabajo se propone un algoritmo de identificación para obtener una estimación de la función de distribución a *priori* del problema tradicional de inferencia bayesiana. Se utiliza en enfoque de estimación empírica bayesiana modelando la función de distribución a *priori* como una mezcla finita de distribuciones gaussianas. El problema de estimación se resuelve usando un algoritmo basado en la maximización de la esperanza considerando simultáneamente la información de un conjunto de experimentos independientes. El algoritmo propuesto presenta una buena exactitud en la estimación a medida que se aumenta el número de experimentos. Finalmente, los beneficios del algoritmo propuesto se muestran a través de ejemplos numéricos de simulación.



***Palabras clave:*** Inferencia bayesiana, Estimación empírica bayesiana, Modelo de mezclas gaussianas, Distribución a *priori*, Maximizar la esperanza.

Title:

**Empirical Bayes estimation utilizing Gaussian mixture models**


ABSTRACT

The classical problem of Bayesian estimation has been addressed in different frameworks of science and technology. The key idea comes from the well-known Bayes Theorem, where we need to obtain an estimation of a stochastic process utilizing measurements and a *prior* distribution of the random variable. Nevertheless, the estimation accuracy can be far from the *true* value when the assumption of the *prior* distribution does not correspond to the *true* distribution. In this thesis an identification algorithm to obtain an estimation of the *prior* distribution in the classical problem of Bayesian inference is developed. The Empirical Bayes approach is considered to obtain the *prior* distribution as a finite Gaussian mixture model. An Expectation-Maximization based algorithm is used to obtain an estimate of the Gaussian mixture model parameters. This approach shows an accurate estimation of the *prior* distribution when the number of experiments is increased. The benefits of the proposed algorithm are illustrated via numerical simulations.



***Keywords:*** Bayesian inference, Empirical Bayes estimation, Gaussian mixture model, *Prior* distribution, Expectation-Maximization.

# Contenido

# Lista de Tablas

# Lista de Figuras

# CAPÍTULO 1
# INTRODUCCIÓN

## 1.1. Motivación

La estimación bayesiana es un tipo de inferencia estadística en la que se usan observaciones de una variable aleatoria para inferir la probabilidad de que un supuesto o hipótesis pueda ser cierto. Este concepto ha sido aplicado en diversas áreas como inferencia estadística [1–3], sistemas lineales con datos cuantizados [4], sistemas no lineales [5,6], control [7], comunicaciones [8,9] y astronomía [10,11], entre otros. La idea principal se basa en obtener una estimación de un proceso aleatorio usando: i) la información contenida en un conjunto de observaciones relacionadas a una variable aleatoria y ii) el conocimiento previo de la función de densidad de probabilidad (pdf) del proceso. La formulación del problema de estimación se basa en el teorema de Bayes el cual se expresa como:

$$p(\theta|\mathscr{D}) = \frac{p(\mathscr{D}|\theta)p(\theta)}{p(\mathscr{D})}, \tag{1.1}$$

donde $\mathscr{D}$ es un conjunto de observaciones o mediciones del proceso estocástico $\theta$, $p(\mathscr{D}|\theta)$ es la función de verosimilitud, $p(\theta)$ es la pdf a *priori* (*prior distribution*) y $p(\theta|\mathscr{D})$ es la pdf a *posteriori* (*posterior distribution*) del proceso estocástico. Es importante resaltar que $p(\mathscr{D})$ puede ser considerada como una "constante"de normalización que no es necesario calcularla directamente [12].

La metodología de estimación bayesiana permite obtener estimadores clásicos a partir de la caracterización de la pdf a *posteriori* del proceso estocástico en función del error de estimación, por ejemplo, maximizar la función de distribución a *posteriori* (MAP) [13], maximizar la función de verosi-

militud (ML) [14, 15], minimizar el error cuadrático medio de estimación (MMSE) [16], minimizar el valor absoluto del error de estimación (MAVE) [17], entre otros. Un enfoque práctico ha sido estudiado en sistemas de comunicación inalámbrica para analizar el comportamiento aleatorio del canal de comunicación [8, 9]. De manera similar, el enfoque bayesiano ha sido usado en el diseño de clasificadores para cuantificar la calidad percibida en sistemas de voz sobre protocolo de internet (VoIP) [18] o bien en sistemas jerárquicos de aprendizaje [19]. Todas estas formulaciones requieren conocer (o asumir) un modelo de la pdf a *priori* a fin de obtener una metodología de estimación.

El principal problema del paradigma de estimación bayesiana es determinar de manera apropiada la estructura o modelo de la pdf a *priori*, $p(\theta)$. Típicamente, la selección se hace en base a información conocida acerca del proceso en estudio o bien sobre la experiencia en el área de aplicación [20–22]. Otro enfoque corresponde a la estimación empírica de Bayes (EB), la cual consiste en utilizar la estimación por ML a fin de obtener los parámetros que definen la pdf a *priori* [23, 24]. Este planteamiento considera una suposición inicial sobre el modelo de la pdf a *priori*, donde el desempeño del algoritmo de estimación puede comprometerse cuando $p(\theta)$ no corresponde con la suposición inicial considerada. Por lo tanto, es necesario considerar escenarios más flexibles que permitan estudiar la estimación de la pdf a *priori*, por ejemplo, aproximación usando un modelo de suma finita de distribuciones gaussianas (GMMs).

Tradicionalmente, las GMMs han sido usadas para aproximar funciones de distribución desconocidas o no-gaussianas en áreas como filtros no lineales [25], seguimientos de trayectorias [26, 27], clasificación [28], inferencia estadística [3, 29] entre otros. Recientemente, las GMMs han sido usadas para obtener algoritmos de identificación para sistemas dinámicos con el enfoque de ML [30] y con un enfoque bayesiano [31, 32]. Además, el problema de identificación en sistemas estáticos ha sido abordado usando mezcla

finita de distribuciones, por ejemplo, en la deconvolución de la velocidad de rotación estelar [29,33]. Por otro lado, es posible obtener la formulación clásica del algoritmo de maximización de la esperanza (EM) con GMMs para el problema de estimación por ML en el marco empírico bayesiano. Para ello se utiliza el enfoque de datos aumentados en donde se incorpora una variable no observada con el fin de maximizar de forma iterativa el valor esperado condicional de la función de verosimilitud del conjunto de datos completos (mediciones observadas y no observadas) [34]. Sin embargo, este planteamiento requiere observaciones directas de la variable estocástica $\theta$, las cuales no están disponibles para el problema de estimación empírica bayesiana.

Por todo esto, el objetivo principal de este trabajo es proponer una metodología de identificación usando el enfoque EB para obtener una estimación de la pdf a *priori*, $p(\theta)$, en (1.1) como la suma finita de distribuciones gaussianas. Se propone un algoritmo basado en ML para estimar los parámetros de la GMM que define la pdf a *priori* del proceso estocástico $\theta$. Se utiliza un algoritmo basado en EM para resolver el problema de estimación por ML en forma iterativa.

## 1.2. Contribuciones de la Tesis

Las principales contribuciones de la tesis son las siguientes:

1. Se propone una metodología de identificación para la función de distribución a *priori* como una suma finita de distribuciones gaussianas usando el enfoque EB.

2. Se propone un algoritmo basado en EM para resolver el problema de optimización asociado al algoritmo de estimación EB. Con el algoritmo propuesto se obtiene una forma cerrada de los estimadores de los parámetros que definen la GMM.

3. La implementación del algoritmo propuesto se hace utilizando aproximación numérica para resolver las ecuaciones integrales que definen los estimadores de la GMM.

## 1.3. Publicación asociada a la tesis

Artículo en Conferencia:

R. Orellana, R. Carvajal and J.C. Agüero, "Empirical Bayes estimation utilizing finite Gaussian Mixture Models," in *2019 IEEE CHILEAN Conference on Electrical, Electronics Engineering, Information and Communication Technologies (CHILECON)*, Valparaíso, Chile, 2019, pp. 1–6.

**Abstract:** In this paper we develop an identification algorithm to obtain an estimation of the *prior* distribution in the classical problem of Bayesian inference. We consider the Empirical Bayes approach to obtain the *prior* distribution approximation by a finite Gaussian mixture. An Expectation-Maximization based algorithm is used to obtain an estimate of the Gaussian mixture parameters. Our approach shows a good approximation of the *prior* distribution when the number of experiments is increased. We illustrate the estimation performance of our proposal with numerical simulations.

## 1.4. Notación

La notación que se utiliza en el documento es la siguiente: $p(x|\eta)$ representa la función de densidad de probabilidad de la variable aleatoria $x$ conociendo el vector de parámetros $\eta$. $\eta_0$ denota el vector de parámetros verdadero, $\hat{\eta}$ una estimación de $\eta_0$ y $\hat{\eta}^{(k)}$ denota la $k$-ésima iteración para la estimación del parámetro. $(\cdot)^T$ denota el operador traspuesto. $\mathbb{R}$ representa el conjunto de números reales. Finalmente, $\mathbb{E}\{\cdot\}$ representa el operador esperanza.

## 1.5. Organización del documento

La presente tesis esta organizada por capítulos de la siguiente manera:

- El Capítulo 2 presenta los conceptos fundamentales relacionados a la representación de funciones de distribución de probabilidad como la suma finita de funciones de distribución gaussianas. El objetivo es considerar esta representación para aproximar la función de distribución a *priori* en el enfoque EB cuando ésta no corresponde a una GMM.
- El Capítulo 3 muestra los conceptos relacionados a la estimación por ML usando GMM. Adicionalmente, se formula el problema de estimación usando el enfoque EB para obtener los parámetros que definen la GMM.
- El Capítulo 4 presenta el desarrollo de un algoritmo basado en EM para el problema de interés, obteniendo los estimadores para los parámetros que definen la GMM que aproxima la función de distribución a *priori*.
- El Capítulo 5 presenta los ejemplos de simulación numérica para analizar el desempeño del algoritmo propuesto.
- El Capítulo 6 presenta las conclusiones de la tesis y las opciones de trabajo futuro para dar continuidad a la investigación.

# CAPÍTULO 2
# MODELO DE MEZCLAS GAUSSIANAS

## 2.1. Introducción

En este capítulo se estudian los conceptos básicos asociados a funciones de distribución de probabilidad dadas por la suma finita de distribuciones conocidas, en particular funciones de distribución gaussianas. Se muestra la fundamentación teórica para aproximar funciones de distribución usando modelos de mezclas gaussianas. Por último, se muestran algunos criterios para la selección del número de componentes en los modelos de mezclas gaussianas.

## 2.2. Formulación básica

En estadística, un modelo de mezclas de funciones de distribución permite determinar la presencia de sub-poblaciones dentro de una población general, esto sin requerir observaciones individuales de cada una de ellas [35]. La Figura 2.1 muestra un ejemplo considerando la mezcla de tres funciones de distribución gaussiana.

La función de densidad de probabilidad en el caso de una mezcla finita de distribuciones de una variable aleatoria $n_y$-dimensional discreta $Y$, viene dada por:

$$p(y) = \sum_{i=1}^{\kappa} \lambda_i p_i(y), \tag{2.1}$$

donde $\kappa$ es el número de componentes de la mezcla de distribuciones, $\lambda_i$ es el factor de ponderación de la mezcla sujeto a $\sum_{i=1}^{\kappa} \lambda_i = 1$, $0 \leq \lambda_i \leq 1$ y $p_i(y)$ son las funciones de densidad que forman la mezcla de distribuciones. Típicamente, en los modelos de mezclas de distribuciones se considera que

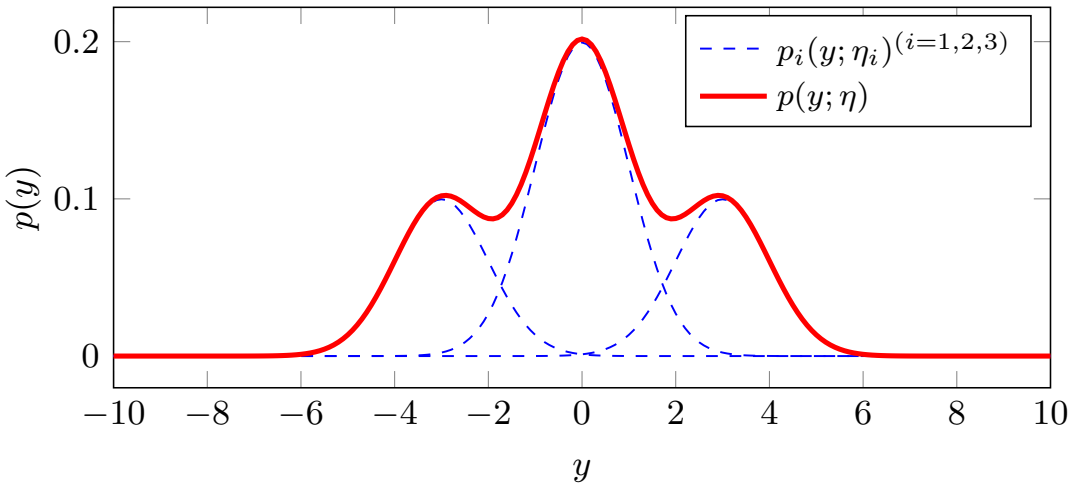



Figura 2.1: Ejemplo de una mezcla de tres funciones de distribución gaussiana.

las componentes son parametrizadas por un vector de parámetros $\gamma_i$. En ese caso, la mezcla finita de distribuciones se puede escribir como:

$$p(y;\eta) = \sum_{i=1}^{\kappa} \lambda_i p_i(y;\gamma_i), \tag{2.2}$$

donde $\eta = [\lambda_1 \ \gamma_1^T \ \ldots \lambda_M \ \gamma_M^T]^T$ representa el vector de parámetros que definen el modelo de mezclas de distribuciones. Generalmente, las componentes de la mezcla, $p_i(y;\gamma_i)$, pertenecen a la misma familia de funciones de distribución, por ejemplo, funciones de distribución gaussianas [34,36]. Sin embargo, en algunas aplicaciones la función de distribución corresponde a la mezcla de funciones de densidad diferentes, por ejemplo, el problema de procesamiento y detección de imágenes multi-espectrales [37].

## 2.3. Modelos de mezclas gaussianas para aproximar funciones de distribución

Típicamente, las GMMs han sido usadas en áreas como filtros no lineales [12,38,39], seguimientos de trayectorias [26,27], clasificación [28] entre otros. Recientemente, este enfoque ha sido usado para proponer algoritmos de identificación de sistemas dinámicos con fuentes de ruido de distribu-

ción no-gaussiana. Para esto, se considera la aproximación de la función de distribución no-gaussiana usando una GMM [30]. Considerando funciones de distribución gaussianas en (2.2) para $p_i(y;\gamma_i) = \phi(y;\mu_i,\Sigma_i)$, se tiene que la GMM viene dada por:

$$p(y;\eta) = \sum_{i=1}^{\kappa} \lambda_i \phi(y;\mu_i,\Sigma_i), \tag{2.3}$$

donde $\phi(y;\mu_i,\Sigma_i)$ corresponde a una pdf gaussiana dada por:

$$\phi(y;\mu_i,\Sigma_i) = \frac{1}{(2\pi)^{n_y/2}|\Sigma_i|^{1/2}} \exp\left\{-\frac{1}{2}(y-\mu_i)^T\Sigma_i^{-1}(y-\mu_i)\right\}, \tag{2.4}$$

donde $\mu_i \in \mathbb{R}^{n_y\times 1}$ es la media y $\Sigma_i \in \mathbb{R}^{n_y\times n_y}$ es la matriz de covarianza.
Las GMMs han representado una solución frecuente para aproximar funciones de distribución desconocidas [40]. Basado en el teorema de aproximación de Wiener [41, 42] es posible aproximar cualquier pdf con soporte compacto a través de una suma finita de distribuciones gaussianas [43]. Toda esta formulación se resume a continuación:

**Lema 1.** *Cualquier función de densidad de probabilidad $p(x|\eta)$ de una variable aleatoria n-dimensional $x$, puede ser aproximada en el espacio $L_1(\mathbb{R}^n)$ por una distribución de probabilidad de la siguiente forma:*

$$p(x|\eta) \approx \sum_{i=1}^{\kappa} \lambda_i \phi(x;\mu_i,\Sigma_i), \tag{2.5}$$

*donde $\phi(x;\mu_i,\Sigma_i)$ representa una distribución gaussiana n-dimensional de media $\mu_i$ y varianza $\Sigma_i$, $\kappa \geq 2$, $\lambda_i > 0$ sujeto a $\sum_{i=1}^{\kappa}\lambda_i = 1$.*

*Demostración.* Ver [42, Teorema 3]. □

Es importante señalar que el resultado del Lema 1 puede extenderse a mezcla de funciones de distribución conocidas que tengan traslación. Este hecho se fundamenta en la formulación del problema de Wiener [43] el cual

muestra que para un cierto conjunto $\mathscr{M}$ en el espacio $L(-\infty, \infty)$, toda función de la forma

$$\sum_{i.k} c_{i.k} f_i(x + \xi_{i,k})$$

donde $c_{i,k}$ es cualquier número, $\xi_{i,k}$ es un número real y $f_i \in \mathscr{M}$, pertenece al espacio $L$ y la totalidad de estas funciones representan una combinación lineal en el espacio $L$.

