

PONTIFICIA UNIVERSIDAD CATÓLICA DEL PERÚ

Escuela de Posgrado



Modelo de regresión lineal usando una mixtura de
distribuciones senh-normal

Tesis para optar por el grado académico de Maestro en
Estadística

que presenta:

Roussel Simpson Palomino Ore

Asesor:

Dr. Luis Enrique Benites Sánchez

Lima, Julio 2023

Informe de Similitud

Yo, Luis Enrique Benites Sánchez, docente de la Escuela de Posgrado de la Pontificia Universidad Católica del Perú, asesor de la tesis de investigación titulado: Modelo de regresión lineal usando una mixtura de distribuciones senh-normal, del autor Roussel Simpson Palomino Ore, dejo constancia de lo siguiente:

- El mencionado documento tiene un índice de puntuación de similitud de 19%. Así lo consigna el reporte de similitud emitido por el software Turnitin el 19/07/2023.
- He revisado con detalle dicho reporte y la Tesis o Trabajo de Suficiencia Profesional, y no se advierte indicios de plagio.
- Las citas a otros autores y sus respectivas referencias cumplen con las pautas académicas.

Lugar y fecha: Lima, Julio 2023

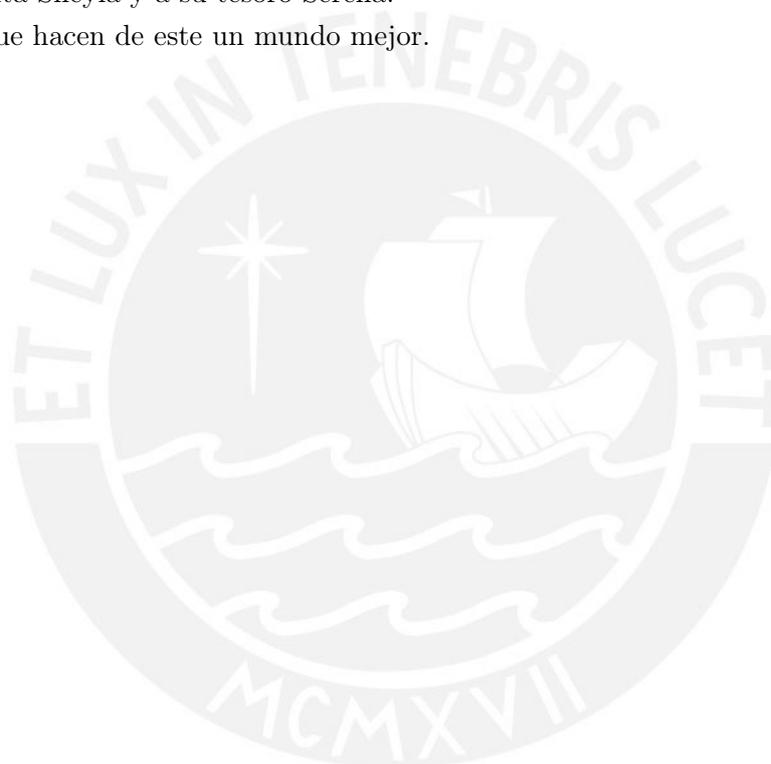
Apellidos y nombres del asesor: Benites Sánchez, Luis Enrique	
DNI:42987865	Firma: 
ORCID:0000000159987098	

Dedicatoria

A mis padres Alberto y Betsabé, por su inmenso amor.

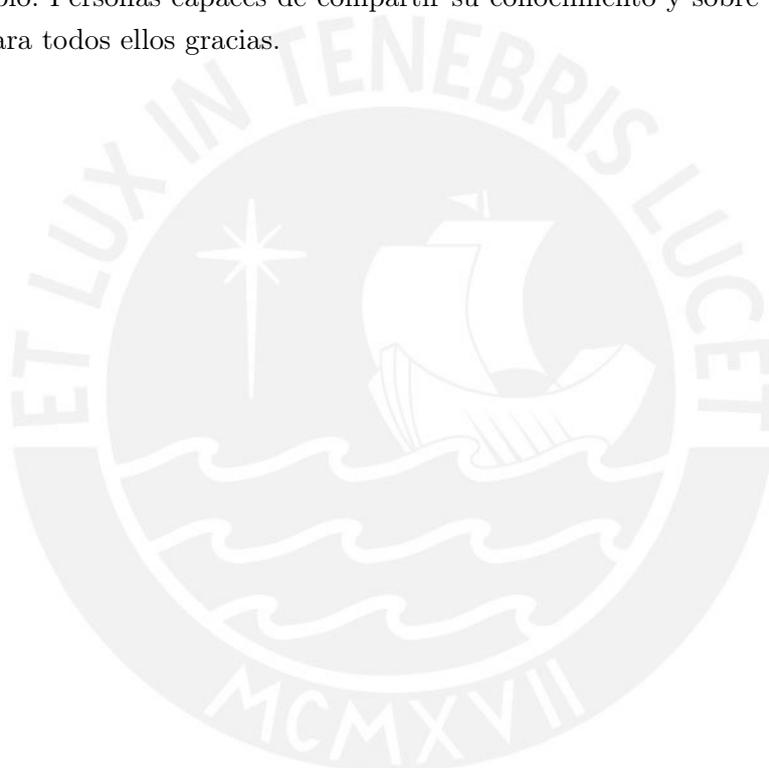
A mi hermanita Sheyla y a su tesoro Serena.

A todos los que hacen de este un mundo mejor.



Agradecimientos

Son muchas las personas que a lo largo de mi vida y carrera profesional me han brindado su apoyo y ejemplo. Personas capaces de compartir su conocimiento y sobre todo sus ganas de superación, para todos ellos gracias.



Resumen

La distribución Senohiperbólico-Normal, denominada también como una variación de la distribución Birnbaum-Saunders, surgió inicialmente para estimar el deterioro en la calidad de los materiales sujetos a estrés. Asimismo, los modelos de mixtura han suscitado considerable interés en el campo de estadística debido a que permiten lidiar con situaciones en las que el comportamiento de los errores de un modelo con ajuste lineal se aleja significativamente de la normalidad. Esta tesis aborda los dos temas mencionados mediante la presentación de un modelo de ajuste lineal usando una mixtura de distribuciones Senohiperbólico Normal o Log-Birnbaum-Saunders. Esta propuesta es una familia versátil de distribuciones de probabilidad que posibilita representar datos que presentan multimodalidad además de provenir de poblaciones heterogéneas. Para conseguir los estimadores de máxima verosimilitud se emplea el algoritmo EM con maximización condicional. Asimismo, se llevan a cabo estudios de simulación y análisis de conjuntos de datos reales para demostrar la utilidad del método propuesto. Por último, se implementa la propuesta del algoritmo y los métodos en el software R.

Palabras-clave: Modelos de regresión lineal, Mixtura finita, Log-Birnbaum-Saunders, ECM algoritmo.

Abstract

The Senohyperbolic-Normal distribution, also known as a variation of the Birnbaum-Saunders distribution, was initially developed to estimate the deterioration in the quality of materials subjected to stress. Likewise, mixture models have attracted considerable interest in the field of statistics because they allow dealing with situations where the error behavior of a linearly fitted model deviates significantly from normality. This thesis addresses the two aforementioned issues by presenting a linear fitting model using a mixture of Senohyperbolic Normal or Log-Birnbaum-Saunders distributions. This proposal is a versatile family of probability distributions that makes it possible to represent data that present multimodality as well as coming from heterogeneous populations. The EM algorithm with conditional maximization is used to obtain the maximum likelihood estimators. Simulation studies and analysis of real data sets are also carried out to demonstrate the usefulness of the proposed method. Finally, the proposed algorithm and methods are implemented in R software.

Keywords: Linear regression models, Finite Mixture, Log-Birnbaum-Saunders, ECM algorithm.

Índice general

Índice de figuras	IX
Índice de cuadros	X
1. Introducción	1
1.1. Consideraciones Preliminares	1
1.2. Objetivos del documento de trabajo	3
1.3. Estructura del documento	3
2. Distribución Senh-Normal	4
2.1. Distribución acumulada de la distribución Senh-Normal	4
2.2. Propiedades de la Distribución Senh-Normal	5
3. Modelo lineal Senh-Normal con mixtura	6
3.1. Algoritmo de maximización de esperanza condicional	7
3.2. Cálculo de los parámetros a través del algoritmo ECM	7
3.2.1. Criterio de detención	9
3.2.2. Valores iniciales	9
3.2.3. Selección del modelo	10
3.2.4. Desviación estándar estimado	10
4. Simulación	12
4.1. Recuperación de parámetros (estudio de simulación 1)	12
4.2. Propiedades asintóticas de las estimaciones de ECM (estudio de simulación 2)	14

5. Aplicaciones	17
5.1. Instituto Australiano de Deportes(AIS)	17
5.2. Clima de Szeged, Hungría, entre 2006 y 2016	19
6. Conclusiones y sugerencias	22
6.1. Conclusiones	22
A. Anexo de Demostraciones	23
A.1. Demostraciones del Capítulo 3	23
Bibliografía	29



Índice de figuras

4.1. Funciones de densidad. (a) Escenario 1 y (b) Escenario 2 del estudio de simulación.	13
4.2. Estudio de simulación 2. Sesgo promedio de los estimadores β , μ , p_1 y α . . .	15
4.3. Estudio de simulación 2. Promedio de RMSE de los estimadores β , μ , p_1 y α .	16
5.1. Histograma de los residuales ordinarios superpuestos con las densidades de los residuales considerando los modelos MR-MF-SHN y MR-MF-N para $g = 2$ componentes.	18
5.2. Histograma de los residuales ordinarios superpuestos con las densidades de los residuales considerando los modelos MR-MF-SHN y MR-MF-N para $g = 5$ componentes.	21

Índice de cuadros

4.1. Simulación 1 - Escenario 1: Media y MC Sd son las respectivas medias y desviaciones estándar estimados del modelo MR-MF-SHN basado en 1000 muestras. IM SE es el resultado promedio del residuo estándar aproximado conseguido con el procedimiento propuesto. Los valores reales están entre paréntesis. . . .	13
4.2. Simulación 1 - Escenario 2: Media y MC Sd son las respectivas medias y desviaciones estándar estimados del modelo MR-MF-SHN basado en 1000 muestras. IM SE es el resultado promedio del residuo estándar aproximado conseguido con el procedimiento propuesto. Los valores reales están entre paréntesis. . . .	14
5.1. Cuadro comparativo del máximo log-verosimilitud, el criterio de información de Akaike (AIC) y el criterio de información bayesiano (BIC) para los modelos MR-MF-SHN y MR-MF-N usando el conjunto de datos AIS.	18
5.2. Estimaciones de los parámetros del modelo MR-MF-SHN con $g = 2$. SE denota los errores estándar correspondientes usando el conjunto de datos AIS.	18
5.4. Valores estimados de los parámetros del modelo MR-MF-SHN con parámetro $g=5$. Se denota los errores estándar correspondientes empleando la muestra de datos históricos del clima de Szeged, Hungría.	20
5.3. Comparación del máximo log-verosimilitud, el criterio de información de Akaike (AIC) y el criterio de información bayesiano (BIC) para los modelos MR-MF-SHN y MR-MF-N usando el conjunto de datos históricos del clima de Szeged, Hungría, entre 2006 y 2016.	20

Capítulo 1

Introducción

1.1. Consideraciones Preliminares

Una de las suposiciones de un modelo de ajuste lineal es que el error tiene una distribución normal. No obstante, se tiene conocimiento de que hay fenómenos que no cumplen con esta suposición Johnson (1949) estableció como una alternativa relevante, la generación de distribuciones no normales mediante la transformación de variables aleatorias normales. Entre estas se incluyen las distribuciones Birnbaum-Saunders (BS) (Birnbaum y Saunders, 1969), log-normal (Johnson y Kotz, 1970) y Senohiperbólico-Normal (SHN) (Rieck y Nedelman, 1991), entre otras.

