

PONTIFICIA UNIVERSIDAD CATÓLICA DEL PERÚ

ESCUELA DE POSGRADO



**Mixtura finita de una distribución Birnbaum-Saunders  
basado en la familia de mixtura en parámetros de escala de  
distribuciones normal asimétrica**

TESIS PARA OPTAR POR EL GRADO ACADÉMICO DE MAGÍSTER EN  
ESTADÍSTICA

**AUTOR**

**Daniel Elias Gavidia Pucllas**

**ASESOR**

**Dr. Luis Enrique Benites Sánchez**

**Miembros del jurado**

**Dr. Cristian Luis Bayes Rodriguez**

**Dr. Luis Enrique Benites Sánchez**

**Dra. Zaida Jesus Quiroz Cornejo**

Junio, 2021

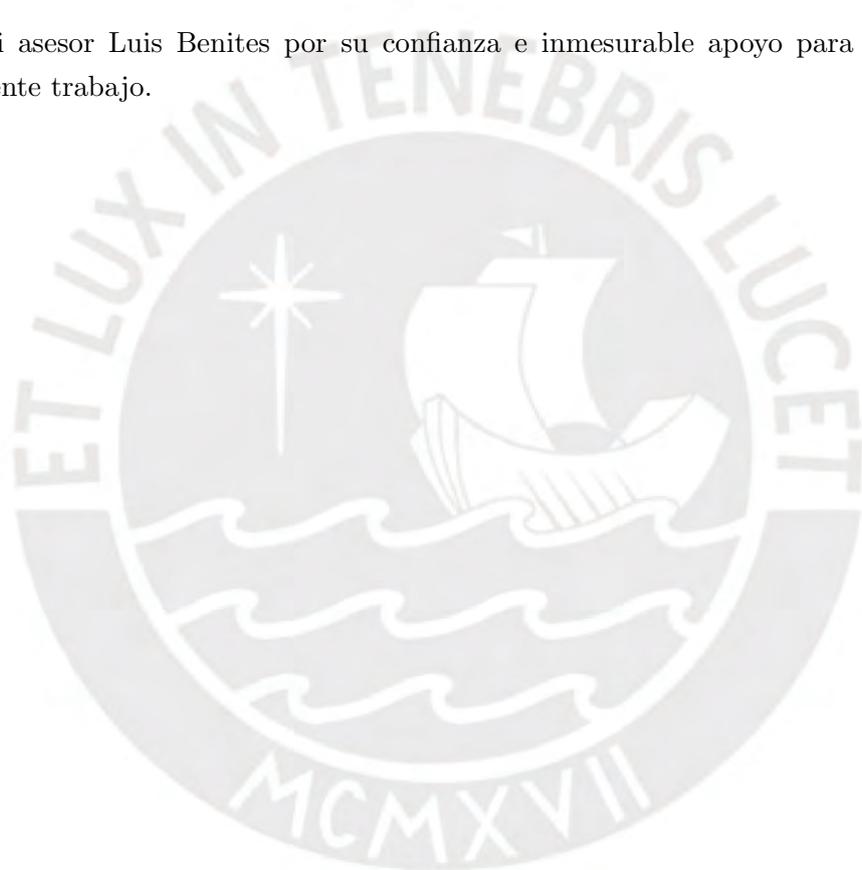
## Dedicatoria

El presente trabajo es dedicado a mi padre Q.E.P.D.



## Agradecimientos

- A mi madre Zeida por su apoyo incondicional.
- A mi tío Luis Gavidia, por sus consejos y constante apoyo.
- A mi amigo Luis Bautista por sus consejos y enseñanzas.
- A mi asesor Luis Benites por su confianza e inmesurable apoyo para desarrollar el presente trabajo.