## 2.4. Selección de modelos de mezclas gaussianas

Determinar el número de componentes de una mezcla finita de distribuciones dada por (2.2) representa un problema importante, dado la relación directa del número de componentes de la distribución con los problemas de *identificabilidad* en los modelos de mezclas de distribuciones [34].
Una solución viable corresponde a utilizar los criterios de información tradicionalmente usados para determinar el orden del modelo en la estimación de sistemas dinámicos, por ejemplo, el criterios de información de Akaike (AIC) [44] o el criterio de información bayesiana (BIC) [45]. Sin embargo, las propiedades asintóticas de estros criterios de información dependen de condiciones de regularidad que requieren que el modelo sea identificable. A pesar de esto, el BIC ha sido usado en el caso de mezclas de funciones de distribución gaussianas manteniendo sus propiedades de consistencia [46, 47]. De manera similar, es posible utilizar criterios de información como pendiente heurística (SH) [48–50], criterio de información de desviación (DIC) [51, 52], longitud del último mensaje (MML) [53, 54], entre otros. Para efectos de su aplicación al problema de estimación de mezclas de distribuciones, se diferencian en la forma de escoger el factor de penalización de la función de verosimilitud de acuerdo al número de componentes de la mezcla.
Por otro lado, el problema de selección del orden de una GMM también ha sido tratado con un enfoque de aproximación cuasi-Bayesiano [55, 56]. Típicamente se utilizan técnicas basadas en: i) descartar factores de pon-

deración de la mezcla de distribuciones (*pruning*), ii) similitud entre las distribuciones que componen la mezcla (*joining*) y iii) selección del número de componentes de la mezcla en función de la integral cuadrática del error (*ISE-GMM*) [25].

La técnica de *pruning* consiste en descartar aquellas componentes de la GMM cuyos factores de ponderación, $\lambda_i$, sean menores a un valor umbral $\varepsilon$ y mantener las componentes asociadas que cumplan $\lambda_i \geq \varepsilon$. En este caso, el valor umbral se en base al conocimiento previo del problema asociado a la aproximación usando GMM.

En cuanto a la técnica de *joining* se basa en medir la similitud que existe entre las diferentes componentes que integran la GMM. Para ello, se determina el grado de similitud entre dos componentes $i$ y $j$ de la GMM usando la distancia de Mahalonobis $d_{ij}$ [57]:

$$d_{ij}^2 = \frac{\lambda_i \lambda_j}{\lambda_i + \lambda_j}(\mu_i - \mu_j)^T \Sigma^{-1}(\mu_i - \mu_j), \tag{2.6}$$

donde $\Sigma$ es la varianza combinada de la GMM, $\lambda_i$ y $\lambda_j$ son los factores de ponderación , y $\mu_i$ y $\mu_j$ son las medias de las componentes gaussianas $i$ y $j$ de la GMM. Este tipo de criterio favorece a fusionar aquellas componentes de la GMM asociadas a factores de ponderación bajos con respecto a los que tienen factores de ponderación altos [25]. Por lo tanto, aquellas componentes que tengan una distancia $d_{ij}$ pequeña se pueden fusionar en una distribución gaussiana, $\phi(\cdot; \mu_c, \Sigma_c)$, con factor de ponderación $\lambda_c$ tal que [56, 57]:

$$\lambda_c = \lambda_i + \lambda_j, \tag{2.7}$$

$$\mu_c = \frac{1}{\lambda_i + \lambda_j}\left(\lambda_i \mu_i + \lambda_j \mu_j\right), \tag{2.8}$$

$$\Sigma_c = \frac{1}{\lambda_i + \lambda_j}\left\{\lambda_i \Sigma_i + \lambda_j \Sigma_j + \frac{\lambda_i \lambda_j}{\lambda_i + \lambda_j}(\mu_i - \mu_j)^T(\mu_i - \mu_j)\right\}. \tag{2.9}$$

Por ultimo, la técnica basada en el ISE considera la selección del orden de la GMM considerando la información de toda las componentes de la GMM y resolviendo el siguiente problema de minimización [56]:

$$\eta_r = \arg\min \int_{\mathbb{R}^{n_y}} (p(y;\eta_h) - p(y;\eta_r))^2 dy, \tag{2.10}$$

donde $p(y;\eta_h)$ representa la GMM original con parámetros $\eta_h$ y $p(y;\eta_r)$ representa la GMM con un número reducido de componentes y parametrizada por $\eta_r$. Este tipo de estrategia combina de manera simultanea las ideas de las técnicas de *pruning* y *joining*. El problema de minimización puede ser tratado usando técnicas basadas en el descenso del gradiente [56, 57].

## 2.5. Conclusiones

En este capítulo se mostró la formulación básica de las funciones de distribución de probabilidad modeladas como mezclas finitas de distribuciones conocidas. Se abordaron los fundamentos teóricos que establecen el uso de mezclas finitas de distribuciones gaussianas para aproximar funciones de distribución desconocidas. Para el problema de interés, este enfoque puede usarse para aproximar la función de distribución a *priori* en el enfoque de estimación empírica bayesiana. Finalmente, se mostraron algunas técnicas que permiten seleccionar el orden del modelo para el caso de mezclas de distribuciones gaussianas.

# CAPÍTULO 3
# MÁXIMA VEROSIMILITUD USANDO MODELOS DE MEZCLAS GAUSSIANAS

## 3.1. Introducción

En este capítulo se muestra la relación entre la técnica de máxima verosimilitud y la estimación empírica bayesiana. Seguidamente se formula el algoritmo de estimación con enfoque empírico bayesiano usando mezclas de distribuciones gaussianas. Finalmente se muestran las conclusiones.

## 3.2. Método de máxima verosimilitud y estimación empírica bayesiana

La estimación por máxima verosimilitud (ML) es una técnica que permite obtener una estimación de los parámetros que definen un modelo. Los parámetros se determinan tal que maximizan la probabilidad de que el proceso descrito por el modelo puedan generar los datos observados [58]. Suponiendo que se tiene un conjunto, $\mathscr{D}$[1], de observaciones independientes e identicamente distribuidas (i.i.d) de una variable aleatoria $y$ con una pdf, $p(y|\theta)$, entonces la función de verosimilitud viene dada por:

$$\mathcal{L}(\theta) = p(\mathscr{D}|\theta) = \prod_{t=1}^{N} p(y_t|\theta). \tag{3.1}$$

Típicamente se utiliza el logaritmo de la función de verosimilitud como:

$$\ell(\theta) = \sum_{t=1}^{N} \log\left[p(y_t|\theta)\right]. \tag{3.2}$$

[1] $\mathscr{D}$ se refiere a un conjunto de $N$ mediciones de la variable aleatoria y se denota como $\mathscr{D} = y_{1:N} = \{y_1, \ldots, y_N\}$.

Entonces, el estimador de máxima verosimilitud se obtiene resolviendo:

$$\hat{\theta}_{\mathrm{ML}} = \arg \max_{\theta} \ell(\theta). \tag{3.3}$$

En algunos casos es posible obtener una expresión explícita para el estimador de ML en función del conjunto de observaciones, sin embargo, muchas veces hay que utilizar algoritmo numéricos de optimización para resolver el problema de estimación [59].

Por otro lado, la estimación por ML corresponde a un caso particular del problema de estimación bayesiana basado en la maximización de la función de distribución a *posteriori* (MAP) de un vector de parámetros estocástico $\theta$. Este se puede formular como:

$$\hat{\theta}_{\mathrm{MAP}} = \arg \max_{\theta} \sum_{t=1}^{N} \log \left[p(y_t|\theta)\right] + \log[p(\theta)], \tag{3.4}$$

donde $p(\theta)$ corresponde a la función de distribución a *priori* del vector de parámetros $\theta$. Es importante resaltar que el estimador en (3.4) combina la información de la función de verosimilitud y la función de distribución $p(\theta)$. Además, asumir una función de distribución a *priori* difusa, es decir $p(\theta)$ como una distribución uniforme, corresponde al estimador por ML [1, 2].

Tradicionalmente, la selección de la función de distribución a *priori* en el problema de estimación MAP se hace en base a información conocida acerca del proceso en estudio o bien sobre la experiencia en el área de aplicación [20–22].

Por otro lado, es posible usar el enfoque de estimación EB, el cual utiliza la técnica de ML con el fin de obtener los parámetros que definen la función de distribución a *priori* [23, 24]. La formulación consiste en considerar la variable estocástica desconocida $\theta$ como una variable escondida o latente (*hidden variable*), y tiene una función de distribución, $p(\theta|\eta)$, definida por un vector de parámetros $\eta$. Entonces, la función de verosimilitud se puede

obtener marginalizando la variable latente como:

$$p(\mathscr{D}|\eta) = \int_{-\infty}^{\infty} p(\mathscr{D}|\theta)p(\theta|\eta)d\theta. \tag{3.5}$$

El problema de estimación EB puede resolverse usando el algoritmo tradicional de maximización de la esperanza [60]. Sin embargo, es necesario considerar una suposición inicial sobre la estructura de la pdf a *priori*, $p(\theta|\eta)$, donde la exactitud de la estimaciones se compromete cuando la suposición inicial no corresponde al comportamiento verdadero de $p(\theta)$ [30].

## 3.3. Estimación empírica bayesiana usando modelos de mezclas gaussianas

### 3.3.1. Formulación del estimador por máxima verosimilitud

El objetivo principal es caracterizar la función de distribución $p(\theta|\eta)$ en (3.5) con una estructura de GMM en (2.5) usando el enfoque de estimación por ML. Para esto, el vector de parámetros $\theta$ es considerado como una variable escondida (*hidden variable*) cuya pdf esta parametrizada por un vector de parámetros determinísticos $\eta$ que definen la GMM. Entonces, el vector de parámetros a estimar viene dado por:

$$\eta = [\underbrace{\lambda_1, \mu_1, \Sigma_1}_{\eta_1}, \ldots, \underbrace{\lambda_\kappa, \mu_\kappa, \Sigma_\kappa}_{\eta_\kappa}]. \tag{3.6}$$

De manera similar, se define $\eta_o$ como el vector de parámetros verdadero. Por otro lado, se asume que se realizan $M$ experimentos independientes tal que el proceso estocástico $\theta$ toma diferentes valores $\theta^{[r]}$, $r = 1, \ldots, M$ en cada experimento. Además, se considera que se tiene un conjunto de mediciones, $\mathscr{D}$, diferente para cada experimento, es decir, $\mathscr{D}^{[r]}$, $r = 1, \ldots, M$. Este tipo de formulación ha sido utilizada de manera similar para el modelado de la incertidumbre en sistemas dinámicos [61, 62].

Asumiendo que $\mathscr{D}^{[r]}$ es un conjunto de mediciones independientes e idénticamente distribuidas (i.i.d) para cada experimento, se obtiene la función de verosimilitud como:

$$\mathcal{L}(\eta) = p(\mathscr{D}^{[1]}, \ldots, \mathscr{D}^{[M]}|\eta) \tag{3.7}$$

$$= \prod_{r=1}^{M} \int_{-\infty}^{\infty} p(\mathscr{D}^{[r]}|\theta^{[r]}, \eta) p(\theta^{[r]}|\eta) d\theta^{[r]}. \tag{3.8}$$

donde $\mathcal{L}(\eta)$ es la función de verosimilitud. Considerando que $p(\theta^{[r]}|\eta)$ en (3.8) es una GMM, entonces se obtiene:

$$\mathcal{L}(\eta) = \prod_{r=1}^{M} \sum_{i=1}^{\kappa} \int_{-\infty}^{\infty} p(\mathscr{D}^{[r]}|\theta^{[r]}, \eta) \lambda_i \phi(\theta^{[r]}; \mu_i, \Sigma_i) d\theta^{[r]}, \tag{3.9}$$

donde $\phi(\cdot)$ es una función de distribución gaussiana de media $\mu_i$, matriz de covarianza $\Sigma_i$ y factor de ponderación $\lambda_i$. Entonces, el logaritmo de la función de verosimilitud es el siguiente:

$$\ell(\eta) = \sum_{r=1}^{M} \log \left[ \sum_{i=1}^{\kappa} \lambda_i \int_{-\infty}^{\infty} p(\mathscr{D}^{[r]}|\theta^{[r]}, \eta) \phi(\theta^{[r]}; \mu_i, \Sigma_i) d\theta^{[r]} \right]. \tag{3.10}$$

Finalmente, el estimador de máxima verosimilitud se obtiene resolviendo:

$$\hat{\eta}_{\mathrm{ML}} = \arg \max_{\eta} \ \ell(\eta) \quad \text{sujeto a} \ \ 0 \leq \lambda_i \leq 1, \quad \sum_{i=1}^{\kappa} \lambda_i = 1. \tag{3.11}$$

**Observación 1.** *El vector de parámetros $\eta = \{\lambda_i, \mu_i, \Sigma_i\}_{i=1}^{\kappa}$ no depende del índice $r$. La notación $\theta^{[r]}$ indica que la realización de $\theta$ depende del experimento $r$, pero todas la realizaciones están definidas en el mismo espacio de probabilidad con la misma función de distribución parametrizada por $\eta$.* $\bigtriangledown$

**Observación 2.** *Se asume que el vector de parámetros verdadero, $\eta_0$, satis-*

*face las condiciones de regularidad que garantizan que la solución $\hat{\eta}_{ML}$ del problema de optimización en* (3.11) *converge (en probabilidad casi segura) a la solución verdadera $\eta_o$ cuando $N \to \infty$.* $\bigtriangledown$

**Observación 3.** *Es importante resaltar que formulación de la estimación por ML para el problema de interés en* (3.9) *permite obtener un resultado de estimación del vector de parámetros que define la GMM usando de manera conjunta la información de $M$ experimentos independientes.* $\bigtriangledown$

### 3.3.2. Función de verosimilitud usando modelos de mezclas gaussianas

El problema de estimación por ML se obtiene encontrando el vector de parámetros que maximiza la función de verosimilitud en (3.10). La solución puede ser difícil de obtener cuando se incrementa el número de componentes de la GMM, dado que la función de verosimilitud puede tener máximos locales. Además, el problema de estimación usando GMMs permite obtener múltiples soluciones de acuerdo al número de componentes de la GMM y las permutaciones de las mismas [63]. Esto implica que se debe manejar restricciones sobre los parámetros que definen la GMM, por ejemplo, ordenar de forma ascendente los factores de ponderación $\lambda_i$ de la GMM.
Para ilustrar el comportamiento de la función de verosimilitud en (3.10), la cual va a depender de la cantidad de componentes $\kappa$ de la GMM y de la cantidad de experimentos $M$, se considera que la función de distribución a *priori* esta definida por una GMM de cuatro componentes ($\kappa$=4). Los valores de los parámetros corresponden a igual factor de ponderación $\lambda_i =$ 0,25 para cada componente, igual varianza $\Sigma_i = 10$, y las medias vienen dadas por $\mu_1 = -5$, $\mu_2 = 5$, $\mu_3 = -15$ y $\mu_4 = 15$. La Figura 3.1 muestra la función de verosimilitud para $M = 1$ y variaciones en los valores de las medias de dos componentes de la GMM ($\mu_1$ y $\mu_4$) considerando que el