La distribución BS se originó en el campo de la ingeniería de materiales, debido a su adecuación para describir procesos de degradación acumulativa. Esto la convierte en un modelo versátil que puede aplicarse a otras áreas (Birnbaum y Saunders, 1969).

En el trabajo de Leiva et al. (2008) y Leiva et al. (2010) se presentan diversos ejemplos de aplicaciones en distintos campos, como ciencias de la salud, ambientales, forestales, demográficas, actuariales, financieras, entre otras áreas.

Birnbaum y Saunders (1969) demostró que bajo ciertas condiciones, la función de distribución acumulada (fda) de una variable aleatoria T , que representa el tiempo total hasta que ocurre una falla, puede expresarse mediante la siguiente ecuación:

$$F_T(t; \alpha, \beta) = P(T \leq t) = \Phi \left[\frac{1}{\alpha} \left(\sqrt{\frac{t}{\beta}} - \sqrt{\frac{\beta}{t}} \right) \right], \quad t > 0.$$

Donde T sigue una distribución BS con parámetro de localización $\alpha > 0$ y escala $\beta > 0$ representada por $T \sim \text{BS}(\alpha, \beta)$.

Un aspecto crucial en la construcción de esta distribución es que la variable aleatoria exhibe

el siguiente comportamiento:

$$Z = \frac{1}{\alpha} \left(\sqrt{\frac{T}{\beta}} - \sqrt{\frac{\beta}{T}} \right) \sim N(0, 1).$$

Por lo tanto, la distribución BS guarda una estrecha relación con la distribución normal mediante la siguiente relación:

$$T = \beta \left[\frac{\alpha Z}{2} + \sqrt{\left(\frac{\alpha Z}{2} \right)^2 + 1} \right]^2,$$

donde $Z \sim N(0, 1)$. donde Z sigue una distribución normal estándar $Z \sim N(0, 1)$. Al tener en cuenta la función de distribución acumulada (fda) de la variable aleatoria T , dada en la ecuación (1.1), su correspondiente función de densidad de probabilidad (fsp) se define como sigue:

$$f_T(t) = \phi(a_t(\alpha, \beta)) A_t(\alpha, \beta),$$

donde $\phi(\cdot)$ representa la función de densidad de probabilidad (fdp) de la distribución normal estándar, mientras que $a_t(\alpha, \beta)$ y $A_t(\alpha, \beta)$ se definen como sigue:

$$a_t(\alpha, \beta) = \frac{1}{\alpha} \left(\sqrt{\frac{t}{\beta}} - \sqrt{\frac{\beta}{t}} \right) \quad \text{y} \quad A_t(\alpha, \beta) = \frac{d}{dt} a_t(\alpha, \beta) = \frac{t^{-3/2}(t + \beta)}{2\alpha\beta^{1/2}}.$$

Para esta investigación nos basaremos en el trabajo de Bartolucci y Scaccia (2005) los cuales propusieron un modelo de regresión en el cual los errores siguen una mixtura de distribuciones normales, cuya definición es la siguiente:

$$Y_i = \beta_0 + \mathbf{x}_i \boldsymbol{\beta} + \epsilon_i, \quad \epsilon_i \sim \sum_{j=1}^g p_j N(\mu_j, \sigma_j), \quad i = 1, \dots, n,$$

donde Y_i son las variables respuestas, $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})^\top$ es un conjunto de variables explicativas en forma de vector, $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_p)^\top$ es un vector que contiene los parámetros lineales desconocidos, los p_j son pesos positivos que suman uno, los términos μ_j que satisfacen la restricción $\sum_{j=1}^g p_j \mu_j = 0$. Siguiendo esta misma línea se propone en este documento un modelo de regresión cuyo error es una mixtura de distribuciones SHN, el trabajo realizado por Benites et al. (2016) realizaron similar planteamiento de modelos de regresión lineal con mixtura finita de colas pesadas. La ventaja de estos modelos de regresión son su capacidad de modelar datos con mayor flexibilidad cuando presentan multimodalidad y provienen de poblaciones heterogéneas siendo en estos casos conveniente considerar una mixtura finita de distribuciones. Entre los tópicos a ser estudiados consideramos: la estimación de parámetros utilizando un algoritmo de maximización con esperanza condicional (ECM), análisis de simulación y aplicaciones con información real para mostrar la utilidad práctica del modelo.

1.2. Objetivos del documento de trabajo

El propósito de este documento es plantear un modelo de ajuste lineal usando una mixtura finita de distribuciones SHN que será denominado Modelo de regresión lineal usando una mixtura de distribuciones senh-normal (MR-MF-SHN), esto incluye la metodología para la estimación de parámetros y su aplicación con información real. Específicamente:

- Plantear y poner en práctica un algoritmo ECM para llevar a cabo la estimación de los parámetros.
- Llevar a cabo análisis de simulación para demostrar el rendimiento del modelo planteado.
- Utilizar nuestro modelo con información real para ejemplificar la metodología empleada.
- Implementar el modelo propuesto en el lenguaje R e incorporarlo al paquete `bssn`.

1.3. Estructura del documento

En el capítulo 2, se presenta los conceptos iniciales relacionados a la distribución senh-normal. En el capítulo 3 se examina el modelo planteado, modelo de regresión lineal usando mixtura de distribución senh-normal desde un enfoque estadístico tradicional, teniendo en cuenta el algoritmo ECM para el cálculo de sus variables. En el capítulo 4 se presenta las conclusiones de la simulación para la obtención de los parámetros, precisión de las estimaciones de los parámetros y propiedad de consistencia de los estimadores. Al finalizar, se presenta el capítulo 5 donde se demuestra la implementación del modelo con un conjunto de datos reales.

Capítulo 2

Distribución Senh-Normal

En relación al trabajo realizado por Sánchez et al. (2002) se menciona que no todos los fenómenos de la naturaleza pueden ser descritos por la distribución normal, siendo necesarias otras distribuciones para modelar ciertos comportamientos aleatorios. Históricamente, se construyeron distribuciones no normales asumiendo que una transformación de la variable aleatoria tenía una distribución normal estándar. También se han hecho transformaciones que pueden ser representadas por polinomios. En la presente sección se considerará la transformación logarítmica planteada por Sánchez et al. (2002) en una variable aleatoria T que se distribuye de acuerdo a una distribución BS.

2.1. Distribución acumulada de la distribución Senh-Normal

La función de de la variable aleatorio Y esta dado por:

$$F(y) = \Phi \left[\frac{2}{\alpha} \sinh \left(\frac{y - \gamma}{\sigma} \right) \right], \quad (-\infty < y < \infty). \quad (2.1)$$

y de densidad

$$f(y) = \left(\frac{2}{\alpha\sigma\sqrt{2\pi}} \right) \cosh \left(\frac{y - \gamma}{\sigma} \right) \exp \left[-\frac{2}{\alpha^2} \sinh^2 \left(\frac{y - \gamma}{\sigma} \right) \right],$$

con distribución SHN, denotada por $\text{SHN}(\alpha, \gamma, \sigma)$, donde γ es parámetro de posición, σ el parámetro de tamaño y α parámetro de forma; α y σ son positivos, mientras que γ pertenece a los reales.

A continuación se establecerá la conexión existente entre la distribución SHN y la distribución BS, demostrado en Sánchez et al. (2002).

Teorema 2.1.1. *Sea $T \sim \text{BS}(\alpha, \beta)$. Entonces, $\ln(T) \sim \text{SHN}(\alpha, \gamma, \sigma = 2)$, donde $\gamma = \ln(\beta)$.*

2.2. Propiedades de la Distribución Senh-Normal

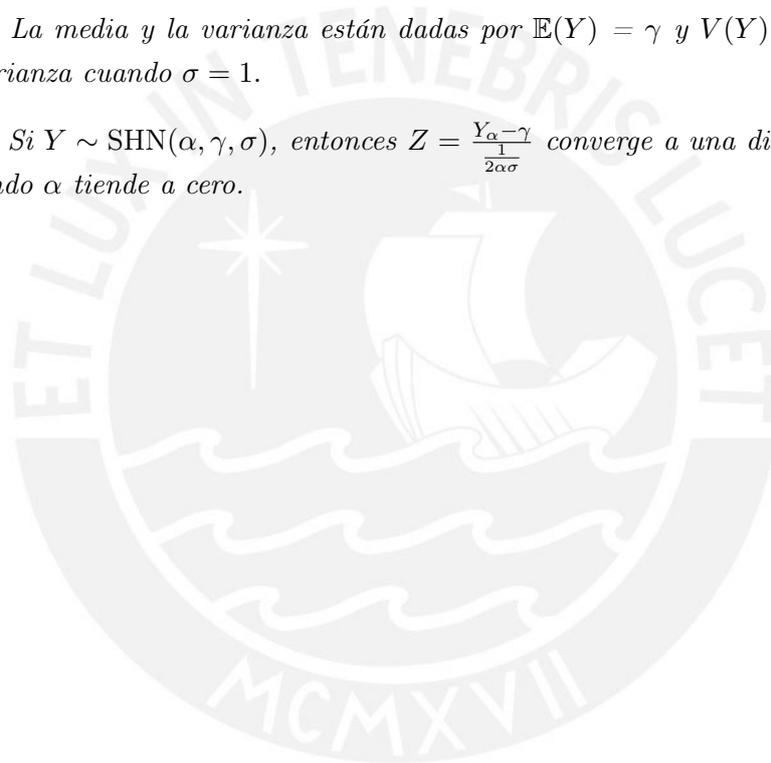
La distribución SHN tiene algunas propiedades que se verán en los siguientes lemas (Rieck, 1989) :

Lema 2.2.1. *La distribución SHN es simétrica respecto al parámetro de posición γ .*

Lema 2.2.2. *La distribución SHN es estrictamente unimodal para $\alpha \leq 2$ y bimodal para $\alpha > 2$.*

Lema 2.2.3. *La media y la varianza están dadas por $\mathbb{E}(Y) = \gamma$ y $V(Y) = \sigma^2\omega(\alpha)$, donde $\omega(\alpha)$ es la varianza cuando $\sigma = 1$.*

Lema 2.2.4. *Si $Y \sim \text{SHN}(\alpha, \gamma, \sigma)$, entonces $Z = \frac{Y - \gamma}{\sigma}$ converge a una distribución Normal estándar cuando α tiende a cero.*