## Resumen

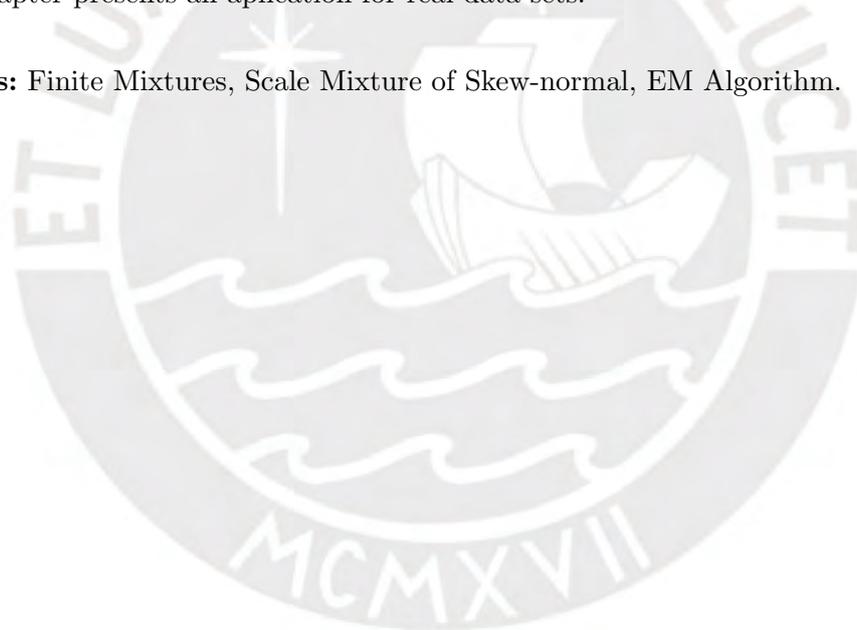
La presente tesis muestra la distribución mixtura de distribuciones Birnbaum-Saunders basados en mixturas de escala normal asimétrica (MF-BS-MENA). Este modelo es una extensión a la propuesta de Maehara (2018a) para datos unimodales basados en distribuciones con mixtura de escala normal asimétrica utilizada para modelar datos con percentiles extremos y altamente concentrados a la izquierda de la distribución. El modelo propuesto permite modelar datos con dos o más componentes de mixtura de distribuciones asimétricas como la  $t$  de Student asimétrica (TA), la Slash asimétrica (SLA), y la normal contaminada asimétrica (NCA). Para estimar los parámetros del modelo propuesto se presenta un método de estimación basado en el algoritmo de maximización condicional de la esperanza (una extensión del algoritmo EM). Además, se desarrollan simulaciones que muestran la precisión de las estimaciones y los errores estándar. Por último, se realizan aplicaciones con un conjunto de datos reales.

**Palabras-clave:** Mixturas Finitas, Mixtura de Escala Normal Asimétrica, Algoritmo EM.

## Abstract

The following thesis presents the finite mixtures of Birbaums-Saunders distributions based on the scale mixture of skew-normal distributions, which are called FM-BS-SMSN. This model is an extension of Maehara (2018a) study of unimodal data based on scale mixture of skew-normal distributions which are used to model extreme percentiles in the left-tail of the distribution. The proposed model can fit two or more mixture of components of skewed distributions like Skew-t, Skew-slash, and skew-contaminated normal. The proposed method to accomplish the estimation of model parameters is the expectation of conditional maximization (an extension of EM algorithm). Moreover, simulations and applications are presented to illustrate the robustness of the proposed estimation method and standar errors. Finally, the last chapter presents an aplication for real data sets.

**Keywords:** Finite Mixtures, Scale Mixture of Skew-normal, EM Algorithm.



# Índice general

<b>Lista de Abreviaturas</b>	<b>VIII</b>
<b>Índice de figuras</b>	<b>IX</b>
<b>Índice de cuadros</b>	<b>X</b>
<b>1. Introducción</b>	<b>1</b>
1.1. Consideraciones Preliminares . . . . .	1
1.2. Objetivos . . . . .	3
1.3. Organización del Trabajo . . . . .	3
<b>2. Conceptos preliminares</b>	<b>4</b>
2.1. Distribución Birnbaum-Saunders . . . . .	4
2.1.1. Función de densidad de probabilidad . . . . .	5
2.1.2. Propiedades . . . . .	5
2.2. Mixturas finitas . . . . .	7
2.3. Distribución MENA (Mixtura de Escala Normal Asimétrica) . . . . .	8
2.4. Algoritmo EM . . . . .	8
<b>3. Modelo propuesto</b>	<b>10</b>
3.1. Introducción . . . . .	10
3.1.1. Casos especiales . . . . .	11
3.2. MF-BS-MENA . . . . .	12
3.2.1. Descripción del modelo . . . . .	12
3.2.2. Estimación de máxima verosimilitud . . . . .	13
3.2.3. Estimación de parámetros mediante el algoritmo ECM . . . . .	13
3.2.4. Algoritmo EM para el MF-BS-MENA . . . . .	15
3.2.5. Especificación de los valores iniciales . . . . .	17
3.2.6. Matriz de información observada . . . . .	17
<b>4. Estudio de Simulación</b>	<b>19</b>
<b>5. Aplicación</b>	<b>26</b>
<b>6. Conclusiones</b>	<b>30</b>
6.1. Conclusiones . . . . .	30
6.2. Sugerencias para investigaciones futuras . . . . .	30