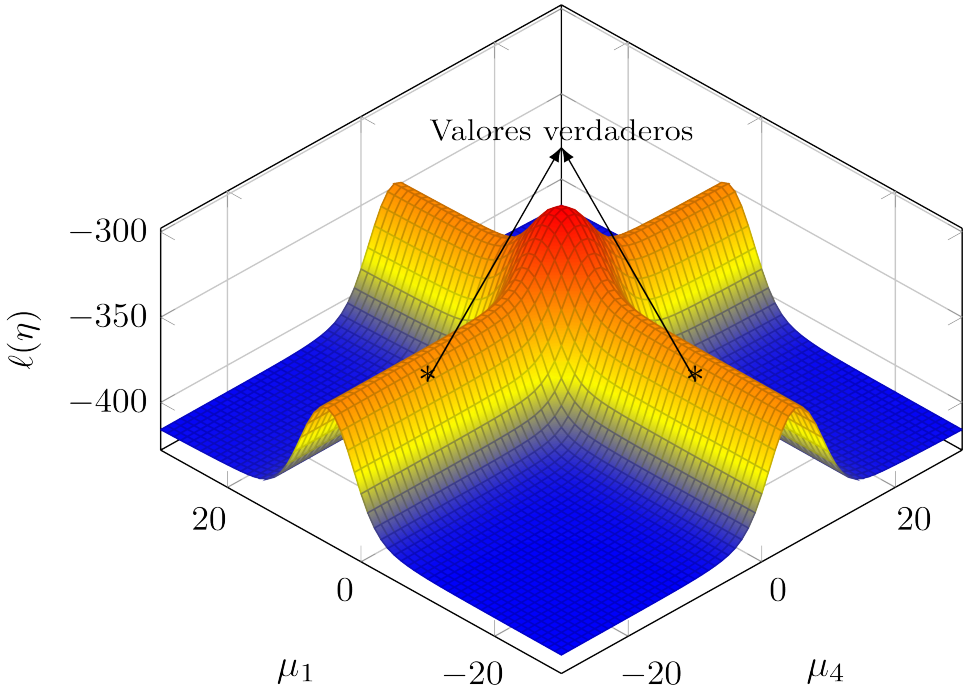



Figura 3.1: Logaritmo de la función de verosimilitud para $M = 1$ y variaciones de las medias $\mu_1$ y $\mu_4$.

resto de los parámetros permanecen en sus valores verdaderos. Se puede observar la presencia de máximos locales, además se muestra la dificultad de obtener una buena estimación usando la información de un experimento simple ($M = 1$) dado que el punto máximo en la función de verosimilitud no corresponde a los valores verdaderos.

De manera similar la Figura 3.2 muestra la función de verosimilitud para $M = 20$. De igual manera se muestra la presencia de máximos locales, además de los puntos óptimos que producen el mismo valor del logaritmo de la función de verosimilitud debido a la permutación de las componentes. Sin embargo, ambos puntos corresponden a una solución del estimador por ML para las componentes de la GMM. Para resolver este tipo de problema es usar técnicas de optimización global, sin embargo esto implica un mayor costo computacional. Por otro lado, la exactitud de las estimaciones usando técnicas de optimización local, por ejemplo algoritmos de EM, está sujeto a los valores de inicialización de los parámetros que definen la GMM,

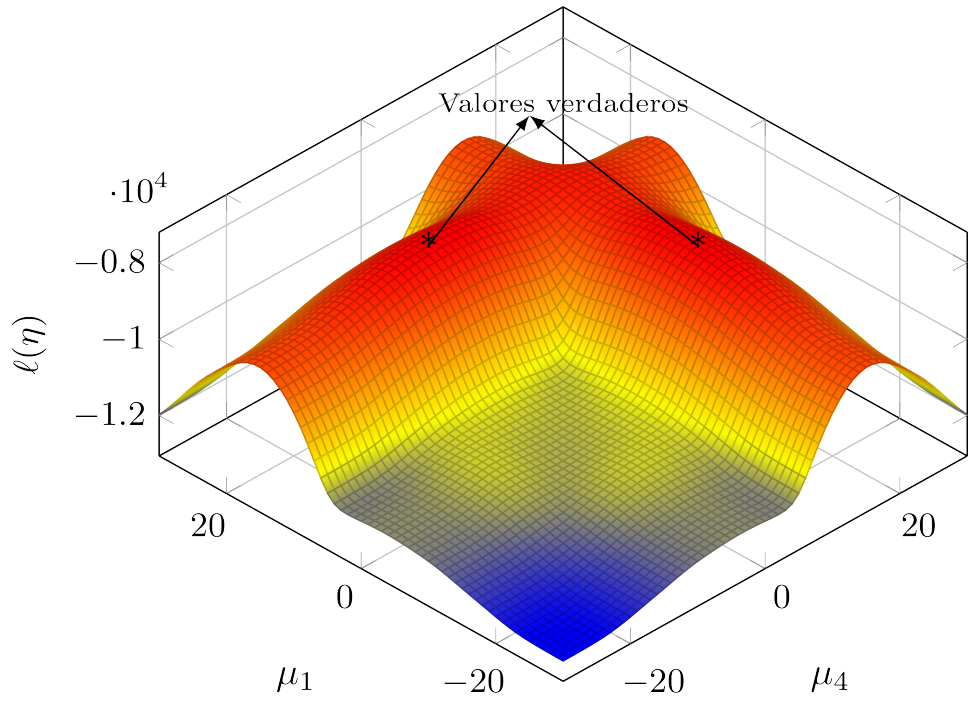



Figura 3.2: Logaritmo de la función de verosimilitud para $M = 20$ y variaciones de las medias $\mu_1$ y $\mu_4$.

los cuales condicionan el desempeño y velocidad de convergencia de estos algoritmos, ver por ejemplo [64–67].

## 3.4. Conclusiones

En este capítulo se mostró la relación entre la técnica de máxima verosimilitud y el enfoque empírico bayesiano para obtener una estimación de la función de distribución a *priori* en el problema clásico de inferencia estadística. Además, se mostró la formulación del problema de estimación empírica bayesiana usando una GMM para definir la estructura de la función de distribución a *priori*. Para ello, se consideró usar de manera simultanea la información de un conjunto finito de experimentos independientes a fin de obtener un estimador por ML con restricciones dadas por la estructura de la GMM. Finalmente, se analizó el comportamiento de la función de verosimilitud para la estrategia de estimación EB usando GMM. Se observó que la presencia de máximos locales condicionan la exactitud de las

estimaciones, por lo que es necesario establecer una buena inicialización para usar técnicas de optimización convencionales.

# CAPÍTULO 4
# ALGORITMO BASADO EN MAXIMIZACIÓN DE LA ESPERANZA USANDO MODELOS DE MEZCLAS GAUSSIANAS

## 4.1. Introducción

En este capítulo se muestra la formulación clásica del algoritmo EM para el problema de interés. Seguidamente, se propone un algoritmo basado en EM usando GMMs para obtener la función de distribución a *priori* en el problema de estimación empírica bayesiana. Se muestra la forma de obtener los estimadores de los parámetros que definen la GMM.

## 4.2. Algoritmo de estimación de maximización de la esperanza (EM)

El algoritmo de maximización de la esperanza (EM) ha sido usado para identificar diferentes sistemas, por ejemplo, sistemas continuos usando datos muestreados [68], sistemas con datos cuantizados [69–75], sistemas en espacio de estados usando datos incompletos [76], telecomunicaciones [77], modelado de la incertidumbre en sistemas dinámicos [78], sistemas estáticos [3, 29, 33], sistemas bilineales en espacio de estado [79] y sistemas no lineales [80], entre otros.

El objetivo principal del algoritmo EM es obtener el estimador por ML en (3.11) del parámetro $\eta$ utilizando el enfoque de datos aumentados o datos completos [60]. En este caso, los datos completos están formados por el conjunto de mediciones $\mathscr{D}^{[r]}$ y el conjunto de datos no observados (o variable escondida) $\theta$. El procedimiento iterativo para obtener una estimación consiste en construir una función auxiliar, $\mathcal{Q}(\eta, \hat{\eta}^{(k)})$, en función de los da-

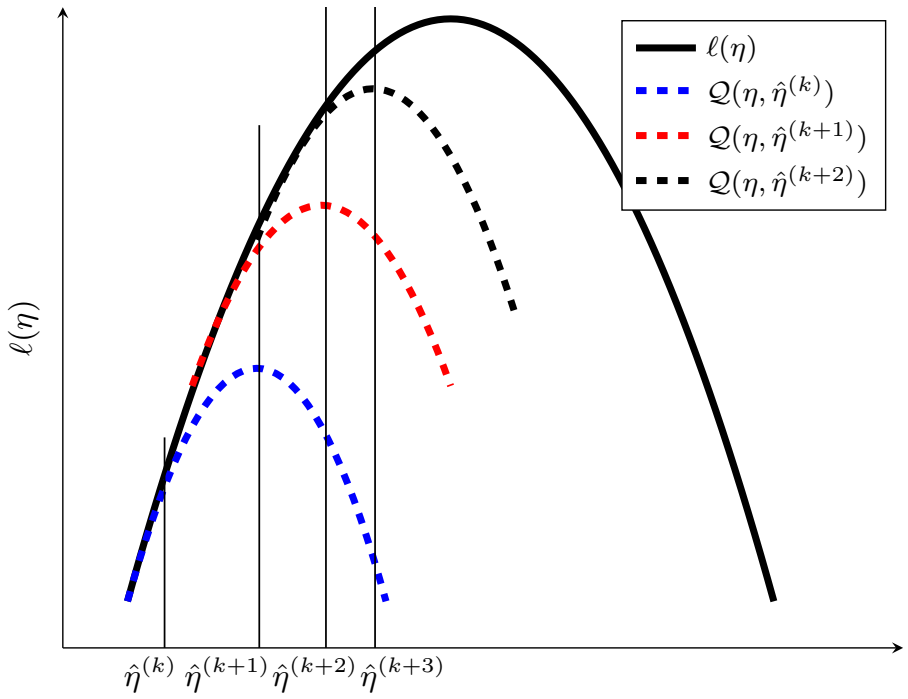



Figura 4.1: Iteraciones del algoritmo de maximización de la esperanza (EM).

tos completos y la estimación actual $\hat{\eta}^{(k)}$ (*E-step*), y seguidamente obtener una nueva estimación $\hat{\eta}^{(k+1)}$ maximizando la función auxiliar (*M-step*) [60]. La Figura 4.1 muestra el procedimiento iterativo del algoritmo EM. Para el problema de interés en (3.11), el algoritmo iterativo EM viene dado por:

E-step:

$$\mathcal{Q}(\eta, \hat{\eta}^{(k)}) = \mathbb{E}\left\{ \sum_{r=1}^{M} \log \left[ \sum_{i=1}^{\kappa} \lambda_i \int_{-\infty}^{\infty} p(\mathscr{D}^{[r]} | \theta^{[r]}, \eta) \phi(\theta^{[r]}; \mu_i, \Sigma_i) d\theta^{[r]} \right] \middle| \mathscr{D}^{[r]}, \hat{\eta}^{(k)} \right\}, \tag{4.1}$$

M-step:

$$\hat{\eta}^{(k+1)} = \arg\max_{\eta} \ \mathcal{Q}(\eta, \hat{\eta}^{(k)}) \quad \text{sujeto a} \quad 0 \leq \lambda_i \leq 1, \quad \sum_{i=1}^{\kappa} \lambda_i = 1. \tag{4.2}$$

Es importante resaltar que obtener la función auxiliar en (4.1) tiene un alto grado de dificultad, dado que requiere resolver el logaritmo de una sumatoria de ecuaciones integrales en el problema de optimización [29].

Además, la complejidad varía a medida que se incrementa las componentes de la GMM dado que se aumenta el tamaño del espacio de búsqueda [30]. Por todo esto, en la siguiente sección se propone una metodología basada en la formulación mostrada en [3] con el fin de obtener un algoritmo basado en EM usando GMM.

## 4.3. Formulación del algoritmo basado en EM usando modelos de mezclas gaussianas

Para la formulación del algoritmo de estimación, usando (3.10) se define lo siguiente:

$$K(\theta^{[r]}, \eta_i) = \lambda_i p(\mathscr{D}^{[r]}|\theta^{[r]}, \eta)\phi(\theta^{[r]}; \mu_i, \Sigma_i). \tag{4.3}$$

Entonces, el logaritmo de la función de verosimilitud en (3.10) se puede re-escribir como:

$$\ell(\eta) = \sum_{r=1}^{M} \log[\mathcal{V}^{[r]}(\eta)], \tag{4.4}$$

donde

$$\mathcal{V}^{[r]}(\eta) = \sum_{i=1}^{\kappa} \int_{-\infty}^{\infty} K(\theta^{[r]}, \eta_i) d\theta^{[r]}. \tag{4.5}$$

Una formulación similar ha sido estudiada en diversas aplicaciones asociadas a la optimización de un funcional de costo dado por una ecuación integral con una variable escondida [3,29,33]. En estos trabajos se formulan algoritmos basados en EM [60] definiendo una función auxiliar adecuada para el funcional de costo integral. Definiendo $\mathcal{B}^{[r]}(\eta) = \log[\mathcal{V}^{[r]}(\eta)]$, es posible obtener [3]:

$$\mathcal{B}^{[r]}(\eta) = \mathcal{Q}^{[r]}(\eta, \hat{\eta}^{(k)}) - \mathcal{H}^{[r]}(\eta, \hat{\eta}^{(k)}), \tag{4.6}$$

donde

$$\mathcal{Q}^{[r]}(\eta, \hat{\eta}^{(k)}) = \sum_{i=1}^{\kappa} \int_{-\infty}^{\infty} \log[K(\theta^{[r]}, \eta_i)] \frac{K(\theta^{[r]}, \hat{\eta}_i^{(k)})}{\mathcal{V}^{[r]}(\hat{\eta}^{(k)})} d\theta^{[r]}, \tag{4.7}$$

$$\mathcal{H}^{[r]}(\eta, \hat{\eta}^{(k)}) = \sum_{i=1}^{\kappa} \int_{-\infty}^{\infty} \log \left[ \frac{K(\theta^{[r]}, \eta_i)}{\mathcal{V}^{[r]}(\eta)} \right] \frac{K(\theta^{[r]}, \hat{\eta}_i^{(k)})}{\mathcal{V}^{[r]}(\hat{\eta}^{(k)})} d\theta^{[r]}. \tag{4.8}$$

Con el objetivo de construir la función auxiliar para el algoritmo iterativo, se tiene el siguiente resultado:

**Lema 2.** *La función $\mathcal{H}^{[r]}(\eta, \hat{\eta}^{(k)})$ en (4.8) es una función decreciente para cualquier valor de $\eta$ y satisface la siguiente desigualdad:*

$$\mathcal{H}^{[r]}(\eta, \hat{\eta}^{(k)}) - \mathcal{H}^{[r]}(\hat{\eta}^{(k)}, \hat{\eta}^{(k)}) \leq 0 \tag{4.9}$$

*Demostración.* Ver en [3, Sección 3.1]. □

Finalmente, del Lema 2 se obtiene un algoritmo iterativo basado en EM de la siguiente forma:

$$\bar{\mathcal{Q}}(\eta, \hat{\eta}^{(k)}) = \sum_{r=1}^{M} \mathcal{Q}^{[r]}(\eta, \hat{\eta}^{(k)}), \tag{4.10}$$

$$\hat{\eta}^{(k+1)} = \arg \max_{\eta} \ \bar{\mathcal{Q}}(\eta, \hat{\eta}^{(k)}), \tag{4.11}$$

sujeto a $0 \leq \lambda_i \leq 1, \ \sum_{i=1}^{\kappa} \lambda_i = 1$.

Es importante señalar que (4.10) y (4.11) están relacionados con la formulación clásica (*E-step* y *M-step* respectivamente) del algoritmo iterativo de EM [60]. Por otro lado, la formulación presentada en [3] es una variación del algoritmo EM que no se limita a funciones de distribución para resolver problemas de estimación por máxima verosimilitud, en donde el problema de estimación puede tener una solución explícita.