Capítulo 3

Modelo de regresión lineal basado en la mixtura de distribuciones Senh-Normal

Basándose en las ideas desarrolladas por Bartolucci y Scaccia (2005) se propone el modelo de regresión lineal en el cual el error sigue una distribución MF-SHN, denotado por MR-MF-SHN. Se considera el siguiente modelo de regresión

$$Y_i = \beta_0 + \mathbf{x}_i^\top \boldsymbol{\beta} + \varepsilon_i, \quad i = 1, \dots, n, \quad (3.1)$$

en el cual Y_i son los registros de la variable a explicar, $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})^\top$ son variables independientes $(p+1) \times 1$ y $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_p)^\top$ es el de vector coeficientes de regresión. Además, se asume que

$$\varepsilon_i \sim \sum_{j=1}^g p_j \phi_{\text{SHN}}(\varepsilon_i | \alpha_j, \mu_j, \sigma_j = 2), \quad i = 1, \dots, n, \quad j = 1, \dots, g, \quad (3.2)$$

en el que $\phi_{\text{SHN}}(\cdot)$ representa la función de densidad de probabilidad (fdp) de una distribución SHN($\alpha_j, \mu_j, \sigma_j = 2$) donde $\sum_{j=1}^g p_j = 1$ y los μ_j deben cumplir la restricción de identificación $\sum_{j=1}^g p_j \mu_j = 0$. En consecuencia, debido a la linealidad de las distribución de SHN y teniendo en cuenta que $E(\varepsilon_i) = 0$, la función de densidad de probabilidad de Y_i es

$$f(y_i | \boldsymbol{\theta}) = \sum_{j=1}^g p_j \phi_{\text{SHN}}(y_i | \alpha_j, \mu_{ij}, \sigma_j = 2), \quad \mu_{ij} = \beta_0 + \mathbf{x}_i^\top \boldsymbol{\beta} + \mu_j = \vartheta_j + \mathbf{x}_i^\top \boldsymbol{\beta}, \quad (3.3)$$

en el cual $\mu_{ij} = \mathbf{x}_i^\top \boldsymbol{\beta} + \vartheta_j$, $\vartheta_j = \beta_0 + \mu_j$ y $\boldsymbol{\theta} = (\alpha_1, \dots, \alpha_g, \boldsymbol{\beta}^\top, (p_1, \dots, p_{g-1})^\top, \vartheta_1, \dots, \vartheta_g)^\top$ es el vector con todos los parámetros. Dado $\mathbf{y} = (y_1, \dots, y_n)^\top$, la forma de log-verosimilitud

se representa por

$$\ell(\boldsymbol{\theta}|\mathbf{y}) = \sum_{i=1}^n \log \left(\sum_{j=1}^g p_j \phi_{\text{SHN}}(y_i | \alpha_j, \mu_{ij}, \sigma_j = 2) \right). \quad (3.4)$$

Para cada i y j , consideramos la variable indicadora latente Z_{ij} tal que:

$$P(Z_{ij} = 1) = 1 - P(Z_{ij} = 0) = p_j, \quad \sum_{j=1}^g p_j = 1 \quad \text{y} \quad y_i | Z_{ij} = 1 \sim \text{SHN}(\alpha_j, \mu_{ij}, \sigma = 2). \quad (3.5)$$

Se puede apreciar que luego de integrar $\mathbf{Z}_i = (Z_{i1}, \dots, Z_{ig})^\top$ se obtiene la densidad marginal de una distribución SHN y $\mathbf{Z}_1, \dots, \mathbf{Z}_n$ representan vectores aleatorios independientes, cada uno de ellos posee una distribución multinomial con función de densidad de probabilidad

$$f(\mathbf{z}_i) = p_1^{z_{i1}} p_2^{z_{i2}} \dots (1 - p_1 - \dots - p_{g-1})^{z_{ig}},$$

la cual se denota por $\mathbf{Z}_i \sim M(1; p_1 \dots, p_g)$. Dichos vectores latentes aparecen en la representación jerárquica que presentaremos a continuación, la cual es usada para construir el algoritmo ECM. De la expresión (3.5), el MR-MF-SHN puede ser representado como:

$$Y_i | Z_{ij} = 1 \stackrel{\text{ind}}{\sim} \text{SHN}(\alpha_j, \mu_{ij}, \sigma = 2), \quad (3.6)$$

$$\mathbf{Z}_i \stackrel{\text{iid}}{\sim} M(1; p_1 \dots, p_g), \quad i = 1, \dots, n, \quad j = 1, \dots, g. \quad (3.7)$$

3.1. Algoritmo de maximización de esperanza condicional

En estadística, un algoritmo de optimización de valor esperado condicional (Expectation Conditional Mean, ECM) (Meng y Rubin (1993) y Sexton y Swensen (2000)) es un proceso iterativo para encontrar estimadores de máxima verosimilitud, donde el proceso depende de valores inferidos no observados. La iteración ECM es una variación del algoritmo de Esperanza-Maximización (EM), en donde primero se realiza un paso de esperanza (E), que genera una función para la esperanza de la log-verosimilitud utilizando los estimadores actuales de los parámetros, y luego se agrega un paso de maximización condicional (MC), que calcula los parámetros optimizando el valor esperado encontrado en el paso (E). El cálculo de estos parámetros se emplean para establecer la distribución de las variables no observables en el siguiente paso (E).

3.2. Cálculo de los parámetros a través del algoritmo ECM

La puesta en práctica del algoritmo ECM, que permite obtener los estimadores de MV de los parámetros del MR-MF-SHN. De (3.6) y (3.7), se obtiene la subsiguiente ecuación de logaritmo de verosimilitud completa

$$\ell_c(\boldsymbol{\theta}|\mathbf{y}, \mathbf{z}) = c + \sum_{i=1}^n \sum_{j=1}^g Z_{ij} \left\{ \log p_j + \log(\xi_{1ij}) - \frac{1}{2} \xi_{2ij}^2 \right\},$$

en el cual

$$\xi_{1ij} = \frac{z}{\alpha_j} \cosh\left(\frac{Y_i - \mu_{ij}}{2}\right); \xi_{2ij} = \frac{2}{\alpha_j} \sinh\left(\frac{Y_i - \mu_{ij}}{2}\right),$$

y c es un coeficiente independiente del conjunto de parámetros $\boldsymbol{\theta}$. Se establece la siguiente esperanza condicional

$$\widehat{z}_{ij} = E[Z_{ij}|\widehat{\boldsymbol{\theta}}, y_i],$$

y haciendo uso de las características de la esperanza condicional, se concluye que:

$$\widehat{z}_{ij} = \frac{\widehat{p}_j \phi_{\text{SHN}}(y_i|\alpha_j, \mu_{ij}, \sigma_j = 2)}{\sum_{j=1}^g \widehat{p}_j \phi_{\text{SHN}}(y_i|\alpha_j, \mu_{ij}, \sigma_j = 2)}$$

Así, la función- Q dada por $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)}) = E[\ell_c(\boldsymbol{\theta})|\mathbf{y}, \widehat{\boldsymbol{\theta}}^{(k)}]$ puede ser escrita como,

$$Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)}) = c + \sum_{i=1}^n \sum_{j=1}^g \left(\widehat{z}_{ij}^{(k)} \left(\log(p_j) + \log(\xi_{1ij}) - \frac{1}{2} \xi_{2ij}^2 \right) \right).$$

Luego, el método ECM para obtener los estimadores de MV de $\boldsymbol{\theta}$ está compuesto de la siguiente forma (los cuales están demostrados en el anexo de Demostraciones de este documento):

Paso-E: Considerando que $\widehat{\boldsymbol{\theta}}^{(k)}$, determinar \widehat{z}_{ij} , cuando $i = 1, \dots, n$ y $j = 1, \dots, g$.

Paso-MC 1: Fijando $\widehat{\boldsymbol{\beta}}^{(k)}$, $\widehat{\boldsymbol{\vartheta}}_j^{(k)}$ y actualizamos $\widehat{p}_j^{(k)}$, $\widehat{\alpha}_j^{(k)}$ que tienen las siguientes expresiones:

$$\widehat{p}_j^{(k+1)} = \frac{1}{n} \sum_{i=1}^n \widehat{Z}_{ij}^{(k)}$$

$$\alpha_j^{2(k)} = \frac{\sum_{i=1}^n \widehat{z}_{ij}^{(k)} \xi_{3ij}^2}{\sum_{i=1}^n \widehat{z}_{ij}^{(k)}}$$

Paso-MC 2: Fijando $\widehat{p}_j^{(k+1)}$, $\widehat{\alpha}_j^{(k+1)}$ y actualizando $\widehat{\boldsymbol{\beta}}^{(k+1)}$, $\widehat{\boldsymbol{\vartheta}}_j^{(k+1)}$ por maximización de la función función- Q tenemos que

$$\widehat{\boldsymbol{\beta}}^{(k+1)} = \arg \max_{\boldsymbol{\beta}} Q(\widehat{\boldsymbol{\alpha}}^{(k+1)}, \boldsymbol{\beta}, \widehat{\boldsymbol{p}}^{(k+1)}, \widehat{\boldsymbol{\vartheta}}^{(k)}|\widehat{\boldsymbol{\theta}}^{(k)}),$$

$$\widehat{\boldsymbol{\vartheta}}^{(k+1)} = \arg \max_{\boldsymbol{\vartheta}} Q(\widehat{\boldsymbol{\alpha}}^{(k+1)}, \boldsymbol{\beta}^{(k+1)}, \widehat{\boldsymbol{p}}^{(k+1)}, \boldsymbol{\vartheta}|\widehat{\boldsymbol{\theta}}^{(k)}).$$

De acuerdo con Bartolucci y Scaccia (2005) se obtienen los parámetros estimados de β_0 y μ_j

según este procedimiento:

$$\widehat{\beta}_0^{(k+1)} = \sum_{j=1}^g \widehat{p}_j^{(k+1)} \widehat{\vartheta}_j^{(k+1)}$$

y para μ_j , $j = 1, \dots, g$, como

$$\widehat{\mu}_j = \widehat{\vartheta}_j - \widehat{\beta}_0.$$

Este procedimiento se realiza de forma repetida hasta que se alcance un criterio de finalización adecuado

3.2.1. Criterio de detención

Para verificar la convergencia del algoritmo EM para el MR-MF-SHN, se utiliza el criterio de detención según la aceleración de Aitken:

$$|\ell_\infty^{(k+1)} - \ell^{(k+1)}| < \varepsilon,$$

en el cual $\ell^{(k+1)}$ es el logaritmo de verosimilitud observado en $\boldsymbol{\theta}^{(k)}$ y ε es el la margen aceptable. Para los cálculos numéricos que se mostrarán en la siguiente sección y las aplicaciones de este documento, se considerará un $\varepsilon = 10^{-6}$. Considerando que los estimadores de máxima verosimilitud (EMV) convergen, $\widehat{\boldsymbol{\theta}}$, también $\ell_\infty^{(k+1)}$ es asintóticamente calculado del logaritmo de verosimilitud en la iteración $k + 1$, (McLachlan y Krishnan, 2008), en el cual,

$$\ell_\infty^{(k+1)} = \ell^{(k)} + \frac{\ell^{(k+1)} - \ell^{(k)}}{1 - c^{(k)}},$$

con $c^{(k)}$ representando el acelerador de Aitken en la iteración k , denotado por,

$$c^{(k)} = \frac{\ell^{(k+1)} - \ell^{(k)}}{\ell^{(k)} - \ell^{(k-1)}}.$$

El método expuesto anteriormente también es válido para un caso más sencillo, es decir, al considerar $Z_{ij} = 1$, ($g = 1$).

3.2.2. Valores iniciales

Los modelos de mixtura pueden presentar una función de verosimilitud con varias modas, por tal motivo, el proceso de estimación de máxima verosimilitud a través del algoritmo EM puede no alcanzar una solución global máxima si los valores iniciales están alejados de los valores reales de las variables. A continuación, se resume el procedimiento para obtener los valores iniciales.

- Ejecutar el algoritmo de agrupamiento k -medias (Benites et al., 2016) o k -medoids (Kaufman y Rousseeuw, 1990) para inicializar con respecto a un centro de clúster elegido

al azar;

- Inicializar el indicador $\widehat{Z}_j^{(0)} = \{\widehat{z}_{ij}^{(0)}\}_{i=1}^G$ de acuerdo con el algoritmo de agrupamiento escogido en el ítem anterior;
- Establecer las proporciones de la mezcla de la siguiente forma: $\widehat{p}_j^{(0)} = (1/n) \sum_{i=1}^n \widehat{z}_{ij}^{(0)}$. Esto es, el valor de inicio para p_j ;
- Para conseguir los valores de inicio para β_0 y β se emplea el método de mínimos cuadrados ordinarios. De acuerdo a los grupos obtenidos usando el método de agrupamiento se calculan los valores iniciales para p_j , μ_j y α_j .