**Observación 4.** *El algoritmo propuesto difiere de las soluciones clási-*

*cas con EM usando GMM donde el vector de parámetros desconocido $\theta$ es determinístico. En este caso, $\theta$ es un proceso aleatorio para el cual se considera aproximar $p(\theta)$ a través de una GMM usando conjuntamente la información de un número finito $M$ de experimentos.* $\bigtriangledown$

## 4.4. Estimadores para los parámetros del modelo de mezclas gaussianas

Para la optimización de la función auxiliar $\bar{\mathcal{Q}}(\eta, \hat{\eta}^{(k)})$ en (4.10) se pueden obtener estimadores en forma cerrada para el vector de parámetros $\eta$ que define la GMM. De manera similar a [3, 29], la optimización con respecto a $\eta$ se obtiene de la siguiente manera:

**Teorema 1.** *El vector de parámetros $\hat{\eta}$ que optimiza la función auxiliar $\bar{\mathcal{Q}}(\eta, \hat{\eta}^{(k)})$ en (4.10) con respecto a $\eta$ viene dado por:*

$$\hat{\lambda}_i^{(k+1)} = \frac{\mathcal{P}(\theta, \hat{\eta}_i^{(k)})}{\sum_{l=1}^{\kappa} \mathcal{P}(\theta, \hat{\eta}_l^{(k)})} \tag{4.12}$$

$$\hat{\mu}_i^{(k+1)} = \frac{\mathcal{M}(\theta, \hat{\eta}_i^{(k)})}{\mathcal{P}(\theta, \hat{\eta}_i^{(k)})} \tag{4.13}$$

$$\hat{\Sigma}_i^{(k+1)} = \frac{\mathcal{S}(\theta, \hat{\eta}_i^{(k)})}{\mathcal{P}(\theta, \hat{\eta}_i^{(k)})} \tag{4.14}$$

*donde*

$$\mathcal{P}(\theta, \hat{\eta}_i^{(k)}) = \sum_{r=1}^{M} \int_{-\infty}^{\infty} \frac{K(\theta^{[r]}, \hat{\eta}_i^{(k)})}{\mathcal{V}^{[r]}(\hat{\eta}^{(k)})} d\theta^{[r]}, \tag{4.15}$$

$$\mathcal{M}(\theta, \hat{\eta}_i^{(k)}) = \sum_{r=1}^{M} \int_{-\infty}^{\infty} \theta^{[r]} \frac{K(\theta^{[r]}, \hat{\eta}_i^{(k)})}{\mathcal{V}^{[r]}(\hat{\eta}^{(k)})} d\theta^{[r]}, \tag{4.16}$$

$$\mathcal{S}(\theta, \hat{\eta}_i^{(k)}) = \sum_{r=1}^{M} \int_{-\infty}^{\infty} (\theta^{[r]} - \hat{\mu}_i^{(k)})(\theta^{[r]} - \hat{\mu}_i^{(k)})^T \frac{K(\theta^{[r]}, \hat{\eta}_i^{(k)})}{\mathcal{V}^{[r]}(\hat{\eta}^{(k)})} d\theta^{[r]}. \tag{4.17}$$

*Demostración.* Ver el Apéndice A. □

Finalmente, el algoritmo propuesto se resume en los pasos mostrados en la Figura 4.2. Cabe resaltar que los estimadores del Teorema 1 se pueden aproximar de forma numérica usando métodos de integración por Monte Carlo, ver por ejemplo [81, 82].

## 4.5. Métodos de integración numérica por Monte Carlo

En general, el área de inferencia estadística está relacionado en la resolución de problemas de optimización y estimación asociados a ecuaciones integrales [81], por ejemplo, problemas de estimación con enfoque de ML [71,73,74] o bien formulación de algoritmos de estimación bayesiana [3, 23, 24]. En muchos de estos casos no es posible obtener una solución analítica del problema, por lo que es necesario estudiar técnicas que permitan aproximar numéricamente este tipo ecuaciones integrales.

La integración por Monte Carlo (MC) es una técnica de integración que permite aproximar numéricamente una ecuación integral. El cálculo se fundamenta en usar un conjunto de puntos aleatorios en los cuales el integrando es evaluado [81,83]. En la literatura existen diversas técnicas de integración por MC, por ejemplo *importance sampling* [84], métodos secuenciales de MC (*Particle filtering*) [85], aproximación por suma de Riemann [86], entre otros.

Una metodología clásica de integración por MC corresponde a la propuesta en [87], el cual establece una solución para evaluar una integral dada por:

$$\mathbb{E}\{h(X)\} = \int_{-\infty}^{\infty} h(x)p(x)dx, \tag{4.18}$$

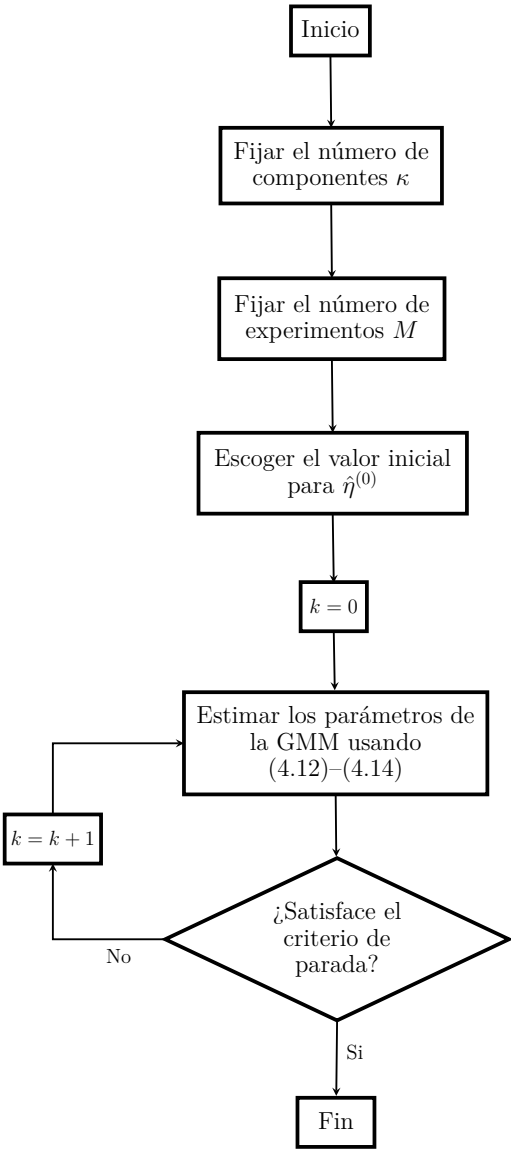



Figura 4.2: Algoritmo iterativo para estimar los parámetros de la GMM.

donde $p(x)$ es la pdf de la variable aleatoria $X$ y $h(x)$ es una función de la variable aleatoria $X$. Utilizando un conjunto finito de muestras $(X_1, \ldots, X_m)$ de la función de distribución de probabilidad definida por $p(x)$, se aproxima numéricamente la integral como:

$$\bar{h}_m^{(MH)} = \frac{1}{m} \sum_{j=1}^{m} h(X_j). \tag{4.19}$$

Por otro lado, la metodología de *importance sampling* o también llamado

muestreo ponderado, establece una aproximación similar usando las llamadas "funciones de importancia"(*importance functions*) [88]. Este método se basa en una representación alternativa de (4.18) como:

$$\mathbb{E}\left\{h(X)\right\} = \int_{-\infty}^{\infty} h(x)\frac{p(x)}{g(x)}g(x)dx, \tag{4.20}$$

donde $g(x)$ es una función de distribución dada. Entonces usando un conjunto finito de muestras $(X_1, \ldots, X_m)$ de la pdf definida por $g(x)$, la integral (4.20) se aproxima como:

$$\bar{h}_m^{(IS)} = \frac{1}{m}\sum_{j=1}^{m}\frac{p(X_j)}{g(X_j)}h(X_j). \tag{4.21}$$

Es importante resaltar que este método establece muy pocas restricciones en cuanto a la selección de la distribución instrumental $g(x)$, por lo que pueden elegirse distribuciones fáciles de simular. Además es posible usar la misma muestra generada por $g(x)$ la cual la hace una técnica robusta de aproximación [81]. Por otro lado, los estimadores (4.19) y (4.21) convergen a la integral en (4.18) con probabilidad casi segura por la ley de los grandes números.

## 4.6. Conclusiones

En este capítulo se mostró una metodología para obtener una estimación de la función de distribución a *priori* como una mezcla finita de distribuciones gaussianas. Para ello se utiliza el enfoque de estimación empírica bayesiana con el fin estimar los parámetros que definen la GMM. Se formuló un algoritmo basado en EM en donde se plantea una metodología para obtener una función auxiliar que considere la información conjunta de un número finito de experimentos independientes. Además, se obtienen estimadores en forma cerrada para cada uno de los parámetros que definen la GMM los cuales pueden ser calculados usando métodos de integración numérica.

# CAPÍTULO 5
# EJEMPLOS NUMÉRICOS DE SIMULACIÓN

## 5.1. Introducción

En este capítulo se muestra el desempeño del algoritmo propuesto a través de un ejemplo numérico de simulación. Se considera un sistema estocástico en el cual se quiere aproximar la distribución a *priori* del parámetro desconocido como una GMM. Se analiza la inicialización del algoritmo propuesto y la selección del número de componentes de la GMM. Finalmente, se obtiene la función de distribución a *posteriori* con el fin de comparar el desempeño del algoritmo propuesto con el enfoque clásico de usar una función de distribución a *priori* difusa.

## 5.2. Descripción general

En esta sección se muestra la descripción del sistema para ilustrar el funcionamiento del algoritmo propuesto con el enfoque de EB. Para esto se considera que las mediciones $\mathscr{D}^{[r]} = y_{1:N}^{[r]}$ vienen dadas por el siguiente sistema estocástico:

$$y_t^{[r]} = \theta^{[r]} u_{t-1} + \omega_t^{[r]}, \tag{5.1}$$

donde $t = 1, \ldots, N$, $r = 1, \ldots, M$, la señal de entrada determinística $u_t \sim \mathscr{N}(0, \sigma_u^2)$, $\sigma_u^2 = 10$, $\omega_t \sim \mathscr{N}(0, \sigma_\omega^2)$, $\sigma_\omega^2 = 1$ es una señal de ruido blanco y $\theta^{[r]}$ es un vector de parámetros desconocidos con una función de distribución no-gaussiana. El ruido $\omega_t^{[r]}$ representa una realización diferente para cada experimento independiente $r$.
El vector de parámetros a estimar es $\eta = \{\lambda_i, \mu_i, \Sigma_i\}_{i=1}^{\kappa}$. La simulación se lleva a cabo de la siguiente forma:

(a) Se utiliza la aproximación clásica de MC en (4.18) para calcular los

estimadores de la GMM.

(b) El número de datos es $N = 500$.

(c) El número de simulaciones de Monte Carlo[1] (MC) son 100.

(d) El criterio de convergencia del algoritmo viene dado por:

$$\left\|\hat{\eta}^{(k)} - \hat{\eta}^{(k-1)}\right\| / \left\|\hat{\eta}^{(k)}\right\| < 1 \times 10^{-6}$$

o se alcance el número máximo de 100 iteraciones del algoritmo EM.

Se analiza una estrategia de inicialización del algoritmo propuesto y se establece un criterio de selección para el número de componentes de la GMM. Adicionalmente, se obtiene la función de distribución a *posteriori*, $p(\theta|\mathscr{D})$, usando (1.1) tal que:

$$p(\theta|\mathscr{D}) \propto p(\mathscr{D}|\theta)p(\theta). \tag{5.2}$$

En este sentido, se comparan las distribuciones a *posteriori* obtenidas cuando se considera la aproximación de una GMM para $p(\theta)$ con al algoritmo propuesto y el enfoque clásico de una distribución $p(\theta)$ difusa (*difusse prior*) [1, 2].

## 5.3. Ejemplo 1: Estimación de una distribución a *priori* dada por una GMM.

Para este caso se considera que la función de distribución *verdadera* de $\theta$ viene dada por una GMM para ($\kappa = 2$):

$$p(\theta)^{(\text{True})} = \lambda_1 \phi(\theta; \mu_1, \Sigma_1) + \lambda_2 \phi(\theta; \mu_2, \Sigma_2), \tag{5.3}$$

[1] Cada simulación de Monte Carlo corresponde a una estimación obtenida usando $M$ experimentos independientes.

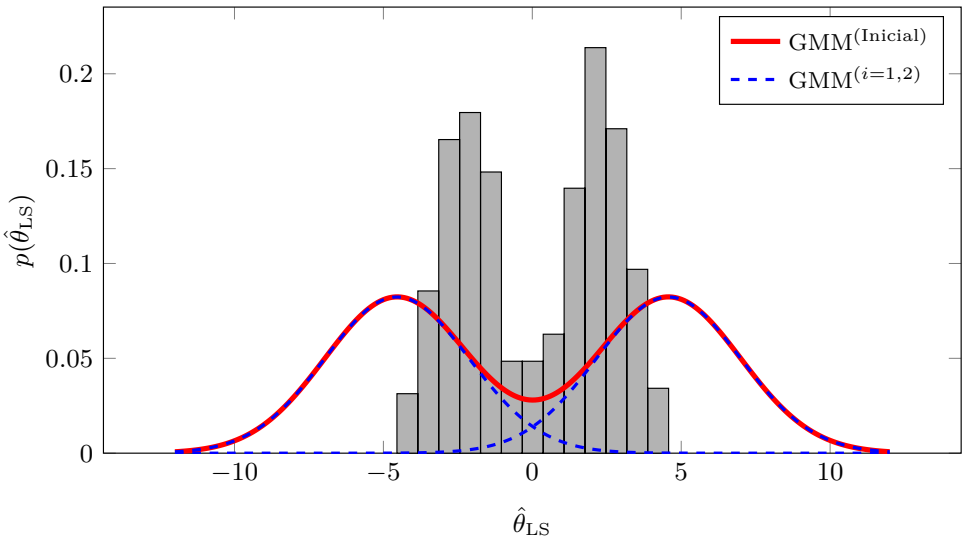



Figura 5.1: Histograma de la estimaciones de $\hat{\theta}_{\text{LS}}$ obtenidas para el Ejemplo 1 y la GMM inicial para $\kappa = 2$.

donde $\lambda_1 = \lambda_2 = 0{,}5$, $\Sigma_1 = \Sigma_2 = 1$, $\mu_1 = -2{,}2$ y $\mu_2 = 2{,}2$. Los datos de esta distribución son generados con *Slice Sampler* [89]. Se utiliza la aproximación clásica de MC en (4.18) para calcular los estimadores de la GMM.

### 5.3.1. Inicialización del algoritmo propuesto

En primer lugar es necesario obtener una inicialización para los parámetros que definen la GMM. Para ello se asume que el parámetro $\theta$ es determinístico y se obtienen estimaciones para cada experimento independiente utilizando $M = 500$. Para las estimaciones del parámetro se utiliza la técnica ML, el cual para este caso particular corresponde al estimador por mínimos cuadrados, $\hat{\theta}_{\text{LS}}$. En la Figura 5.1 las barras grises representan el histograma para las estimaciones obtenidas usando mínimos cuadrados. Las lineas segmentadas azules y la línea roja representan las componentes de la mezcla gaussiana y la GMM respectivamente, para $\kappa = 2$. Los valores iniciales para $\eta^{(0)}$, en función de la cantidad de componentes $\kappa$, se escogen de la siguiente forma:

- Cada componente $\Sigma_i$ se inicializa en el valor de la covarianza muestral

Tabla 5.1: Estimaciones de la GMM para el Ejemplo 1 usando $M = 200$ experimentos y diferentes valores de $\kappa$

| Componentes de GMM | $\kappa = 2$ | | | $\kappa = 3$ | | | $\kappa = 4$ | | |
|---|---|---|---|---|---|---|---|---|---|
| | $\lambda_i$ | $\mu_i$ | $\Sigma_i$ | $\lambda_i$ | $\mu_i$ | $\Sigma_i$ | $\lambda_i$ | $\mu_i$ | $\Sigma_i$ |
| $i = 1$ | 0,4678 | −2,1914 | 1,0636 | 0,4864 | −2,0782 | 1,2670 | 0,2944 | −2,6280 | 0,6173 |
| $i = 2$ | 0,5322 | 2,2046 | 1,0892 | 0,2012 | 2,3929 | 0,3945 | 0,1434 | −1,2590 | 1,3018 |
| $i = 3$ | – | – | – | 0,3124 | 2,2540 | 0,6261 | 0,4746 | 2,3962 | 0,4261 |
| $i = 4$ | – | – | – | – | – | – | 0,0877 | −0,0846 | 1,9988 |

de las estimaciones $\hat{\theta}_{\text{LS}}$.

- Las medias de las componentes de la GMM, $\mu_i$, se inicializan en los valores equiespaciados entre el valor máximo y mínimo de las estimaciones $\hat{\theta}_{\text{LS}}$.

- Los factores de ponderación de la GMM, $\lambda_i$, se escogen de igual valor tal que $\sum_{i=1}^{\kappa} \lambda_i = 1$.

### 5.3.2. Selección del número de componentes de la GMM

Para la selección del orden la GMM, se obtienen las estimaciones usando el algoritmo propuesto para diferentes valores de $\kappa$ para la GMM. Para ello se fija un número de experimentos en $M = 200$. La Figura 5.2 muestra las distribuciones estimadas usando GMM para valores de $\kappa = \{2, 3, 4\}$. Se observa como las estimaciones obtenidas logran ajustar una estructura de una GMM de dos componentes.

La Tabla 5.1 muestra los valores de la estimaciones para los valores de $\kappa = \{2, 3, 4\}$. Para $\kappa = 3$, se analiza la similitud entre las componentes $i = \{2, 3\}$ usando la metodología *joining* [25] para la selección del número de componentes de la GMM. Entonces, se encuentra la distancia de Mahalonobis [57] de acuerdo a (2.6) obteniendo $d_{23}^2 = 4{,}15 \times 10^{-4}$. Por lo tanto, de acuerdo a este criterio de selección es posible fusionar los pares de distribuciones consideradas estableciendo una GMM de dos componentes, $\kappa = 2$, definida por (2.7)–(2.9). Se obtiene un factor de ponderación

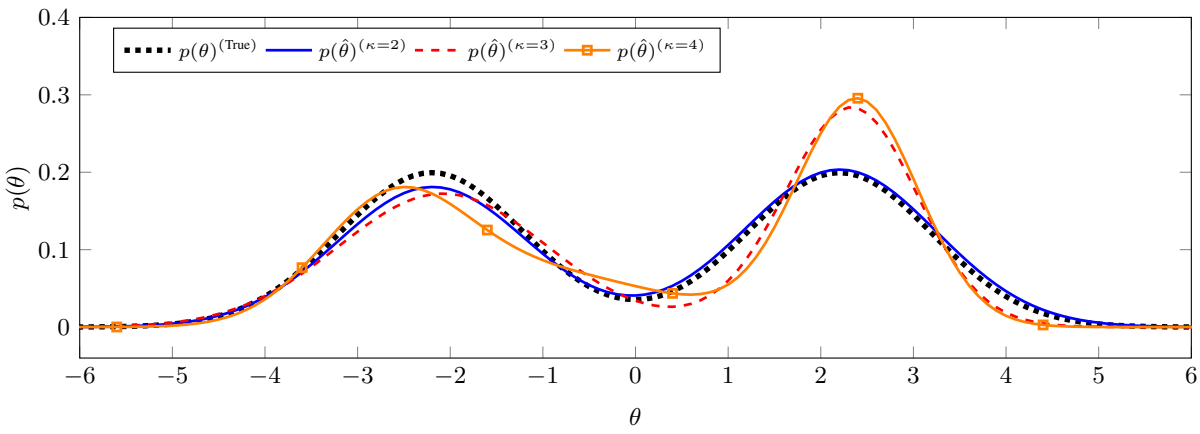



Figura 5.2: Estimación de la función de distribución a *priori* para $M = 200$ y diferentes valores de $\kappa$ en el Ejemplo 1.

$\lambda_2 = 0{,}5136$, media de la componente de $\mu_2 = 2{,}3084$ y covarianza de la GMM de $\Sigma_2 = 0{,}5400$. Para $\kappa = 4$, es posible descartar la componente $i = 4$ de acuerdo a la metodología *prunning*. Por otro lado, se aplica la técnica de *joining* evaluando la similitud entre los pares de componentes $i = \{1, 2\}$. Entonces, se encuentra la distancia de Mahalonobis [57] de acuerdo a (2.6) obteniendo $d_{12}^2 = 3{,}17 \times 10^{-2}$. Por lo tanto, de acuerdo a este criterio de selección es posible fusionar los pares de distribuciones consideradas estableciendo una GMM de dos componentes, $\kappa = 2$, definida por factor de ponderación $\lambda_1 = 0{,}4377$, media de la componente de $\mu_1 = -2{,}1796$ y covarianza de la GMM de $\Sigma_1 = 1{,}2543$. Finalmente se concluye que el mejor modelo que ajusta la distribución a *priori* es una GMM con $\kappa = 2$.

### 5.3.3. Resultados de las simulaciones de Monte Carlo

La Tabla 5.2 muestra los resultados de las estimaciones de los parámetros de la GMM usando una estructura $\kappa = 2$ para diferentes números de experimentos ($M = \{5, 200\}$). La Figura 5.3 muestra el promedio de la estimación de la función de distribución a *priori* para todas las simulaciones de MC, todo esto para diferentes números de experimentos. Se observa un mejor ajuste a la distribución *verdadera* a medida que se incrementa el

Tabla 5.2: Estimaciones de las simulaciones de MC para los parámetros de la GMM en el Ejemplo 1.

| Comp./ Exper. | $M = 5$ | | | $M = 200$ | | |
|---|---|---|---|---|---|---|
| | $\lambda_i$ | $\mu_i$ | $\Sigma_i$ | $\lambda_i$ | $\mu_i$ | $\Sigma_i$ |
| $i = 1$ | $0{,}4876 \pm 0{,}2320$ | $-1{,}7743 \pm 1{,}2218$ | $0{,}7291 \pm 0{,}9820$ | $0{,}5010 \pm 3{,}26 \times 10^{-2}$ | $-2{,}1863 \pm 0{,}1287$ | $0{,}9982 \pm 0{,}2017$ |
| $i = 2$ | $0{,}5124 \pm 0{,}2320$ | $1{,}9331 \pm 1{,}2827$ | $0{,}6824 \pm 0{,}9818$ | $0{,}4990 \pm 3{,}26 \times 10^{-2}$ | $2{,}1988 \pm 0{,}1005$ | $0{,}9703 \pm 0{,}1931$ |

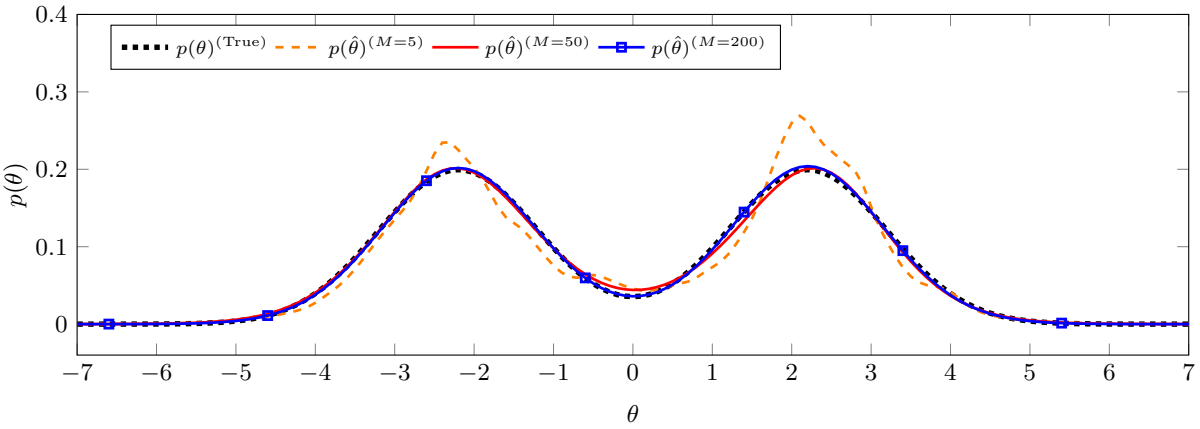



Figura 5.3: Estimación de la función de distribución a *priori* para el Ejemplo 1.

número de experimentos en el algoritmo propuesto.

La Figura 5.4 muestra la función de distribución a *posteriori* promedio de todas las simulaciones de MC, utilizando la estimación obtenida con el algoritmo propuesto para $M = 200$. La región gris sombreada representa el área determinada por el resultado de todas la estimaciones de MC usando el algoritmo propuesto. Se observa un mejor ajuste a la distribución *verdadera* usando una GMM en comparación a la obtenida cuando se considera una estructura difusa para la distribución a *priori* de $\theta$. De manera similar, la Figura 5.5 muestra la función de distribución acumulada (CDF) a *posteriori* obteniendo una mejor aproximación usando el algoritmo propuesto.

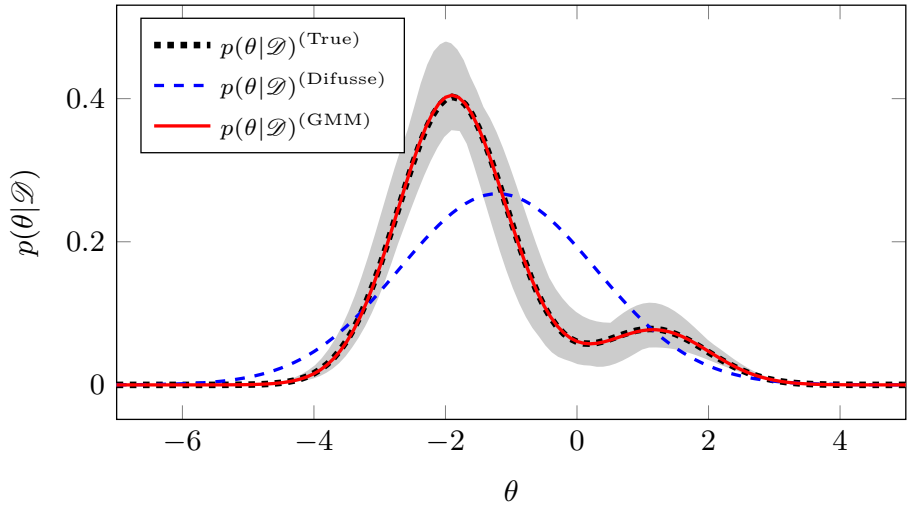



Figura 5.4: Función de distribución a *posteriori* obtenida para el Ejemplo 1 con $M = 200$.

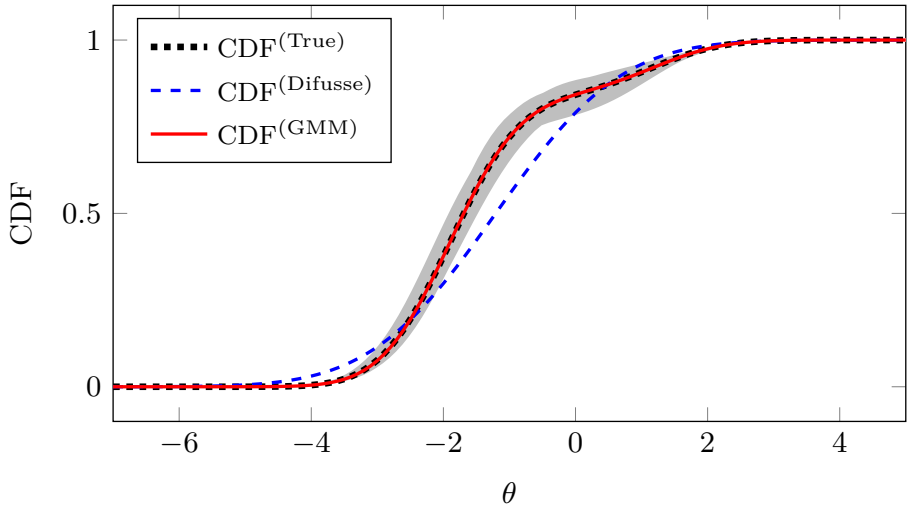



Figura 5.5: CDF a *posteriori* obtenida para el Ejemplo 1 con $M = 200$.

## 5.4. Ejemplo 2: Estimación de una distribución a *priori* uniforme.

Se considera que la función de distribución *verdadera* de $\theta$ viene dada por:

$$p(\theta)^{(\text{True})} = \frac{1}{b-a} I_{[a,b]}(\theta), \tag{5.4}$$

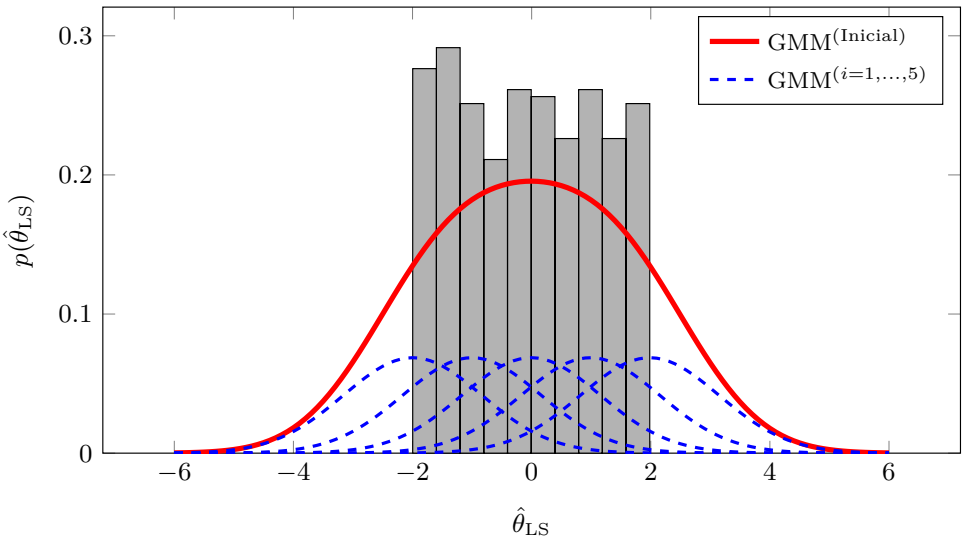



Figura 5.6: Histograma de la estimaciones de $\hat{\theta}_{\text{LS}}$ obtenidas para el Ejemplo 2 y la GMM inicial para $\kappa = 5$.

donde $a = -2$, $b = 2$, $I_{[a,b]}$ es una función indicatriz tal que:

$$I_{[a,b]} = \begin{cases} 1, & a \leq \theta \leq b \quad (5.5a) \\ 0, & \text{caso contrario} \quad (5.5b) \end{cases}$$

### 5.4.1. Inicialización del algoritmo propuesto

La inicialización del algoritmo propuesto corresponde al procedimiento descrito en el ejemplo anterior. Se asume que el parámetro $\theta$ es determinístico y se obtienen estimaciones independientes usando los datos de cada experimento ($M = 500$). Las estimaciones corresponden a las obtenidas con el estimador por mínimos cuadrados, $\hat{\theta}_{\text{LS}}$. En la Figura 5.6 las barras grises representan el histograma para las estimaciones obtenidas usando mínimos cuadrados. Las lineas segmentadas azules y la línea roja representan las componentes de la mezcla gaussiana y la GMM respectivamente, para $\kappa = 5$. En base a los resultados se establece que los valores iniciales para $\eta^{(0)}$, en función de la cantidad de componentes $\kappa$, se escogen de la siguiente forma:

- Cada componente $\Sigma_i$ se inicializa en el valor de la covarianza muestral

Tabla 5.3: Estimaciones de la GMM para el Ejemplo 2 usando $M = 150$ experimentos y diferentes valores de $\kappa$

| Componentes de GMM | $\kappa = 3$ | | | $\kappa = 5$ | | | $\kappa = 7$ | | |
|---|---|---|---|---|---|---|---|---|---|
| | $\lambda_i$ | $\mu_i$ | $\Sigma_i$ | $\lambda_i$ | $\mu_i$ | $\Sigma_i$ | $\lambda_i$ | $\mu_i$ | $\Sigma_i$ |
| $i = 1$ | 0,3399 | −1,3206 | $1{,}73 \times 10^{-1}$ | 0,1115 | −1,7989 | $2{,}08 \times 10^{-2}$ | 0,0600 | 1,8523 | $2{,}20 \times 10^{-3}$ |
| $i = 2$ | 0,4837 | 0,2063 | $3{,}38 \times 10^{-1}$ | 0,1991 | −1,1500 | $4{,}90 \times 10^{-2}$ | 0,0198 | 1,5575 | $4{,}07 \times 10^{-4}$ |
| $i = 3$ | 0,1764 | 1,4468 | $1{,}08 \times 10^{-1}$ | 0,4895 | 0,1102 | $3{,}30 \times 10^{-1}$ | 0,2996 | −0,9859 | $1{,}59 \times 10^{-1}$ |
| $i = 4$ | – | – | – | 0,0491 | 1,8516 | $1{,}6 \times 10^{-3}$ | 0,0631 | −1,9483 | $6{,}93 \times 10^{-4}$ |
| $i = 5$ | – | – | – | 0,1507 | 1,2656 | $6{,}71 \times 10^{-2}$ | 0,3673 | 0,2841 | $1{,}34 \times 10^{-1}$ |
| $i = 6$ | – | – | – | – | – | – | 0,1210 | 1,0708 | $1{,}96 \times 10^{-2}$ |
| $i = 7$ | – | – | – | – | – | – | 0,0691 | −1,6756 | $1{,}40 \times 10^{-2}$ |

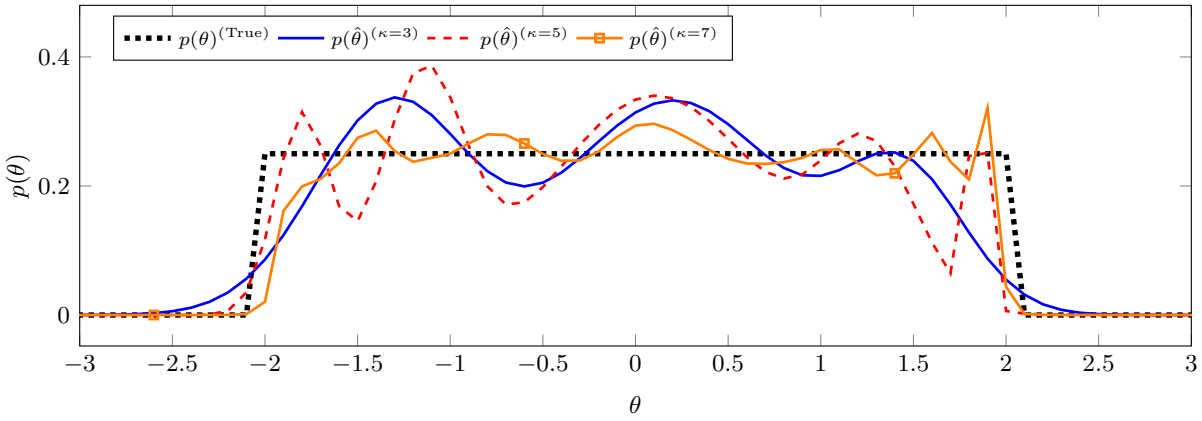



Figura 5.7: Estimación de la función de distribución a *priori* para $M = 150$ y diferentes valores de $\kappa$ en el Ejemplo 2.

de las estimaciones $\hat{\theta}_{\text{LS}}$.

- Las medias de las componentes de la GMM, $\mu_i$, se incializan en los valores equiespaciados entre el valor máximo y mínimo de las estimaciones $\hat{\theta}_{\text{LS}}$.

- Los factores de ponderación de la GMM, $\lambda_i$, se escogen de igual valor tal que $\sum_{i=1}^{\kappa} \lambda_i = 1$.

### 5.4.2. Selección del número de componentes de la GMM

Para la selección del orden la GMM, se obtienen las estimaciones usando diferentes valores de $\kappa$ para la GMM. Para ello se fija un número de experimentos en $M = 150$. La Figura 5.7 muestra como la GMM se ajusta

Tabla 5.4: Valores de BIC para la selección de la GMM en el Ejemplo 2

| **Componentes** | **BIC** |
|---|---|
| $\kappa = 3$ | $1{,}0451 \times 10^5$ |
| $\kappa = 5$ | $1{,}0440 \times 10^5$ |
| $\kappa = 7$ | $0{,}4645 \times 10^5$ |

a la distribución uniforme a medida que se consideran mayor cantidad de componentes de la mezcla de gaussianas.

Por otro lado, la Tabla 5.3 muestra los valores estimados para la GMM usando valores de $\kappa = \{3, 5, 7\}$. En este caso particular no es evidente aplicar técnicas de selección de estructuras de GMM como *prunning* o *joining*, ya que descartar alguna componente puede disminuir la exactitud de la aproximación de la distribución *verdadera*. Para la selección del número de componentes de la GMM se utiliza el criterio de información bayesiana (BIC), tal que:

$$\text{BIC} = \arg \min_{\mathscr{M}} \min_{\eta} \frac{1}{NM} \left[ -2\ell(\eta) + \rho_{\mathscr{M}} \log(NM) \right], \tag{5.6}$$

donde $\mathscr{M}$ es el conjunto de modelos a considerar y $\rho_{\mathscr{M}}$ es la dimensión del vector de parámetros $\eta$ para el modelo $\mathscr{M}$. La Tabla 5.4 muestra los valores de BIC para diferente cantidad de componentes de la GMM. De acuerdo a este criterio el mejor modelo que ajusta la distribución a *priori* corresponde a $\kappa = 7$. Sin embargo, para este caso particular una GMM que considere $\kappa > 7$ podría ajustar con mejor exactitud la distribución *verdadera* que corresponde a una distribución uniforme. Finalmente, en este ejemplo se selecciona una GMM de $\kappa = 7$.

### 5.4.3. Resultados de las simulaciones de Monte Carlo

La Figura 5.8 muestra la aproximación de la distribución a *priori* usando una estructura GMM con $\kappa = 7$ para diferentes números de experimentos. En ella se muestra el promedio de las estimaciones de las funciones de

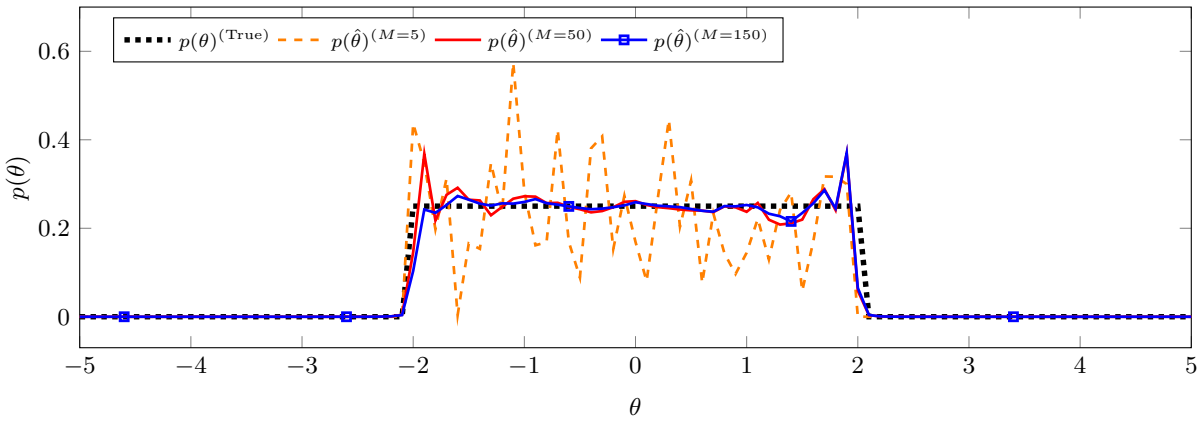



Figura 5.8: Estimación de la función de distribución a *priori* para el Ejemplo 2.

Tabla 5.5: Valores de MISE de la estimación de la distribución a *priori* para el Ejemplo 2

| **Experimentos** | **MISE** |
|---|---|
| $M = 5$ | $7{,}745 \times 10^{-1}$ |
| $M = 50$ | $2{,}740 \times 10^{-2}$ |
| $M = 150$ | $9{,}600 \times 10^{-3}$ |

distribución a *priori* obtenidas de las simulaciones de MC, logrando una mejor aproximación cuando se incrementa el número de experimentos. Para complementar este análisis, se calcula la integral del error cuadrático medio (MISE) para cuantificar el error de estimación de $p(\hat{\theta})$ considerando diferentes número de experimentos como se muestra en la Tabla 5.5. Al igual que el ejemplo anterior, se observa como la estimación de $p(\theta)$ se ajusta a la distribución *verdadera* cuando se consideran un mayor número de experimentos. El MISE se obtiene como:

$$\text{MISE} = \frac{1}{n_{\text{MC}}} \sum_{j=1}^{n_{\text{MC}}} \left( \frac{1}{N} \sum_{l=1}^{N} \left( p_j(\theta_l|\hat{\eta}) - p_j(\theta_l)^{(\text{True})} \right)^2 \right), \tag{5.7}$$

donde $p_j(\theta_l)^{(\text{True})}$ representa la pdf a *priori* verdadera evaluada en la $l$-ésima muestra de la $j$-ésima simulación de MC, $p_j(\theta_l|\hat{\eta})$ es la pdf a *priori* estimada usando una GMM para la $j$-ésima simulación de MC evaluada en la $i$-ésima muestra, $n_{\text{MC}}$ es el número de simulaciones de MC y $N$ es la longitud de los datos. La Figura 5.9 muestra claramente el impacto sobre

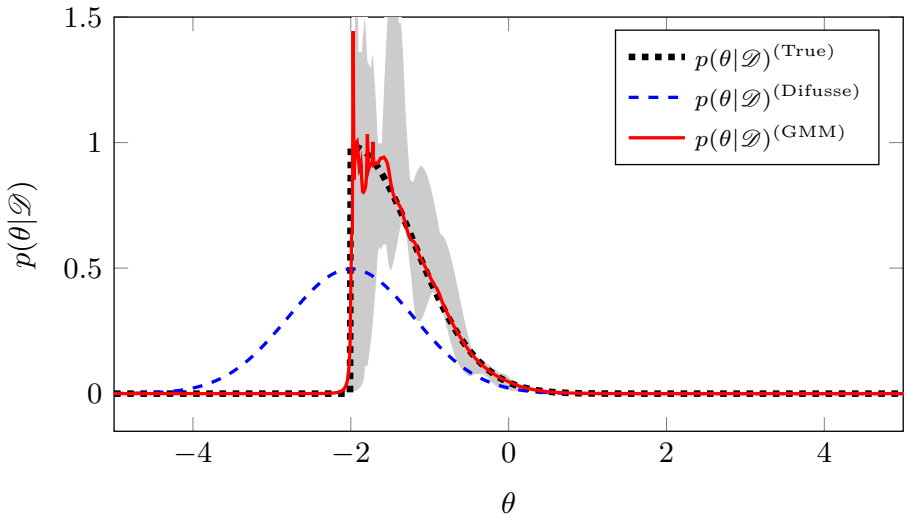



Figura 5.9: Función de distribución a *posteriori* obtenida para el Ejemplo 2 con $M = 150$.

la función de distribución a *posteriori* si se considera $p(\theta)$ estimada para $M = 150$. La región gris sombreada representa el área determinada por el resultado de todas la estimaciones de MC usando el algoritmo propuesto. De la misma manera, la Figura 5.10 muestra un mejor ajuste para la CDF a *posteriori* cuando se utiliza la aproximación de GMM para $p(\theta)$. En ambos casos se observa el beneficio de las estimaciones usando una aproximación por GMM en comparación con ele enfoque tradicional de considerar una distribución a *priori* difusa.

## 5.5. Conclusiones

En este capítulo se mostró el funcionamiento de algoritmo propuesto a través de ejemplos numéricos de simulación en donde se consideran dos casos: i) una distribución de probabilidad a *priori* no-gaussiana y ii) una distribución uniforme. Para cada ejemplo se estableció una metodología de inicialización del algoritmo propuesto. Adicionalmente, se analizó el uso de criterios de selección para la cantidad de componentes de la GMM. Finalmente, un análisis usando simulaciones de MC mostró un ajuste adecuado de la distribución a *priori* usando una suma finita de distribuciones gaussia-

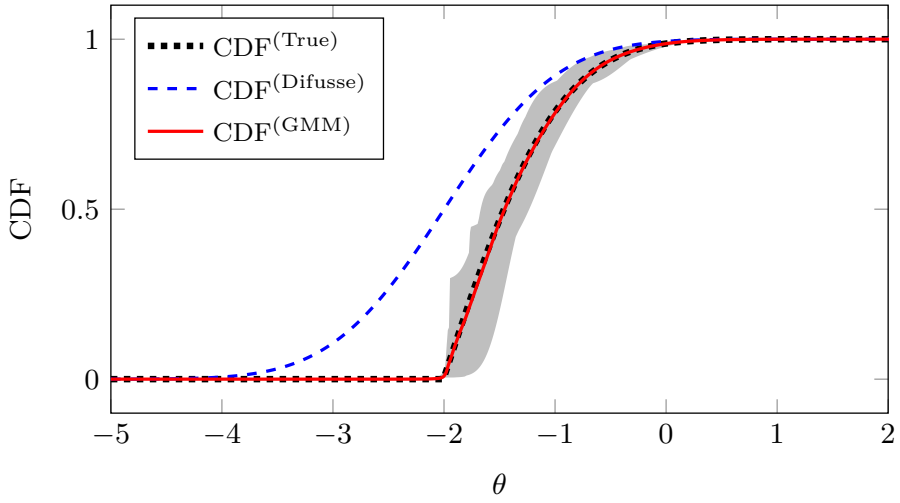



Figura 5.10: CDF a *posteriori* obtenida para el Ejemplo 2 con $M = 150$.

nas, obteniendo mejor resultado al incrementar el número de experimentos en el algoritmo.

# CAPÍTULO 6
# CONCLUSIONES

## 6.1. Conclusiones

En esta tesis se ha analizado el problema de estimación de la función de distribución a *priori* para la formulación clásica de inferencia estadística bayesiana. Para ello se consideró modelar la función de distribución a *priori* como la suma finita de funciones de distribución gaussianas. Adicionalmente, para distribuciones que no corresponden a la suma finita de distribuciones gaussianas, se consideró la posibilidad de su aproximación usando una GMM. Este enfoque se basa en la fundamentación teórica planteada en el teorema de aproximación de Wiener respecto a la aproximación de funciones de distribución desconocidas.

Para la formulación del algoritmo de estimación se utilizó el enfoque de estimación empírica bayesiana, la cual utiliza la técnica de máxima verosimilitud considerando la variable estocástica desconocida como una variable latente, y así obtener una estimación de los parámetros que definen la función de distribución a *priori* como una GMM. Adicionalmente, un análisis del comportamiento la función de verosimilitud mostró la presencia de máximos y mínimos locales, lo que condiciona la exactitud de las estimaciones de los parámetros que definen la GMM a una buena inicialización en los algoritmos de optimización.

El problema de estimación fue resuelto usando un algoritmo iterativo basado en la maximización de la esperanza (EM). Para ello se estableció una metodología que permite obtener una función auxiliar directamente de la ecuación integral obtenida de la formulación empírica bayesiana usando GMM. El algoritmo propuesto considera la información conjunta de un número finito de experimentos independientes para obtener una estimación

de los parámetros de la distribución a *priori* como una GMM. Se obtuvieron expresiones en forma cerrada para los estimadores que definen la GMM, en donde se utilizan aproximación numérica por métodos de Monte Carlo para resolver las ecuaciones integrales relacionadas a los estimadores de la GMM.