3.2.3. Selección del modelo

Tomando en consideración el desafío de comparar varios modelos de MR-MF-SHN, con distintos valores de densidades de componentes. Emplearemos dos enfoques para la selección de modelos: el criterio de información de Akaike (AIC) más un término de corrección de sesgo (Hurvich y Tsai, 1989), que se representa por (AICc), y el criterio de información Schwarz o bayesiano ajustado (Schlove, 1987), representado por (BICa). Estos criterios se definen de la siguiente manera:

$$\text{AIC}_c = -2\ell(\widehat{\theta}) + \frac{2n\rho}{n - \rho - 1} \quad \text{y} \quad \text{BIC}_a = -2\ell(\widehat{\theta}) + \rho \log \left(\frac{n+2}{24} \right)$$

en el cual $\ell(\widehat{\theta})$ es la expresión del logaritmo de verosimilitud evaluada en el EMV, ρ es el número de parámetros que deben estimarse, y n es el tamaño de la muestra.

3.2.4. Desviación estándar estimado

Una forma sencilla de obtener la estimación de la desviación estándar es aproximando la matriz de covarianza asintótica de $\widehat{\theta}$ mediante la inversa de la matriz de información observada. Consideremos $\mathbf{I}_o(\theta)$ como la matriz de información observada, en el cual $\ell(\theta|\mathbf{y})$ es el procedimiento del logaritmo de verosimilitud observada en (3.4). En este estudio, empleamos el enfoque alternativo propuesto por Basford et al. (1997), el cual se basa en una aproximación de la inversa de la matriz de covarianza por:

$$\mathbf{I}_o(\widehat{\theta}) = \sum_{i=1}^n \widehat{\mathbf{s}}_i \widehat{\mathbf{s}}_i^\top, \quad \text{donde} \quad \widehat{\mathbf{s}}_i = \frac{\partial}{\partial \theta} \log \left(\sum_{j=1}^g \phi_{\text{SHN}}(y_i | \alpha_j, \mu_{ij}, \sigma = 2) \right) \Bigg|_{\theta = \widehat{\theta}}, \quad (3.8)$$

donde

$$\widehat{\mathbf{s}}_i = (\widehat{s}_{i,\alpha}, \widehat{s}_{i,p_1}, \dots, \widehat{s}_{i,p_{g-1}}, \widehat{s}_{i,\beta}, \widehat{s}_{i,\vartheta_1}, \dots, \widehat{s}_{i,\vartheta_g})^\top.$$

las expresiones para los elementos $\widehat{s}_{i,\alpha}, \widehat{s}_{i,p_j}, \widehat{s}_{i,\beta}, \widehat{s}_{i,\vartheta_j}$, $j = 1, \dots, g$, son las siguientes:

$$\widehat{s}_{i,\beta} = \frac{\sum_{j=1}^g p_j D_\beta(y_i; \boldsymbol{\theta}_j)}{f(y_i; \boldsymbol{\theta})}, \quad \widehat{s}_{i,\vartheta_j} = \frac{p_r D_{\vartheta_j}(y_i; \boldsymbol{\theta}_j)}{f(y_i; \boldsymbol{\theta})}, \quad \widehat{s}_{i,\alpha_j} = \frac{p_r D_{\alpha_j}(y_i; \boldsymbol{\theta}_j)}{f(y_i; \boldsymbol{\theta})}$$

y

$$\widehat{s}_{i,p_j} = \frac{1}{f(y_i; \boldsymbol{\theta})} (\phi_{\text{SHN}}(y_i | \alpha_j, \mu_{ij}, \sigma = 2) - \phi_{\text{SHN}}(y_i | \alpha_g, \mu_{ig}, \sigma = 2)),$$

Obteniendo (mayores detalles se encuentran en el anexo de demostraciones de este documento):

$$\begin{aligned} D_\beta(y_i; \boldsymbol{\theta}_j) &= \frac{1}{4} [\phi(\xi_{2ij}) \xi_{2ij} (\xi_{1ij}^2 - 1)] \mathbf{x}_i, \\ D_{\vartheta_j}(y_i; \boldsymbol{\theta}_j) &= \frac{1}{4} [\phi(\xi_{2ij}) \xi_{2ij} (\xi_{1ij}^2 - 1)], \\ D_{\alpha_j}(y_i; \boldsymbol{\theta}_j) &= \frac{1}{2\alpha_j} [\phi(\xi_{2ij}) \xi_{1ij} (\xi_{2ij}^2 - 1)]. \end{aligned}$$

Ahora, regresando a la parametrización original, se calcula la matriz Hessiana para el vector de parámetros originales $\boldsymbol{\theta}^* = (\alpha_1, \dots, \alpha_g, (p_1, \dots, p_{g-1})^\top, \beta_0, \beta^\top, \mu_1, \dots, \mu_g)^\top$. Utilizando el método de delta se obtiene

$$\mathbf{I}_o(\widehat{\boldsymbol{\theta}}^*) = \mathbf{J}(\boldsymbol{\theta}^* | \boldsymbol{\theta}) \mathbf{I}_o(\widehat{\boldsymbol{\Theta}})^{-1} \mathbf{J}(\boldsymbol{\theta}^* | \boldsymbol{\theta})^\top,$$

donde $\mathbf{J}(\boldsymbol{\theta}^* | \boldsymbol{\theta})$ es una matriz de dimensión $(p + 4g) \times (p + 4g - 1)$, definida por:

$$\mathbf{J}(\boldsymbol{\theta}^* | \boldsymbol{\theta}) = \frac{\partial \boldsymbol{\theta}^*}{\partial \boldsymbol{\theta}} = \begin{pmatrix} \mathbf{I}_g & \mathbf{0}_{g \times (g-1)} & \mathbf{0}_{g \times p} & \mathbf{0}_{g \times g} \\ \mathbf{0}_{(g-1) \times g} & \mathbf{I}_{(g-1) \times (g-1)} & \mathbf{0}_{(g-1) \times p} & \mathbf{0}_{(g-1) \times g} \\ \mathbf{0}_{1 \times g} & \mathbf{0}_{1 \times (g-1)} & \mathbf{0}_{1 \times p} & \mathbf{1}_{1 \times g} \\ \mathbf{0}_{p \times g} & \mathbf{0}_{p \times (g-1)} & \mathbf{I}_p & \mathbf{0}_{p \times g} \\ \mathbf{0}_{g \times g} & \boldsymbol{\gamma} & \mathbf{0}_{g \times p} & \mathbf{I}_g \end{pmatrix},$$

donde $\boldsymbol{\gamma} = \frac{\partial \boldsymbol{\mu}}{\partial \mathbf{p}^\top} = \mathbf{A}(\mathbf{p}^{-1})^\top$ es una matriz de dimensión $g \times (g-1)$, con $\mathbf{A} = (\mu_g - \mu_1, \dots, \mu_g - \mu_{g-1})^\top$ y $\mathbf{p}^{-1} = (1/p_1, \dots, 1/p_{g-1})^\top$.

Capítulo 4

Simulación

En este apartado, se llevarán a cabo algunos casos de simulación con el objetivo de ilustrar el rendimiento del modelo planteado. El primer estudio de simulación muestra la consistencia de los errores estándar aproximados para los estimadores de MV de los parámetros a través del algoritmo ECM con cada muestra bajo el criterio de detención en (3.2.1). El segundo estudio de simulación muestra que los estimadores obtenidos a través del algoritmo de ECM tienen buenas propiedades asintóticas.

4.1. Recuperación de parámetros (estudio de simulación 1)

Aquí, se considerarán dos escenarios para el estudio de simulación con el fin de verificar si es posible estimar los valores verdaderos de los parámetros con precisión mediante la aplicación del algoritmo ECM propuesto. Iniciamos con el primer paso para garantizar que el procedimiento funciona correctamente. Para ello se generarán datos artificialmente a partir del siguiente modelo con dos componentes:

Donde Z_{ij} es un componente indicador de Y_i con $P(Z_{ij} = 1) = p_j$, $j = 1, 2$, $\mathbf{x}_i^\top = (x_{i1}, x_{i2})$, tal que $x_{i1} \sim U(0, 1)$ and $x_{i2} \sim U(0, 1)$, $i = 1, \dots, n$, y ε_1 y ε_2 siguen una distribución como en la suposición dada en (3.1). Consideramos los valores de los parámetros como: $\alpha_1 = 0.5$, $\alpha_2 = 1.5$, $\beta_0 = 1$, $\beta_1 = 2$, y $\beta_2 = 3$. Además, consideramos los siguientes escenarios: escenario 1 (componentes bien separados, representados en la Figura 4.1 (a)): $p_1 = 0.2$, $\mu_1 = 4$ y $\mu_2 = -1$, y escenario 2 (componentes ligeramente separados, representados en la Figura 4.1 (b)): $p_1 = 0.2$, $\mu_1 = 3.5$ y $\mu_2 = -0.875$. Para cada conjunto de parámetros y tamaños de muestra, se generaron 1000 muestras de Monte Carlo de tamaño $n = 1000$. Se calcularon las medias y desviaciones estándar (MC Sd) de los estimadores en las 1000 muestras de Monte Carlo, junto con las medias (IM SE) de las desviaciones estándar aproximadas de los estimadores obtenidos mediante el método descrito en la Sección 3.2.4.

Los resultados se muestran en el Cuadro 4.1 y el Cuadro 4.2. Obsérvese que en ambos escenarios (componentes bien separados y poco separados), los resultados sugieren que el modelo

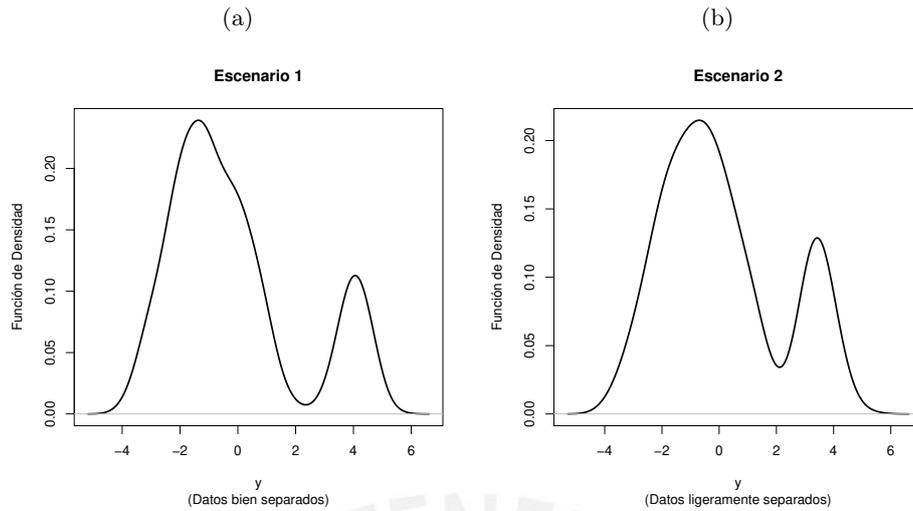


Figura 4.1: Funciones de densidad. (a) Escenario 1 y (b) Escenario 2 del estudio de simulación.

Cuadro 4.1: Simulación 1 - Escenario 1: Media y MC Sd son las respectivas medias y desviaciones estándar estimados del modelo MR-MF-SHN basado en 1000 muestras. IM SE es el resultado promedio del residuo estándar aproximado conseguido con el procedimiento propuesto. Los valores reales están entre paréntesis.