El funcionamiento del algoritmo propuesto fue analizado a través de ejemplos numéricos de simulación, en donde se consideraron dos casos: i) una distribución de probabilidad a *priori* no-gaussiana, en particular la suma de dos distribuciones gaussianas, y ii) una distribución a *priori* uniforme. Se analizó una metodología de inicialización del algoritmo propuesto, además de diferentes criterios de selección del número de componentes para la GMM. En ambos casos se observó un ajuste adecuado usando una suma finita de distribuciones gaussianas, obteniendo resultados con mejor exactitud al incrementar el número de experimentos en el algoritmo. Por otro lado, un análisis considerando simulaciones de Monte Carlo mostró como disminuye el error de estimación a medida que se incrementa el número de experimentos. Además, se mostró el impacto de nuestros resultados al obtener la pdf a *posteriori* y la CDF para el enfoque de estimación bayesiana. Todo esto puede ser utilizado para evaluar el impacto de la función de distribución a *priori* estimada en diferentes técnicas de inferencia bayesiana y sus aplicaciones en áreas como inteligencias artificial, máquinas de aprendizaje entre otras.

## 6.2. Trabajo futuro

El problema teórico resuelto en este trabajo contempló una solución iterativa que permita estimar la función de distribución *priori* como una GMM en el problema clásico de estimación bayesiana. Este planteamiento es un punto de partida para tratar el problema de interés, en el cual se puede seguir analizando lo siguiente:

- Analizar el desempeño del algoritmo considerando los diferentes criterios de selección del número de componentes de la GMM. Además, considerar la posibilidad que el número de componentes de la GMM sea un parámetro a estimar en función de un criterio de selección.
- Analizar la convergencia del algoritmo propuesto basado en EM. Para ello es posible considerar el caso en donde la estructura de la distribución desconocida corresponde a una GMM o bien es una aproximación de otra distribución aproximada como una GMM.
- Extender la formulación propuesta para sistemas dinámicos en donde se utilicen consideraciones de aproximación de distribuciones no-gaussianas usando una GMM, por ejemplo, caracterizar la incertidumbre de modelos lineales dinámicos usando mezclas de distribuciones gaussianas.

# BIBLIOGRAFÍA

[1] A. Spanos, *Probability Theory and Statistical Inference: Econometric Modeling with Observational Data*. Cambridge University Press, 1999.

[2] A. O'Hagan and J. Forster, *Kendall's Advanced Theory of Statistics, volume 2B: Bayesian Inference, second edition*. Arnold, 2004, vol. 2B.

[3] R. Carvajal, R. Orellana, D. Katselis, P. Escárate, and J. C. Agüero, "A data augmentation approach for a class of statistical inference problems," *PLOS ONE*, vol. 13, no. 12, pp. 1–24, 2018.

[4] H. Haimovich, G. Goodwin, and D. E. Quevedo, "Moving horizon Monte Carlo state estimation for linear systems with output quantization," in *42nd IEEE International Conference on Decision and Control*, vol. 5, Dec 2003, pp. 4859–4864.

[5] A. Tejada, O. González, and W. Gray, "On nonlinear discrete-time systems driven by Markov chains," *Journal of the Franklin Institute*, vol. 347, no. 5, pp. 795–805, 2010.

[6] T. Perez, J. Gómez, and S. Junco, "Induction motor parameter and state estimation using nonlinear observers," *Latin American Applied Research*, vol. 30, no. 2, pp. 107–113, 2000.

[7] R. Murray-Smith, D. Sbarbaro, C. Rasmussen, and A. Girard, "Adaptive, cautious, predictive control with Gaussian process priors," *IFAC Proceedings Volumes*, vol. 36, no. 16, pp. 1155–1160, 2003, 13th IFAC Symposium on System Identification.

[8] S. Haykin, K. Huber, and Zhe Chen, "Bayesian sequential state estimation for MIMO wireless communications," *Proceedings of the IEEE*, vol. 92, no. 3, pp. 439–454, March 2004.

[9] H. Chen and C. Qi, "Group Bayesian Sparse Channel Estimation for Massive MIMO Systems," in *IEEE Global Communications Conference*, Dec 2017, pp. 1–6.

[10] L. Araya-Hernández, J. F. Silva, A. Osses, and F. Tobar, "A Bayesian mixture-of-Gaussians model for astronomical observations in interferometry," in *CHILEAN Conference on Electrical, Electronics Engineering, Information and Communication Technologies*, Oct 2017, pp. 1–5.

[11] R. Mendez, R. Clavería, M. Orchard, and J. F. Silva, "Orbits for 18 visual binaries and two double-line spectroscopic binaries observed with HRCAM on the CTIO SOAR 4 m Telescope, Using a New Bayesian Orbit Code Based on Markov Chain Monte Carlo," *The Astronomical Journal*, vol. 154, no. 5, p. 187, oct 2017.

[12] B. Anderson and J. Moore, *Optimal Filtering*. Englewood Cliffs, NJ: Prentice-Hall, 1979.

[13] M. Volkhardt, S. Kalesse, S. Müller, and H. Gross, "Maximum a Posteriori Estimation of Dynamically Changing Distributions." Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 484–491.

[14] L. Ljung, *System Identification: Theory for the User*. Upper Saddle River, NJ, USA: Prentice Hall PTR, 1999.

[15] T. Söderström and P. Stoica, Eds., *System Identification*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1988.

[16] L. A. Dalton and E. R. Dougherty, "Bayesian Minimum Mean-Square Error Estimation for Classification Error—Part I: Definition and the Bayesian MMSE Error Estimator for Discrete Classification," *IEEE Transactions on Signal Processing*, vol. 59, no. 1, pp. 115–129, Jan 2011.

[17] S. V. Vaseghi, *Advanced Digital Signal Processing and Noise Reduction.* New York: John Wiley & Sons, 2008.

[18] F. Rahdari and M. Eftekhari, "Using bayesian classifiers for estimating quality of voip," in *The 16th CSI International Symposium on Artificial Intelligence and Signal Processing*, May 2012, pp. 348–353.

[19] X. Luo and A. Kareem, "Bayesian deep learning with hierarchical prior: Predictions from limited and noisy data," *Structural Safety*, vol. 84, p. 101918, May 2020. [Online]. Available: http://dx.doi.org/10.1016/j.strusafe.2019.101918

[20] R. Risuleo, G. Bottegal, and H. Hjalmarsson, "Modeling and identification of uncertain-input systems," *Automatica*, vol. 105, pp. 130 – 141, 2019.

[21] M. Klasson, S. I. Adalbjörnsson, J. Swärd, and S. V. Andersen, "Conjugate-prior-regularized multinomial pLSA for collaborative filtering," in *25th European Signal Processing Conference*, Aug 2017, pp. 2501–2505.

[22] B. Carlin and T. Louis, "Bayes and Empirical Bayes methods for data analysis," *Statistics and Computing*, vol. 7, no. 2, pp. 153–154, 1997.

[23] N. Everitt, G. Bottegal, and H. Hjalmarsson, "An empirical Bayes approach to identification of modules in dynamic networks," *Automatica*, vol. 91, pp. 144–151, 2018.

[24] J. Yansong, S. Baoku, and Y. Liu, "Application of Empirical Bayes Estimation in Error Model Identification of Two Orthometric Accelerometers," in *3rd International Symposium on Systems and Control in Aeronautics and Astronautics*, June 2010, pp. 218–221.

[25] I. Arasaratnam, S. Haykin, and R. J. Elliott, "Discrete-Time Nonlinear Filtering Algorithms Using Gauss-Hermite quadrature," *Proceedings of the IEEE*, vol. 95, no. 5, pp. 953–977, 2007.

[26] W. Li, Y. Jia, J. Du, and J. Zhang, "PHD filter for multi-target tracking with glint noise," *Signal Processing*, vol. 94, pp. 48 – 56, 2014.

[27] I. Bilik and J. Tabrikian, "Target tracking in glint noise environment using nonlinear non-Gaussian Kalman filter," in *IEEE Conference on Radar*, 2006, pp. 1–6.

[28] J. Janouek, P. Gajdo, M. Radecký, and V. Snáel, "Gaussian Mixture Model Cluster Forest," in *IEEE 14th International Conference on Machine Learning and Applications*, 2015, pp. 1019–1023.

[29] R. Orellana, R. Carvajal, and J. C. Agüero, "Maximum Likelihood Infinite Mixture Distribution Estimation Utilizing Finite Gaussian Mixtures," *IFAC-PapersOnLine*, vol. 51, no. 15, pp. 706–711, 2018, 18th IFAC Symposium on System Identification.

[30] G. Bittner, R. Orellana, R. Carvajal, and J. C. Agüero, "Maximum Likelihood identification for Linear Dynamic Systems with finite Gaussian mixture noise distribution," in *IEEE CHILEAN Conference on Electrical, Electronics Engineering, Information and Communication Technologies*, Oct 2019, pp. 1–6.

[31] H. W. Sorenson and D. L. Alspach, "Recursive Bayesian estimation using Gaussian sums," *Automatica*, vol. 7, no. 4, pp. 465–479, 1971.

[32] J. Dahlin, A. Wills, and B. Ninness, "Sparse Bayesian ARX models with flexible noise distributions," *IFAC-PapersOnLine*, vol. 51, no. 15, pp. 25 – 30, 2018, 18th IFAC Symposium on System Identification.

[33] R. Orellana, P. Escárate, M. Curé, A. Christen, R. Carvajal, and J. C. Agüero, "A method to deconvolve stellar rotational velocities - III.

The probability distribution function via maximum likelihood utilizing finite distribution mixtures," *A&A*, vol. 623, p. A138, 2019.

[34] S. Frühwirth-Schnatter, G. Celeux, and C. Robert, *Handbook of Mixture Analysis*. Chapman and Hall/CRC, 2018.

[35] B. Everitt and D. J. Hand, *Finite mixture distributions*. Chapman and Hall London, 1981.

[36] G. J. McLachlan, S. X. Lee, and S. I. Rathnayake, "Finite mixture models," *Annual Review of Statistics and Its Application*, vol. 6, no. 1, pp. 355–378, 2019.

[37] M. Zanetti, F. Bovolo, and L. Bruzzone, "Rayleigh-Rice Mixture Parameter Estimation via EM Algorithm for Change Detection in Multispectral Images," *IEEE Transactions on Image Processing*, vol. 24, no. 12, pp. 5004–5016, Dec 2015.

[38] A. Wills, J. Hendriks, C. Renton, and B. Ninness, "A Bayesian Filtering Algorithm for Gaussian Mixture Models," *ArXiv e-prints*, 2017.

[39] T. Söderström, *Discrete-Time Stochastic Systems: Estimation and Control*, 2nd ed. Berlin, Heidelberg: Springer-Verlag, 2002.

[40] K. Mengersen, C. Robert, and M. Titterington, *Mixtures: Estimation and Applications*. Wiley, 2011.

[41] N. Wiener, "Tauberian theorems," *Annals of Mathematics*, vol. 33, no. 1, pp. 1–100, 1932.

[42] J. Lo, "Finite-dimensional sensor orbits and optimal nonlinear filtering," *IEEE Transactions on Information Theory*, vol. 18, no. 5, pp. 583–588, 1972.

[43] N. I. Achieser, *Theory of Approximation*. Dover Publications, 1992.

[44] H. Akaike, "A new look at the statistical model identification," *IEEE Transactions on Automatic Control*, vol. 19, no. 6, pp. 716–723, December 1974.

[45] G. Schwarz, "Estimating the dimension of a model," *The Annals of Statistics*, vol. 6, no. 2, pp. 461–464, 1978.

[46] C. Keribin, "Consistent estimation of the order of mixture models," *Sankhyā: The Indian Journal of Statistics, Series A (1961-2002)*, vol. 62, no. 1, pp. 49–66, 2000.

[47] C. Fraley and A. E. Raftery, "Model-based clustering, discriminant analysis, and density estimation," *Journal of the American Statistical Association*, vol. 97, no. 458, pp. 611–631, 2002.

[48] L. Birgé and P. Massart, "Gaussian model selection," *Journal of the European Mathematical Society*, vol. 3, no. 3, pp. 203–268, Aug 2001.

[49] ——, "Minimal penalties for Gaussian model selection," *Probability Theory and Related Fields*, vol. 138, no. 1, pp. 33–73, May 2007.

[50] J.-P. Baudry, C. Maugis, and B. Michel, "Slope heuristics: overview and implementation," *Statistics and Computing*, vol. 22, no. 2, pp. 455–470, Mar 2012.

[51] D. J. Spiegelhalter, N. G. Best, B. P. Carlin, and A. Van Der Linde, "Bayesian measures of model complexity and fit," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 64, no. 4, pp. 583–639, 2002.

[52] G. Celeux, F. Forbes, C. P. Robert, and D. M. Titterington, "Deviance information criteria for missing data models," *Bayesian Anal.*, vol. 1, no. 4, pp. 651–673, 12 2006.

[53] M. A. T. Figueiredo and A. K. Jain, "Unsupervised learning of finite mixture models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 3, pp. 381–396, March 2002.

[54] J. Rissanen, *Optimal Estimation of Parameters*. Cambridge University Press, 2012.

[55] D. Titterington, A. Smith, and U. Makov, *Statistical Analysis of Finite Mixture Distributions*. Wiley, New York, 1985.

[56] J. Williams, "Gaussian mixture reduction for tracking multiple maneuvering targets in clutter," Ph.D. dissertation, 3 2003, air Force Inst. Technol., Wright-Patterson Air ForceBase, OH. [Online]. Available: http://ssg.mit.edu/jlwil/

[57] D. J. Salmond, "Mixture reduction algorithms for target tracking in clutter," in *Signal and Data Processing of Small Targets 1990*, vol. 1305, International Society for Optics and Photonics. SPIE, 1990.

[58] G. A. Young, "Mathematical statistics: An introduction to likelihood based inference," *International Statistical Review*, vol. 87, no. 1, pp. 178–179, 2019.

[59] R. E. Kass, "The geometry of asymptotic inference," *Statist. Sci.*, vol. 4, no. 3, pp. 188–219, 08 1989.

[60] A. P. Dempster, N. M. Laird, and D. Rubin, "Maximum Likelihood from Incomplete Data via the EM Algorithm," *Royal Statistical Society*, vol. 39, no. 1, pp. 1–38, 1977.

[61] L. Ljung and G. C. Goodwin and J. C. Agüero, "Stochastic Embedding revisited: A modern interpretation," in *53rd IEEE Conference on Decision and Control*, Dec 2014, pp. 3340–3345.

[62] L. Ljung and G. C. Goodwin and J.C. Agüero and T. Chen, "Model Error Modeling and Stochastic Embedding," *IFAC-PapersOnLine*,

vol. 48, no. 28, pp. 75 – 79, 2015, 17th IFAC Symposium on System Identification.

[63] G. McLachlan and D. Peel, Eds., *Finite Mixture Models*. NJ, USA: Wiley, 2000.

[64] L. Xu and M. I. Jordan, "On convergence properties of the EM algorithm for Gaussian mixtures," *Neural Computation*, vol. 8, no. 1, pp. 129–151, Jan 1996.

[65] A. Polanski, M. Marczyk, M. Pietrowska, P. Widlak, and J. Polanska, "Initializing the EM Algorithm for Univariate Gaussian, Multi-Component, Heteroscedastic Mixture Models by Dynamic Programming Partitions," *International Journal of Computational Methods*, vol. 15, no. 03, p. 1850012, 2018.

[66] J. Ma, L. Xu, and M. I. Jordan, "Asymptotic convergence rate of the EM algorithm for Gaussian mixtures," *Neural Computation*, vol. 12, no. 12, pp. 2881–2907, Dec 2000.

[67] C. Jin, Y. Zhang, S. Balakrishnan, M. J. Wainwright, and M. I. Jordan, "Local Maxima in the Likelihood of Gaussian Mixture Models: Structural Results and Algorithmic Consequences." Red Hook, NY, USA: Curran Associates Inc., 2016, p. 4123–4131.

[68] J. I. Yuz, J. Alfaro, J. C. Agüero, and G. C. Goodwin, "Identification of continuous-time state-space models from non-uniform fast-sampled data," *IET Control Theory Applications*, vol. 5, no. 7, pp. 842–855, May 2011.