		Escenario 1: (Bien separados)				
Parámetro		$n = 100$	$n = 300$	$n = 500$	$n = 800$	$n = 1000$
$\alpha_1(0,5)$	Media	0.4810	0.4867	0.4914	0.4949	0.4953
	IM SE	0.1290	0.0500	0.0405	0.0289	0.0259
	MC Sd	0.1066	0.0507	0.0381	0.0296	0.0263
$\alpha_2(1,5)$	Media	1.5423	1.4859	1.4938	1.4921	1.4932
	IM SE	0.4938	0.1020	0.1081	0.0422	0.0381
	MC Sd	0.1821	0.0724	0.0546	0.0426	0.0380
$\beta_0(1)$	Media	0.9597	0.9820	0.9795	0.9824	0.9757
	IM SE	0.2033	0.1302	0.1005	0.0761	0.0653
	MC Sd	0.4316	0.1188	0.0862	0.0679	0.0603
$\beta_1(2)$	Media	1.9908	1.9985	2.0003	1.9992	1.9986
	IM SE	0.1003	0.0485	0.0393	0.0293	0.0271
	MC Sd	0.1001	0.0508	0.0396	0.0301	0.0262
$\beta_2(3)$	Media	2.9862	2.9969	2.9965	2.9988	2.9996
	IM SE	0.1282	0.0481	0.0405	0.0308	0.0267
	MC Sd	0.1224	0.0511	0.0412	0.0313	0.0268
$\mu_1(4)$	Media	3.8185	4.0152	4.0258	4.0200	4.0278
	IM SE	1.1104	0.2181	0.1058	0.0827	0.0700
	MC Sd	0.4708	0.1384	0.1013	0.0795	0.0708
$\mu_2(-1)$	Media	-0.8963	-0.9922	-0.9892	-0.9921	-0.9876
	IM SE	0.4939	0.1401	0.0951	0.0686	0.0605
	MC Sd	0.3821	0.0846	0.0592	0.0469	0.0416
$p_1(0,2)$	Media	0.2090	0.1987	0.1972	0.1980	0.1969
	IM SE	0.0705	0.0254	0.0184	0.0135	0.0118
	MC Sd	0.1356	0.0306	0.0218	0.0172	0.0153

MR-MF-SHN propuesto produjo estimaciones satisfactorias.

También podemos ver en esta tabla que el método de estimación de los errores estándar proporciona resultados relativamente cercanos (IM SE y MC Sd), lo que indica que la aproximación asintótica propuesta para las varianzas de los estimaciones MV (A.1) es confiable.

Cuadro 4.2: Simulación 1 - Escenario 2: Media y MC Sd son las respectivas medias y desviaciones estándar estimados del modelo MR-MF-SHN basado en 1000 muestras. IM SE es el resultado promedio del residuo estándar aproximado conseguido con el procedimiento propuesto. Los valores reales están entre paréntesis.

		Escenario 2: (Ligeramente separados)				
Parámetro		$n = 100$	$n = 300$	$n = 500$	$n = 800$	$n = 1000$
$\alpha_1(0,5)$	Media	0.4850	0.4923	0.4999	0.4979	0.5010
	MC Sd	0.1071	0.0535	0.0424	0.0322	0.0275
	IM SE	0.1026	0.0534	0.0411	0.0313	0.0282
$\alpha_2(1,5)$	Media	1.3701	1.4333	1.4482	1.4608	1.4630
	MC Sd	0.0863	0.0437	0.0365	0.0285	0.0261
	IM SE	0.1245	0.0698	0.0541	0.0430	0.0383
$\beta_0(1)$	Media	0.9730	0.9698	0.9565	0.9851	0.9601
	MC Sd	0.1634	0.1052	0.0843	0.0684	0.0555
	IM SE	0.1932	0.1059	0.0809	0.0647	0.0571
$\beta_1(2)$	Media	2.0013	1.9977	1.9962	1.9979	1.9975
	MC Sd	0.0928	0.0523	0.0430	0.0334	0.0291
	IM SE	0.0893	0.0539	0.0429	0.0315	0.0289
$\beta_2(3)$	Media	2.9924	2.9966	2.9986	2.9994	2.9978
	MC Sd	0.1073	0.0493	0.0399	0.0326	0.0286
	IM SE	0.1012	0.0504	0.0411	0.0320	0.0281
$\mu_1(3,5)$	Media	3.5138	3.5260	3.5462	3.5136	3.5403
	MC Sd	0.2892	0.1140	0.0930	0.0747	0.0623
	IM SE	0.2349	0.1263	0.0966	0.0763	0.0677
$\mu_2(-0,875)$	Media	-0.9057	-0.8755	-0.8627	-0.8751	-0.8649
	MC Sd	0.2166	0.0938	0.0765	0.0607	0.0530
	IM SE	0.1272	0.0681	0.0515	0.0415	0.0364
$P_1(0,8)$	Media	0.7934	0.8011	0.8044	0.8006	0.8037
	MC Sd	0.0384	0.0206	0.0169	0.0135	0.0116
	IM SE	0.0516	0.0286	0.0218	0.0175	0.0154

4.2. Propiedades asintóticas de las estimaciones de ECM (estudio de simulación 2)

El enfoque principal en esta simulación es presentar las propiedades asintóticas de los estimadores obtenidos mediante el uso del algoritmo ECM. Como estrategia se generó muestras artificiales del modelo MR-MF-SHN con $x_i^\top = (x_{1i}, x_{2i})$, tal que $x_{1i} \sim U(0, 1)$ y $x_{2i} \sim U(0, 1)$, $i = 1, \dots, n$. Se seleccionaron muestras de tamaños $n = 100, 300, 500$ y 1000 . Los parámetros fueron tomados como $\alpha_1 = 0.5$, $\alpha_2 = 1.5$, $\beta_0 = 1$, $\beta_1 = 2$, $\beta_2 = 3$, $p_1 = 0.2$, $\mu_1 = 4$ y $\mu_2 = -1$. Para cada conjunto de parámetros y cantidad de observaciones, generamos 1000 muestras aleatorias. Para examinar las propiedades asintóticas de los valores obtenidos utilizando el algoritmo ECM, se calculó el sesgo y el error cuadrático medio relativo (RMSE) por cada combinación de tamaño de muestra y valores de los parámetros, para θ_i , son dados por:

$$\text{Sesgo}(\theta_i) = \frac{1}{1000} \sum_{i=1}^{1000} (\hat{\theta}_i^{(j)} - \theta_i) \text{ y } \text{RMSE}(\theta_i) = \sqrt{\frac{1}{1000} \sum_{i=1}^{1000} (\hat{\theta}_i^{(j)} - \theta_i)^2},$$

donde $\hat{\theta}_i^{(j)}$ es el estimador de θ_i para la j -ésima observación. Las conclusiones del sesgo para β_0 , β_1 , β_2 , μ_1 , μ_2 , p_1 , α_1 y α_2 se observan en la Figura 4.2, las conclusiones del RMSE se observan en la Figura 4.3. Se puede ver un patrón de convergencia a cero del sesgo y el RMSE cuando n aumenta para todos los parámetros. Como norma general, podemos decir que el Sesgo y el

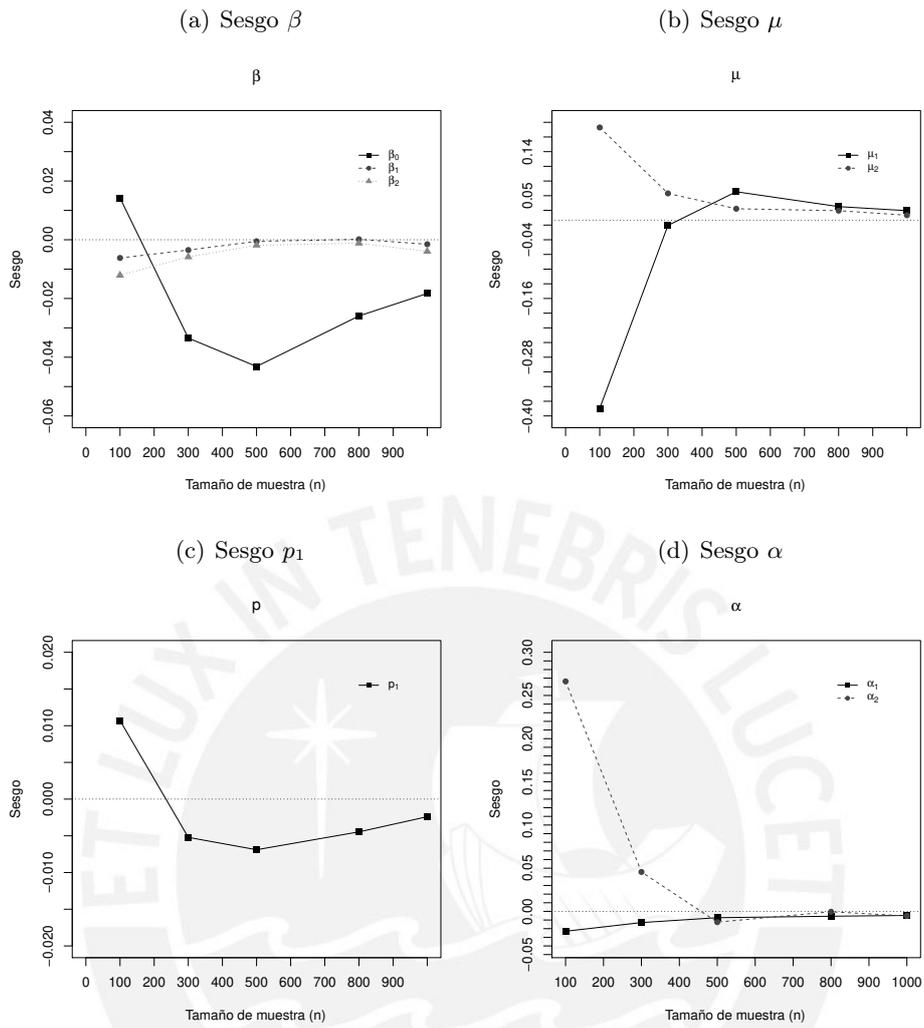


Figura 4.2: Estudio de simulación 2. Sesgo promedio de los estimadores β , μ , p_1 y α

RMSE convergen a cero cuando aumenta el tamaño de las observaciones, indicando que los estimadores basados en el algoritmo de tipo ECM establecido bajo el modelo MR-MF-SHN brindan adecuadas propiedades asintóticas.

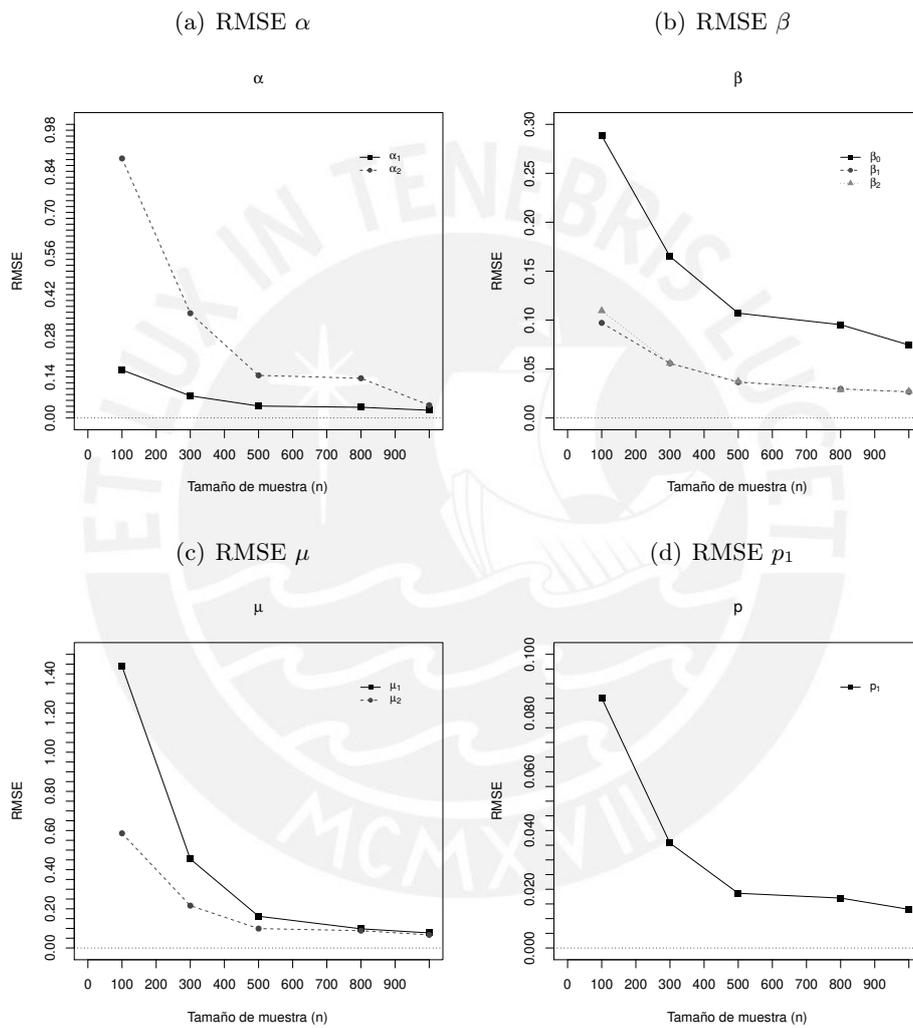


Figura 4.3: Estudio de simulación 2. Promedio de RMSE de los estimadores β , μ , p_1 y α

Capítulo 5

Aplicaciones

En esta capítulo presentaremos los resultados obtenidos de la aplicación del modelo MR-MF-SHN con datos reales, asimismo realizaremos una comparación con un modelo de mixtura finita normal (MR-MF-N).

5.1. Instituto Australiano de Deportes(AIS)

Para la primera aplicación, consideraremos los datos del Australian Institute of Sport (AIS) el cual consta un registro de 202 atletas. Estos valores, fueron analizados previamente por Cook y Weisberg (1994) y posteriormente por Benites et al. (2016) donde se usó un modelo de regresión lineal usando mixtura finita con una distribución de colas pesadas.

Para esta aplicación nos centramos en el porcentaje de grasa corporal (Bfat), que es explicado por la suma de pliegues cutáneos (ssf) y altura en cm (Ht). Por lo tanto, consideraremos el modelo estudiado como:

$$Bfat_i = \beta_0 + \beta_1 ssf_i + \beta_2 Ht_i + \epsilon_i,$$

donde ϵ_i pertenece a la familia MR-MF-SHN para $i = 1, \dots, 202$.

Como primer paso se procedió a ajustar los modelos MR-MF-SHN como está descrito en el capítulo 3. El cuadro 5.1 compara el ajuste con el modelo MR-MF-N para diferentes componentes ($g = 1$ a 5). Observamos que el MR-MF-SHN considerando dos componentes ($g=2$) funciona significativamente mejor que el MR-MF-N, usando los criterios de selección descritos en el capítulo 3.2.3.

El cuadro 5.2 presenta los estimadores de MV de los parámetros para el modelo MR-MF-SHN con $g = 2$, junto con los errores estándar correspondientes (SE).

En la Figura 5.2, se graficó el histograma de los residuales ordinarios junto con las densidades de los residuales de los modelos MR-MF-SHN y MR-MF-N con dos componentes ($g=2$).

Cuadro 5.1: Cuadro comparativo del máximo log-verosimilitud, el criterio de información de Akaike (AIC) y el criterio de información bayesiano (BIC) para los modelos MR-MF-SHN y MR-MF-N usando el conjunto de datos AIS.

Modelo	g	AIC	AICc	Loglik	BIC
FM-N	1	744.5	744.8	-367.2	761.0
FM-N	2	734.4	735.1	-359.2	760.9
FM-N	3	731.1	732.5	-354.6	767.5
FM-SHN	1	1037.3	1037.3	-516.6	1043.9
FM-SHN	2	722.1	722.4	-356.1	738.7
FM-SHN	3	724.5	725.2	-354.2	751.0
FM-SHN	4	731.9	733.3	-354.9	768.3
FM-SHN	5	737.8	740.0	-354.9	784.1

Cuadro 5.2: Estimaciones de los parámetros del modelo MR-MF-SHN con $g = 2$. SE denota los errores estándar correspondientes usando el conjunto de datos AIS.

Parámetro	ML	SE
β_0	16.2905	1.7124
β_1	0.1768	0.0025
β_2	-0.0836	0.0094
α_1	0.8881	0.3460
α_2	3.2334	0.1025
μ_1	-0.2126	0.3592
μ_2	0.4070	0.1405
p_1	0.6569	0.3388

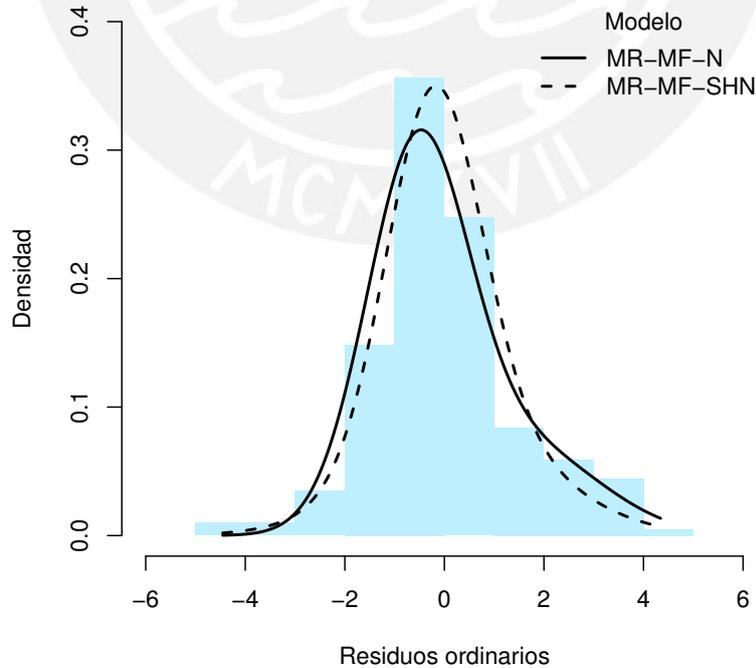


Figura 5.1: Histograma de los residuos ordinarios superpuestos con las densidades de los residuos considerando los modelos MR-MF-SHN y MR-MF-N para $g = 2$ componentes.

Basado en esta representación gráfica, parece una vez más que el modelo MR-MF-SHN tiene un ajuste bastante razonable y mejor que el modelo MR-MF-N con dos componentes ($g=2$). No obstante los valores de los criterios de comparación son algo similares y lo que puede explicar el comportamiento mostrado en la Figura 5.2.

5.2. Clima de Szeged, Hungría, entre 2006 y 2016

Para esta segunda aplicación se utilizará una base inicial de 96,453 registros con datos históricos del clima de Szeged, Hungría, entre 2006 y 2016. Con datos del tipo de clima, tipo de precipitaciones, temperatura, humedad, velocidad del viento, dirección del viento y visibilidad, que fueron utilizados en el artículo (Han et al., 2018). Tal como se realizó en el primer conjunto de datos, se aplicará el modelo MR-MF-SHN en comparación con un modelo de mixtura finita normal (MR-MF-N). Este conjunto de datos fue extraído de una API (Application Programming Interface) proveniente de DarkSky.net, el cual provee datos del clima de cualquier parte del mundo. (Budincsevity, 2016).

Para esta aplicación se tomará en cuenta los registros correspondientes al clima parcialmente nublado y las precipitaciones de tipo nieve, registrando 1,720 datos de tamaño de muestra. Con esta aplicación se pretende estimar la velocidad del viento medido por km/h (WindSpeed), que es explicado por la dirección del viento (WindBearing), la humedad (Humidity), temperatura (Temperature) y la visibilidad en Km. (Visibility). Por lo tanto, consideraremos el modelo estudiado como:

$$WindSpeed_i = \beta_0 + \beta_1 WindBearing_i + \beta_2 Humidity_i + \beta_3 Temperature_i + \beta_4 Visibility_i + \epsilon_i,$$

donde ϵ_i pertenece a la familia MR-MF-SHN para $i = 1, \dots, 1720$.

Posteriormente se procedió a ajustar el modelo MR-MF-SHN. El cuadro 5.3 compara el ajuste con el modelo MR-MF-N para diferentes componentes ($g = 2$ a 5). Observamos que el MR-MF-SHN considerando cinco componentes ($g=5$) funciona significativamente mejor que el MR-MF-N, usando los criterios de selección descritos en el capítulo 3.2.3. Sin embargo note también para el modelo FM-N el modelo que resultó ser mejor es aquel con 3 componentes.

Cuadro 5.4: Valores estimados de los parámetros del modelo MR-MF-SHN con parámetro $g=5$. Se denota los errores estándar correspondientes empleando la muestra de datos históricos del clima de Szeged, Hungría.

Parámetro	$g=5$ ML	$g=5$ SE
β_0	18.3879	19.3152
β_1	0.0021	0.0017
β_2	-10.7878	0.1741
β_3	-12.0182	0.1803
β_4	0.0058	0.0253
α_0	7.8518	2.9694
α_1	36.2538	1.8863
α_2	9.7884	12.3575
α_3	23.5489	5.5231
α_4	24.1992	2.3394
μ_0	-2.7786	9.2051
μ_1	-2.9799	1.1371
μ_2	0.3512	10.9718
μ_3	-4.9603	4.4429
μ_4	5.8762	4.4436

Cuadro 5.3: Comparación del máximo log-verosimilitud, el criterio de información de Akaike (AIC) y el criterio de información bayesiano (BIC) para los modelos MR-MF-SHN y MR-MF-N usando el conjunto de datos históricos del clima de Szeged, Hungría, entre 2006 y 2016.

Modelo	g	AIC	AICc	Loglik	BIC
FM-N	2	10472.0	10472.1	-5226.0	10526.5
FM-N	3	10410.6	10410.8	-5192.3	10481.5
FM-N	4	10421.5	10421.8	-5194.7	10508.7
FM-N	5	10414.5	10414.9	-5188.2	10518.0
FM-SHN	2	11191.5	11191.6	-5590.8	11218.8
FM-SHN	3	10552.1	10552.2	-5268.1	10595.7
FM-SHN	4	10488.8	10488.9	-5233.4	10548.7
FM-SHN	5	10410.0	10410.2	-5191.0	10486.3

El cuadro 5.4 presenta las estimaciones de máxima verosimilitud de los parámetros para el modelo MR-MF-SHN con cinco componentes, junto con los errores estándar correspondientes (SE), obtenidos a través del procedimiento presentado en el capítulo 3.2.4.

Finalmente, en la Figura 5.2 se graficó el histograma de los residuales ordinarios junto con las densidades de los residuales del modelo MR-MF-SHN y MR-MF-N con cinco componentes ($g=5$). Basados en esta representación gráfica, se muestra nuevamente que el modelo MR-MF-SHN tiene un ajuste bastante razonable y mejor que el modelo MR-MF-N con cinco componentes ($g=5$).