[69] J. C. Agüero, G. C. Goodwin, and J. I. Yuz, "System identification using quantized data," in *46th IEEE Conference on Decision and Control*, Dec 2007, pp. 4263–4268.

[70] L. Wang, G. G. Yin, and J.-F. Zhang, "System identification using quantized data," *IFAC Proceedings Volumes*, vol. 39, no. 1, pp. 255 – 260, 2006, 14th IFAC Symposium on Identification and System Parameter Estimation.

[71] J. C. Agüero, B. Godoy, G. C. Goodwin, and T. Wigren, "Scenario-based EM Identification for FIR systems having quantized output data," *IFAC Proceedings Volumes*, vol. 42, no. 10, pp. 66 – 71, 2009, 15th IFAC Symposium on System Identification.

[72] J. C. Agüero, K. González, and R. Carvajal, "EM-based identification of ARX systems having quantized output data," *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 8367 – 8372, 2017, 20th IFAC World Congress.

[73] R. Carvajal, J. C. Agüero, B. Godoy, G. C. Goodwin, and J. I. Yuz, "EM-based identification of sparse FIR systems having quantized data," *IFAC Proceedings Volumes*, vol. 45, no. 16, pp. 553 – 558, 2012, 16th IFAC Symposium on System Identification.

[74] B. Godoy, G. C. Goodwin, J. C. Agüero, D. Marelli, and T. Wigren, "On identification of FIR systems having quantized output data," *Automatica*, vol. 47, no. 9, pp. 1905–1915, 2011.

[75] B. Godoy, J. C. Agüero, R. Carvajal, G. C. Goodwin, and J. Yuz, "Identification of sparse FIR systems using a general quantisation scheme," *Int. J. Control*, vol. 87, no. 4, pp. 874–886, 2014.

[76] J. C. Agüero, T. Wei, J. Yuz, R. Delgado, and G. C. Goodwin, "Dual time–frequency domain system identification," *Automatica*, vol. 48, p. 3031–3041, 12 2012.

[77] R. Carvajal, J. C. Agüero, B. I. Godoy, and G. C. Goodwin, "EM-based channel estimation in OFDM systems with phase noise," in *IEEE Global Telecommunications Conference*, Dec 2011, pp. 1–5.

[78] R. A. Delgado, G. C. Goodwin, R. Carvajal, and J. C. Agüero, "A novel approach to model error modelling using the expectation-maximization algorithm," in *51st IEEE Conference on Decision and Control (CDC)*, Dec 2012, pp. 7327–7332.

[79] S. Gibson, A. Wills, and B. Ninness, "Maximum-likelihood parameter estimation of bilinear systems," *IEEE Transactions on Automatic Control*, vol. 50, no. 10, pp. 1581–1596, Oct 2005.

[80] T. Schön, A. Wills, and B. Ninness, "System identification of nonlinear state-space models," *Automatica*, vol. 47, no. 1, pp. 39 – 49, 2011.

[81] C. Robert. and G. Casella, *Monte Carlo Statistical Methods*. Berlin, Heidelberg: Springer-Verlag, 2005.

[82] L. Shampine, "Vectorized adaptive quadrature in MATLAB," *Journal of Computational and Applied Mathematics*, vol. 211, no. 2, pp. 131–140, 2008.

[83] R. Cools and D. Nuyens, Eds., *Monte Carlo and Quasi-Monte Carlo Methods*. Springer International Publishing, 2016. [Online]. Available: https://doi.org/10.1007%2F978-3-319-33507-0

[84] S. Theodoridis, "Chapter 14 - Monte Carlo Methods," in *Machine Learning*. Oxford: Academic Press, 2015, pp. 707 – 744.

[85] P. Del Moral, "Nonlinear filtering: Interacting particle resolution," *Comptes Rendus de l'Académie des Sciences - Series I - Mathematics*, vol. 325, no. 6, pp. 653 – 658, 1997.

[86] H. Flanders and J. J. Price, "Calculus with Analytic Geometry." Academic Press, 1978, pp. 204 – 261.

[87] N. Metropolis and S. Ulam, "The Monte Carlo Method," *Journal of the American Statistical Association*, vol. 44, no. 247, pp. 335–341, 1949.

[88] H. Kahn and A. W. Marshall, "Methods of Reducing Sample Size in Monte Carlo Computations," *Journal of the Operations Research Society of America*, vol. 1, no. 5, pp. 263–278, 1953.

[89] R. Neal, "Slice Sampling," *The Annals of Statistics*, vol. 31, no. 3, pp. 705–767, 2003.

# APÉNDICE A

## CALCULO DE LOS PARÁMETROS DEL MODELO GMM

Derivando la función auxiliar $\bar{\mathcal{Q}}(\eta, \hat{\eta}^{(k)})$ en (4.10) con respecto a $\mu_i$ e igualando a cero se obtiene:

$$\frac{\partial \bar{\mathcal{Q}}(\eta, \hat{\eta}^{(k)})}{\partial \mu_i} = \sum_{r=1}^{M} \int_{-\infty}^{\infty} \theta^{[r]} \frac{K(\theta^{[r]}, \hat{\eta}_i^{(k)})}{\mathcal{V}^{[r]}(\hat{\eta}^{(k)})} d\theta^{[r]} - \sum_{r=1}^{M} \int_{-\infty}^{\infty} \hat{\mu}_i^{(k+1)} \frac{K(\theta^{[r]}, \hat{\eta}_i^{(k)})}{\mathcal{V}^{[r]}(\hat{\eta}^{(k)})} d\theta^{[r]} = 0 \tag{A.1}$$

Usando las definiciones en (4.15) y (4.16) se tiene:

$$\sum_{r=1}^{M} \int_{-\infty}^{\infty} \theta^{[r]} \frac{K(\theta^{[r]}, \hat{\eta}_j^{(k)})}{\mathcal{V}^{[r]}(\hat{\eta}^{(m)})} d\theta^{[r]} = \hat{\mu}_i^{(k+1)} \mathcal{P}(\theta, \hat{\eta}_i^{(k)})$$

$$\hat{\mu}_i^{(k+1)} = \frac{\mathcal{M}(\theta, \hat{\eta}_i^{(k)})}{\mathcal{P}(\theta, \hat{\eta}_i^{(k)})} \tag{A.2}$$

De manera similar, derivando $\bar{\mathcal{Q}}(\eta, \hat{\eta}^{(k)})$ en (4.10) con respecto a $\alpha = \Sigma_i^{-1}$ e igualando a cero:

$$\frac{\partial \bar{\mathcal{Q}}(\eta, \hat{\eta}^{(k)})}{\partial \alpha} = \hat{\Sigma}_i^{(k+1)} \sum_{r=1}^{M} \int_{-\infty}^{\infty} \frac{K(\theta^{[r]}, \hat{\eta}_i^{(k)})}{\mathcal{V}^{[r]}(\hat{\eta}^{(k)})} d\theta^{[r]} - \tag{A.3}$$

$$\sum_{r=1}^{M} \int_{-\infty}^{\infty} \left(\theta^{[r]} - \hat{\mu}_i^{(k)}\right) \left(\theta^{[r]} - \hat{\mu}_i^{(k)}\right)^T \frac{K(\theta^{[r]}, \hat{\eta}_i^{(k)})}{\mathcal{V}^{[r]}(\hat{\eta}^{(k)})} d\theta^{[r]} = 0$$

Usando (4.15) y (4.17) se obtiene:

$$\hat{\Sigma}_i^{(k+1)}\mathcal{P}(\theta,\hat{\eta}_i^{(k)}) = \sum_{r=1}^{M}\int_{-\infty}^{\infty}\left(\theta^{[r]}-\hat{\mu}_i^{(k)}\right)\left(\theta^{[r]}-\hat{\mu}_i^{(k)}\right)^T\frac{K(\theta^{[r]},\hat{\eta}_i^{(k)})}{\mathcal{V}^{[r]}(\hat{\eta}^{(k)})}d\theta^{[r]}$$

$$\hat{\Sigma}_i^{(k+1)} = \frac{\mathcal{S}(\theta,\hat{\eta}_i^{(k)})}{\mathcal{P}(\theta,\hat{\eta}_i^{(k)})} \tag{A.4}$$

Finalmente, para el parámetro $\lambda_i$ se define $\mathscr{R}(\lambda_i)$:

$$\mathscr{R}(\lambda_i) = \sum_{i=1}^{\kappa}\log[\lambda_i]\left\{\mathcal{P}(\theta,\hat{\eta}_i^{(k)})\right\}, \tag{A.5}$$

sujeto a la restricción $\sum_{j=1}^{\kappa}\lambda_i = 1$. Es importante resaltar que inicialmente no se considera la restricción $0 \leq \lambda_i \leq 1$.

Usando un multiplicador de Lagrange para considerar la restricción de $\lambda_i$ se define:

$$\mathscr{L}(\lambda_i,\gamma) = \sum_{i=1}^{\kappa}\log[\lambda_i]\left\{\mathcal{P}(\theta,\hat{\eta}_i^{(k)})\right\} - \gamma\left(\sum_{i=1}^{\kappa}\lambda_i - 1\right). \tag{A.6}$$

Derivando $\mathscr{L}(\lambda_i,\gamma)$ con respecto a $\lambda_i$ y $\gamma$, entonces igualando a cero se tiene:

$$\frac{\partial\mathscr{L}(\lambda_i,\gamma)}{\partial\lambda_i} = \frac{\mathcal{P}(\theta,\hat{\eta}_i^{(k)})}{\hat{\lambda}_i^{(k+1)}} - \gamma = 0 \tag{A.7}$$

$$\frac{\partial\mathscr{L}(\lambda_i,\gamma)}{\partial\gamma} = \sum_{i=1}^{\kappa}\lambda_i - 1 = 0 \tag{A.8}$$

Entonces:

$$\hat{\lambda}_i^{(k+1)} = \frac{\mathcal{P}(\theta,\hat{\eta}_i^{(k)})}{\gamma} \tag{A.9}$$

Usando (A.8) y sumatoria en $i = 1, \ldots, \kappa$ para (A.9) se tiene:

$$\sum_{i=1}^{\kappa} \hat{\lambda}_i^{(k+1)} = \sum_{i=1}^{\kappa} \frac{\mathcal{P}(\theta, \hat{\eta}_i^{(k)})}{\gamma} = 1. \tag{A.10}$$

Finalmente se obtiene:

$$\hat{\lambda}_i^{(k+1)} = \frac{\mathcal{P}(\theta, \hat{\eta}_i^{(k)})}{\sum_{l=1}^{\kappa} \mathcal{P}(\theta, \hat{\eta}_l^{(k)})}. \tag{A.11}$$

Hay que resaltar que la restricción de $0 \leq \lambda_i \leq 1$ se mantiene en (A.11), a pesar de no ser considerada explícitamente en (A.6).

□

# APÉNDICE B
# CÓDIGO DE SIMULACIÓN EN MATLAB®

**Algoritmo:** Código en Matlab® para estimación de la GMM

```
%============================
% Prueba Prior Estimation - GMM
%============================
clc;
clear all;
close all;

N=1000 % Longitud de datos
term1='data_FIRsystem_N';
term2='_M200_MCsimul';
filename=[term1,int2str(N),term2];
load(filename);
M=200 % Numero de experimentos
Nmc=10000 % Longitud para integracion MC
realz=100 % Realizaciones
%====================
%GMM valores iniciales
%====================
K=2;
mu_0=mean(yt(:));
sigma_0=var(yt(:));
lambda_0=1/K;
gamm0=[ones(K,1).*mu_0 ones(K,1).*sigma_0 ones(K,1).*lambda_0];

nb=1;
ut=u_t(1:end-nb);

for k=1:realz
    clearvars yt;
    yt=y_t(nb+1:end,k*M-M+1:k*M);
    gamma=gamma0;
    for j=1:100
        k,j
        [Vt,dataMC]=Vtfunc_no_GPU(yt,ut,gamma,N,M,K,Nmc);
        [P,Mm,S]=Estep(yt,ut,gamma,Vt,N,M,K,dataMC);
```

```
            gamma_m_1=Mstep(P,Mm,S,K)
              tol1=[gamma_m_1(:)]-[gamma(:)];
              tol=norm(tol1,2)/norm([gamma(:)],2)
              if tol < 5e-5
                  gamma=gamma_m_1;
                  break;
              end
            gamma=gamma_m_1;
        end
        estimate_gamma(:,:,k)=gamma;
end

 term1='est_Prior_GMM_';
 term2='N';
 term3='_M';
 filename=[term1,term2,int2str(N),term3,int2str(M)];
 estimate_gamma=gamma;
 save(filename,'estimate_gamma');

 function [Vt,dataMC]=Vtfunc_no_GPU(yt,ut,gamma,N,M,K,Nmc)
mu=gamma(:,1);
sigma=gamma(:,2);
lambda=gamma(:,3);

dataMC=zeros(length(lambda),Nmc);
 for j=1:length(lambda)
     dataMC(j,:)=sqrt(sigma(j))*randn(1,Nmc)+mu(j);
 end

Vt=zeros(N,M);
for r=1:M
     for t=1:N
          acum=0;
          for j=1:K
               acum=acum+MC_integration( yt(t,r), ut(t), lambda(j), mu(j),...
                    [],1, [], 1, dataMC(j,:));
          end
       Vt(t,r)=acum;
       if Vt(t,r)<eps
            Vt(t,r)=eps;
       end
     end
end
end
```

**Algoritmo:** Código en Matlab® para el E-step

```
function [P,Mm,S]=Estep(yt,ut,gamma,Vt,N,M,K,dataMC)
mu=gamma(:,1);          % medias de la GMM
sigma=gamma(:,2);       % varianza de la GMM
lambda=gamma(:,3);      % pesos de la GMM

for j=1:K
    for r=1:M
        for t=1:N
            intauxP(t,r)=MC_integration( yt(t,r), ut(t), lambda(j), mu(j),...
                [],1, Vt(t,r), 2, dataMC(j,:));
            intauxM(t,r)=MC_integration( yt(t,r), ut(t), lambda(j), mu(j),...
                [],1, Vt(t,r), 3, dataMC(j,:));
            intauxS(t,r)=MC_integration( yt(t,r), ut(t), lambda(j), mu(j),...
                [],1, Vt(t,r), 4, dataMC(j,:));
        end
    end
    P(j,1)=sum(sum(intauxP));
    Mm(j,1)=sum(sum(intauxM));
    S(j,1)=sum(sum(intauxS));
end
end
```

**Algoritmo:** Código en Matlab® para el M-step

```
function gamma_m_1=Mstep(P,Mm,S,K)
lambda_m1=zeros(K,1);     % pesos de la GMM
mu_m1=zeros(K,1);         % medias de la GMM
sigma_m1=zeros(K,1);      % varianzas de la GMM

for j=1:K
    lambda_m1(j,1)=P(j)/(sum(P));
    mu_m1(j,1)=Mm(j,1)/P(j,1);
    sigma_m1(j,1)=(S(j,1))/(P(j,1));%-(mu_m1(j,1))^2;
end
gamma_m_1=[mu_m1 sigma_m1 lambda_m1];
end
```

**Algoritmo:** Código en Matlab® para integración numérica

```
function result=MC_integration(y,u,lambda,mu,sigma,sigma_w,Vt,opt,x)
if opt==1
```

```
py_eta=@(x)(1/sqrt(2*pi*sigma_w)).*(exp(-0.5.*((y-x.*u).^2)./sigma_w));
result=lambda*mean(py_eta(x));
end

if opt==2
    py_eta=@(x)(1/sqrt(2*pi*sigma_w)).*...
        (exp(-0.5.*((y-x.*u).^2)./sigma_w))./(Vt);
    result=lambda*mean(py_eta(x));
end

if opt==3
py_eta=@(x) (x).*(1/sqrt(2*pi*sigma_w)).*...
    (exp(-0.5.*((y-x.*u).^2)./sigma_w))./(Vt);
result=lambda*mean(py_eta(x));
end

if opt==4
   py_eta=@(x) ((x-mu).^2).*(1/sqrt(2*pi*sigma_w)).*...
       (exp(-0.5.*((y-x.*u).^2)./sigma_w))./(Vt);
   result=lambda*mean(py_eta(x));
end

end
```