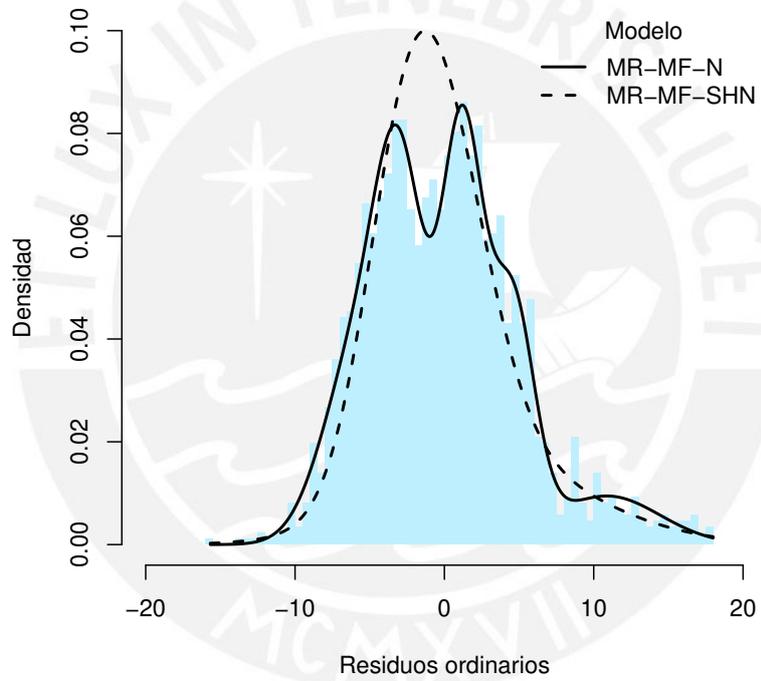


Figura 5.2: Histograma de los residuales ordinarios superpuestos con las densidades de los residuales considerando los modelos MR-MF-SHN y MR-MF-N para $g = 5$ componentes.

Capítulo 6

Conclusiones y sugerencias

6.1. Conclusiones

Del trabajo realizado se concluye:

- Esta investigación propone un nuevo modelo cuyo término de error tiene una mixtura de distribuciones Senh-Normal como una extensión propuesta por Rieck y Nedelman (1991) y Bartolucci y Scaccia (2005).
- El modelo planteado permite modelar con gran flexibilidad, conjunto de datos que presentan multimodalidad y que provienen de poblaciones heterogéneas.
- Para esta estructura se desarrolló el algoritmo ECM para calcular los EMV y la obtención de los errores estándar aproximados utilizando el método planteado por Basford et al. (1997).
- Se elaboraron dos simulaciones para analizar la recuperación de parámetros y las propiedades asintóticas de los estimadores obtenidas a través del algoritmo ECM. Los resultados mostraron resultados satisfactorios en términos de aproximación asintótica, sesgo y RMSE.
- Se realizó la aplicación del modelo propuesto en dos conjuntos de datos reales. El rendimiento del modelo MR-MF-SHN se comparó con el modelo MR-MF-N en diversas condiciones. En el primer conjunto de datos del Instituto Australiano de deportes, los resultados mostraron que para la situación de dos componentes ($g=2$) el modelo planteado se ajusta mucho mejor a los datos, para el segundo conjunto de datos del clima de Szeged, Hungría, entre 2006 y 2016 se identificó que para cinco componentes ($g=5$) se obtiene mejores ajustes con la información. Además, los criterios de selección de modelos indican en ambos casos que el modelo MR-MF-SHN supera al modelo MR-MF-N.

Apéndice A

Anexo de Demostraciones

A.1. Demostraciones del Capítulo 3

Demostrando el Paso MC-1 del algoritmo ECM

Se actualiza $\hat{\theta}^{(k)}$ por optimización, derivando la función $Q(\theta|\hat{\theta}^{(k)})$ con relación a cada parámetro y se equipara a cero para conseguir las ecuaciones de estimación:

$$\hat{p}_j^{(k+1)} = \frac{1}{n} \sum_{i=1}^n \hat{Z}_{ij}^{(k)}$$
$$Q(\theta|\hat{\theta}^{(k)}) = c + \sum_{i=1}^n \sum_{j=1}^g \hat{Z}_{ij}^{(k)} \left\{ \log p_j + \log(\xi_{1ij}) - \frac{1}{2} \xi_{2ij}^2 \right\}$$

Derivadas y expresiones relevantes a considerar:

$$\frac{\partial \xi_{1ij}}{\partial \alpha_j} = -\frac{1}{\alpha_j} \xi_{1ij}$$
$$\frac{\partial \xi_{2ij}^2}{\partial \alpha_j} = -\frac{2}{\alpha_j} \xi_{2ij}^2$$
$$\xi_{2ij}^2 = \frac{1}{\alpha_j^2} \xi_{3ij}^2$$
$$\xi_{3ij} = 2 \sinh\left(\frac{Y_i - \mu_{ij}}{2}\right)$$

Simplificando la expresión:

$$\begin{aligned}
\frac{\partial Q}{\partial \alpha_j} &= \sum_{i=1}^n z_{ij}^{(k)} \left\{ \frac{1}{\xi_{1ij}} \left(-\frac{1}{\alpha_i} \xi_{1ij} \right) - \frac{1}{2} \left(-\frac{2}{\alpha_j} \xi_{2ij}^2 \right) \right\} \\
&= \sum_{i=1}^n \left[-\frac{\widehat{z}_{ij}^{(k)}}{\alpha_j} + \frac{\widehat{z}_{ij}^{(k)}}{\alpha_j^3} \xi_{3ij}^2 \right] \\
&= -\frac{1}{\alpha_j} \sum_{i=1}^n \widehat{z}_{ij}^{(k)} + \frac{1}{\alpha_j^3} \sum_{i=1}^n \widehat{z}_{ij}^{(k)} \xi_{3ij}^2
\end{aligned}$$

Igualando a cero el resultado anterior, obtendríamos:

$$\begin{aligned}
\frac{1}{\alpha_j} \sum_{i=1}^n \widehat{z}_{ij}^{(k)} &= \frac{1}{\alpha_j^3} \sum_{i=1}^n \widehat{z}_{ij}^{(k)} \xi_{3ij}^2 \\
\alpha_j^{2(k)} &= \frac{\sum_{i=1}^n \widehat{z}_{ij}^{(k)} \xi_{3ij}^2}{\sum_{i=1}^n \widehat{z}_{ij}^{(k)}}
\end{aligned}$$

Nótese que $\widehat{\beta}^{(k+1)}$ y $\widehat{v}^{(k+1)}$ no tienen expresiones cerradas, por lo que se obtienen por maximización numérica.

Demostrando el cálculo de las expresiones de los errores estándares aproximados

$$\mathbf{I}_o(\widehat{\boldsymbol{\theta}}) = \sum_{i=1}^n \widehat{\mathbf{s}}_i \widehat{\mathbf{s}}_i^\top, \quad \text{donde} \quad \widehat{\mathbf{s}}_i = \left. \frac{\partial}{\partial \boldsymbol{\theta}} \log \left(\sum_{j=1}^g \phi_{\text{SHN}}(y_i | \alpha_j, \mu_{ij}, \sigma = 2) \right) \right|_{\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}},$$

donde

$$\widehat{\mathbf{s}}_i = (\widehat{s}_{i,\alpha}, \widehat{s}_{i,p_1}, \dots, \widehat{s}_{i,p_{g-1}}, \widehat{s}_{i,\boldsymbol{\beta}}, \widehat{s}_{i,\vartheta_1}, \dots, \widehat{s}_{i,\vartheta_g})^\top.$$

las expresiones para los elementos $\widehat{s}_{i,\alpha}, \widehat{s}_{i,p_j}, \widehat{s}_{i,\beta}, \widehat{s}_{i,\vartheta_j}$, $j = 1, \dots, g$, son las siguientes:

$$\widehat{s}_{i,\beta} = \frac{\sum_{j=1}^g p_j D_\beta(y_i; \boldsymbol{\theta}_j)}{f(y_i; \boldsymbol{\theta})}, \quad \widehat{s}_{i,\vartheta_j} = \frac{p_r D_{\vartheta_j}(y_i; \boldsymbol{\theta}_j)}{f(y_i; \boldsymbol{\theta})}, \quad \widehat{s}_{i,\alpha_j} = \frac{p_r D_{\alpha_j}(y_i; \boldsymbol{\theta}_j)}{f(y_i; \boldsymbol{\theta})}$$

y

$$\widehat{s}_{i,p_j} = \frac{1}{f(y_i; \boldsymbol{\theta})} (\phi_{\text{SHN}}(y_i | \alpha_j, \mu_{ij}, \sigma = 2) - \phi_{\text{SHN}}(y_i | \alpha_g, \mu_{ig}, \sigma = 2)),$$

donde

$$D_{\vartheta_j}(y_i; \boldsymbol{\theta}_j) = \frac{\partial}{\partial \vartheta_j} (\phi_{\text{SHN}}(y_i | \alpha_j, \mu_{ij}, \sigma = 2))$$

$$\phi_{\text{SHN}}(Y_i | \alpha_j, \mu_{ij}, \sigma = 2) = \frac{1}{2} \phi(\xi_{2ij}) \xi_{1ij}$$

$$\xi_{2ij} = \frac{2}{\alpha_j} \sinh \left(\frac{Y_i - v_j - X_i^T \boldsymbol{\beta}}{2} \right)$$

$$\xi_{1ij} = \frac{2}{\alpha_j} \cosh\left(\frac{Y_i - v_j - X_i^T \beta}{2}\right)$$

Derivando respecto a sus parámetros

Primera expresión:

$$\begin{aligned} \frac{\partial \phi_{\text{SHN}}}{\partial \alpha_j} &= \frac{1}{2} \left[\frac{1}{\alpha_j} \xi_{2ij}^2 \phi(\xi_{2ij}) \xi_{1ij} + \phi(\xi_{2ij}) \left(-\frac{1}{\alpha_j} \xi_{1ij} \right) \right] \\ &= \frac{1}{2} \left(\frac{1}{\alpha_j} \xi_{1ij} \phi(\xi_{2ij}) \right) [\xi_{2ij}^2 - 1] \\ &= \frac{1}{2\alpha_j} \xi_{1ij} \phi(\xi_{2ij}) (\xi_{2ij}^2 - 1) \end{aligned}$$

Segunda expresión:

$$\begin{aligned} \frac{\partial \phi_{\text{SHN}}}{\partial \beta} &= \frac{1}{2} \left[\frac{1}{2} \phi(\xi_{2ij}) \xi_{1ij}^2 \xi_{2ij} x_i + \phi(\xi_{2ij}) \left(-\frac{1}{2} \xi_{2ij} x_i \right) \right] \\ &= \frac{1}{4} \phi(\xi_{2ij}) \xi_{1ij}^2 \xi_{2ij} x_i - \frac{1}{4} \phi(\xi_{2ij}) \xi_{2ij} x_i \\ &= \frac{1}{4} \phi(\xi_{2ij}) \xi_{2ij} x_i (\xi_{1ij}^2 - 1) \end{aligned}$$

Tercera expresión:

$$\begin{aligned} \frac{\partial \phi_{\text{SHN}}}{\partial v_j} &= \frac{1}{2} \left[\frac{1}{2} \phi(\xi_{2ij}) \xi_{1ij}^2 \xi_{2ij} + \phi(\xi_{2ij}) \left(-\frac{1}{2} \xi_{2ij} \right) \right] \\ &= \frac{1}{4} \phi(\xi_{2ij}) \xi_{1ij}^2 \xi_{2ij} - \frac{1}{4} \phi(\xi_{2ij}) \xi_{2ij} \\ &= \frac{1}{4} \phi(\xi_{2ij}) \xi_{2ij} (\xi_{1ij}^2 - 1) \end{aligned}$$

Obtenemos:

$$\begin{aligned} D_\beta(y_i; \boldsymbol{\theta}_j) &= \frac{1}{4} [\phi(\xi_{2ij}) \xi_{2ij} (\xi_{1ij}^2 - 1)] \mathbf{x}_i, \\ D_{\vartheta_j}(y_i; \boldsymbol{\theta}_j) &= \frac{1}{4} [\phi(\xi_{2ij}) \xi_{2ij} (\xi_{1ij}^2 - 1)], \\ D_{\alpha_j}(y_i; \boldsymbol{\theta}_j) &= \frac{1}{2\alpha_j} [\phi(\xi_{2ij}) \xi_{1ij} (\xi_{2ij}^2 - 1)]. \end{aligned}$$

Demostrando los valores de la matriz Hessiana

$$\left| \begin{array}{cccc} \frac{\partial \alpha}{\partial \alpha^T} = I_{g \times g} & \frac{\partial \alpha}{\partial p^T} = 0_{g \times (g-1)} & \frac{\partial \alpha}{\partial \beta^T} = 0_{g \times p} & \frac{\partial \alpha}{\partial v^T} = 0_{g \times g} \\ \frac{\partial p}{\partial \alpha^T} = 0_{(g-1) \times g} & \frac{\partial p}{\partial p^T} = I_{(g-1) \times (g-1)} & \frac{\partial p}{\partial \beta^T} = 0_{(g-1) \times p} & \frac{\partial p}{\partial v^T} = 0_{(g-1) \times g} \\ \frac{\partial \beta_0}{\partial \alpha^T} = 0_{1 \times g} & \frac{\partial \beta_0}{\partial p^T} = 0_{1 \times (g-1)} & \frac{\partial \beta_0}{\partial \beta^T} = 0_{1 \times p} & \frac{\partial \beta_0}{\partial v^T} = 1_{1 \times g} \\ \frac{\partial \beta}{\partial \alpha^T} = 0_{p \times g} & \frac{\partial \beta}{\partial p^T} = 0_{p \times (g-1)} & \frac{\partial \beta}{\partial \beta^T} = I_{p \times p} & \frac{\partial \beta}{\partial v^T} = 0_{p \times g} \\ \frac{\partial \mu}{\partial \alpha^T} = 0_{g \times g} & \frac{\partial \mu}{\partial p^T} = A^T (p^{-1})_{g \times (g-1)} & \frac{\partial \mu}{\partial \beta^T} = 0_{g \times p} & \frac{\partial \mu}{\partial v^T} = I_{g \times g} \end{array} \right|$$

$$\frac{\partial \alpha}{\partial v^T} = \begin{vmatrix} \frac{\partial \alpha_1}{\partial v_1} & \frac{\partial \alpha_1}{\partial v_2} & \cdots & \frac{\partial \alpha_1}{\partial v_g} \\ \frac{\partial \alpha_2}{\partial v_1} & \frac{\partial \alpha_2}{\partial v_2} & \cdots & \frac{\partial \alpha_2}{\partial v_g} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial \alpha_g}{\partial v_1} & \frac{\partial \alpha_g}{\partial v_2} & \cdots & \frac{\partial \alpha_g}{\partial v_g} \end{vmatrix} = 0_{g \times g}$$

$$\frac{\partial p}{\partial v^T} = \begin{vmatrix} \frac{\partial p_1}{\partial v_1} & \frac{\partial p_1}{\partial v_2} & \cdots & \frac{\partial p_1}{\partial v_g} \\ \frac{\partial p_2}{\partial v_1} & \frac{\partial p_2}{\partial v_2} & \cdots & \frac{\partial p_2}{\partial v_g} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial p_{g-1}}{\partial v_1} & \frac{\partial p_{g-1}}{\partial v_2} & \cdots & \frac{\partial p_{g-1}}{\partial v_g} \end{vmatrix} = 0_{(g-1) \times g}$$

$$\frac{\partial \beta_0}{\partial \alpha^T} = \begin{vmatrix} \frac{\partial \beta_0}{\partial \alpha_1} & \frac{\partial \beta_0}{\partial \alpha_2} & \cdots & \frac{\partial \beta_0}{\partial \alpha_g} \end{vmatrix} = 0_{1 \times g}$$

$$\frac{\partial \beta_0}{\partial \beta^T} = \begin{vmatrix} \frac{\partial \beta_0}{\partial \beta_1} & \frac{\partial \beta_0}{\partial \beta_2} & \cdots & \frac{\partial \beta_0}{\partial \beta_p} \end{vmatrix} = 0_{1 \times p}$$

$$v_j = \beta_0 + \mu_j$$

$$\beta_0 = v_j - \mu_j$$

$$\sum_{j=1}^g p_j \mu_j = 0$$

$$p_1 \mu_1 + \dots + p_g \mu_g = 0$$

Respecto a lo anterior se procede a derivar

$$\mu_1 = \frac{-p_2\mu_2 - p_3\mu_3 - \dots - p_g\mu_g}{p_1}$$

$$\mu_1 = \frac{-p_2\mu_2 - \dots - (1 - p_1 - p_2 - \dots - p_{g-1})\mu_g}{p_1}$$

$$\frac{\partial\mu_1}{\partial p_1} = \frac{p_1\mu_g - (-p_2\mu_2 - \dots - p_g\mu_g)}{p_1^2}$$

$$\frac{\partial\mu_1}{\partial p_1} = \frac{p_1\mu_g - p_1\mu_1}{p_1^2}$$

$$\frac{\partial\mu_1}{\partial p_1} = \frac{\mu_g - \mu_1}{p_1}$$

$$\frac{\partial\mu_1}{\partial p_2} = \frac{\mu_g - \mu_2}{p_1}$$

$$\frac{\partial\mu_1}{\partial p_{g-1}} = \frac{\mu_g - \mu_{g-1}}{p_1}$$

$$\mu_2 = \frac{-p_1\mu_1 - p_3\mu_3 - \dots - (1 - p_1 - p_2 - \dots - p_{g-1})\mu_g}{p_2}$$

$$\frac{\partial\mu_2}{\partial p_1} = \frac{-\mu_1 + \mu_g}{p_2}$$

$$\frac{\partial\mu_2}{\partial p_2} = \frac{-\mu_2 + \mu_g}{p_2}$$

$$p_1 + p_2 + \dots + p_g = 1$$

$$\frac{\partial\mu}{\partial p^T} = \begin{vmatrix} \frac{\partial\mu_1}{\partial p_1} & \frac{\partial\mu_1}{\partial p_2} & \dots & \frac{\partial\mu_1}{\partial p_{g-1}} \\ \frac{\partial\mu_2}{\partial p_1} & \frac{\partial\mu_2}{\partial p_2} & \dots & \frac{\partial\mu_2}{\partial p_{g-1}} \\ \dots & \dots & \dots & \dots \\ \frac{\partial\mu_g}{\partial p_1} & \frac{\partial\mu_g}{\partial p_2} & \dots & \frac{\partial\mu_g}{\partial p_{g-1}} \end{vmatrix}$$

$$\frac{\partial\mu}{\partial p^T} = \begin{vmatrix} \frac{\mu_g - \mu_1}{p_1} & \frac{\mu_g - \mu_2}{p_1} & \dots & \frac{\mu_g - \mu_{g-1}}{p_1} \\ \frac{\mu_g - \mu_1}{p_2} & \frac{\mu_g - \mu_2}{p_2} & \dots & \frac{\mu_g - \mu_{g-1}}{p_2} \\ \dots & \dots & \dots & \dots \\ \frac{\mu_g - \mu_1}{p_g} & \frac{\mu_g - \mu_2}{p_g} & \dots & \frac{\mu_g - \mu_{g-1}}{p_g} \end{vmatrix}$$

$$\mu_1 = \frac{-p_2\mu_2 - p_3\mu_3 \dots - (1 - p_1 - p_2 \dots - p_{g-1})\mu_g}{p_1}$$

$$\frac{\partial \mu_1}{\partial p_1} = \frac{\mu_g p_1 - p_1 \mu_1}{p_1^2}$$

$$\frac{\partial \mu_1}{\partial p_1} = \frac{\mu_g - \mu_1}{p_1}$$



Bibliografía

- Bartolucci, F. y Scaccia, L. (2005). The use of mixtures for dealing with non-normal regression errors, *Computational Statistics & Data Analysis* **48**(4): 821–834.
- Basford, K., Greenway, D., McLachlan, G. y Peel, D. (1997). Standard errors of fitted component means of normal mixtures, *Computational Statistics* **12**: 1–18.
- Benites, L., Maehara, R. y Lachos, V. H. (2016). Linear regression models with finite mixtures of skew heavy-tailed errors, *Chilean Journal of Statistics* **10**(1): 21–40.
- Birnbaum, Z. W. y Saunders, S. C. (1969). A new family of life distributions, *Journal of Applied Probability* **6**: 637–652.
- Budincsevity, N. (2016). Weather in szeged 2006-2016.
URL: <https://www.kaggle.com/budincsevity/szeged-weather/data>
- Cook, R. D. y Weisberg, S. (1994). *An introduction to regression Graphics*, John Wiley & Sons.
- Han, I., Avron, H. y Shin, J. (2018). Stochastic chebyshev gradient descent for spectral optimization, *Advances in Neural Information Processing Systems*.
- Hurvich y Tsai (1989). Regression and time series model selection in small samples, *Biometrika* **76**: 297–307.
- Johnson, N. L. (1949). Systems of frequency curves generated by methods of translation, *Biometrika* **36**(1-2): 146–176.
- Johnson, N. L. y Kotz, S. (1970). *Distributions in Statistics Continuous Univariate Distributions, Vol. 1*, John Wiley and Sons. New York.
- Kaufman, L. y Rousseeuw, P. (1990). *Finding Groups in Data: An Introduction to Cluster Analysis*, Wiley.
- Leiva, V., Barros, M., Paula, G. y Sanhueza, A. (2008). Generalized birnbaum-saunders distributions applied to air pollutant concentration, *Environmetrics* **19**: 235–249.
- Leiva, V., Vilca, F., Balakrishnan, N. y Sanhueza, A. (2010). A skewed sinh-normal distribution and its properties and application to air pollution, *Communications in Statistics - Theory and Methods* **39**: 426–443.
- McLachlan, G. J. y Krishnan, T. (2008). *The EM algorithm and extensions*, 2nd edn, John Wiley & Sons.
- Meng, X. y Rubin, D. B. (1993). Maximum likelihood estimation via the ecm algorithm: A general framework, *Biometrika* **80**(2): 267–278.

- Rieck, J. R. (1989). *Statistical analysis for the Birnbaum-Saunders fatigue life distribution*, PhD thesis, Clemson University.
- Rieck, J. R. y Nedelman, J. R. (1991). A log-linear model for the birnbaum-saunders distribution, *Technometrics* **33**: 51–60.
- Sclove, L. (1987). Application of model-selection criteria to some problems in multivariate analysis, *Psychometrika* **52**: 333–343.
- Sexton, J. y Swensen, A. R. (2000). Ecm algorithms that converge at the rate of em, *Biometrika* **87**(3): 651–662.
- Sánchez, V. L., Sifuentes, M. C. y Cuello, E. E. (2002). Caracterización y aplicación de la distribución birnbaum saunders como modelo de tiempos de vida, *Editorial de la Universidad Autonoma Agraria Antonio Narro, Mexico* p. 26.