A. Resultados teóricos	31
Bibliografía	35



## Lista de Abreviaturas

GI	Distribución gamma inversa.
BS	Birbaum-Saunders
NA	Normal asimétrica
BS-NA	Birbaum-Saunders normal asimétrica
BS-TA	Birbaum-Saunders $t$ de Student asimétrica
BS-SLA	Birbaum-Saunders Slash asimétrica
BS-NCA	Birbaum-Saunders normal contaminada asimétrica
MF-BS-MEN	Mixtura finita de distribuciones Birbaum-Saunders basadas en la familia de mixtura en escala normal
MF-BS-MENA	Mixtura finita de distribuciones Birbaum-Saunders basadas en la familia de mixtura en escala normal asimétrica
MF-BS-NA	Mixtura finita de una distribución BS Normal asimétrica
MF-BS-TA	Mixtura finita de una distribución BS $t$ -student asimétrica
MF-BS-SLA	Mixtura finita de una distribución BS Slash asimétrica
MF-BS-NCA	Mixtura finita de una distribución BS Normal contaminada asimétrica
EA	Distribución elíptica Asimétrica
IC	Intervalo de confianza
MV	Máxima verosimilitud
i.i.d.	Independiente e idénticamente distribuida
f.d.p.	Función de distribución de probabilidad
f.d.a.	Función de distribución de acumulada
v.a.	Variable aleatoria

## Índice de figuras

2.1.	Gráfico de la densidad de la dsitribución BS con valores dados de $\alpha$ y $\beta = 1$	6
4.1.	Estimaciones de los parámetros de la distribución MF-BS-TA	22
4.1.	Estimaciones de los parámetros de la distribución MF-BS-TA	23
4.2.	Estimaciones de los parámetros de la distribución MF-BS-SLA	24
4.3.	Estimaciones de los parámetros de la distribución MF-BS-NCA	25
5.1.	(a) Histograma con la superposición de la estimación de la f.d.p. por un kernel y (b) diagrama de cajas del índice de masa corporal	26
5.2.	Log-verosimilitud perfilada de $\nu$ para modelos con $G = 2$ componentes para los modelos de (a) MF-BS-TA, (b) y (c) MF-BS-NCA	28
5.3.	(a) y (b) Histograma del índice de masa corporal con la superposición de la estimación de la f.d.p. estimada del modelo de MF-BS-TA con diferentes componentes	29

# Índice de cuadros

4.1. Estimación de parámetros de la distribución MF-BS $t$ de student asimétrica (MF-BS-TA), para 1000 réplicas de diferentes tamaño de muestra sin considerar valores iniciales. Con $\boldsymbol{\alpha} = (0.25, 0.50)^\top$ , $\boldsymbol{\beta} = (5.0, 1.0)^\top$ , $\boldsymbol{\lambda} = (1.0, 3.0)^\top$ , $\mathbf{p} = (0.3, 0.7)^\top$ y $\nu = 4$ . . . . .	20
4.2. Estimación de parámetros de la distribución MF-BS normal contaminada asimétrica (MF-BS-NCA), para 1000 réplicas de diferentes tamaño de muestra sin considerar valores iniciales. Con $\boldsymbol{\alpha} = (0.25, 0.50)^\top$ , $\boldsymbol{\beta} = (5.0, 1.0)^\top$ , $\boldsymbol{\lambda} = (1.0, 3.0)^\top$ , $\mathbf{p} = (0.3, 0.7)^\top$ , $\nu = 0,1$ y $\gamma = 0,1$ . . . . .	21
4.3. Estimación de parámetros de la distribución MF-BS slash asimétrica (MF-BS-SLA), para 1000 réplicas de diferentes tamaño de muestra sin considerar valores iniciales. Con $\boldsymbol{\alpha} = (0.10, 0.50)^\top$ , $\boldsymbol{\beta} = (5.0, 1.0)^\top$ , $\boldsymbol{\lambda} = (-1.0, 3.0)^\top$ , $\mathbf{p} = (0.3, 0.7)^\top$ y $\nu = 4$ . . . . .	21
5.1. Comparación de la máxima log-verosimilitud, BIC para el ajuste del modelo MF-BS-MENA usando el conjunto de datos del índice de masa corporal. El número de parámetros $m$ , el Ranking representa la clasificación de los modelos MF-BS-MENA, iter representa el número de iteraciones hasta la convergencia y tiempo (segundos) de convergencia CPU . . . . .	27
5.2. Estimación de parámetros vía el algoritmo EM con sus correspondientes errores estándar (EE) . . . . .	28

# Capítulo 1

## Introducción

### 1.1. Consideraciones Preliminares

La mixtura finita de distribuciones ha sido ampliamente utilizada para el estudio de datos estructurados de manera heterogénea y para aproximar distribuciones multimodales, asimétricas y con colas pesadas. Estos modelos han sido aplicados en áreas como la medicina, donde destaca el de trabajo realizado por Yau, Lee, Kundu y Ng (2003), quienes realizan una regresión de mixtura de dos componentes con efectos aleatorios para el estudio del tiempo de estadía de recién nacidos. Murphy y Martin (2003) utilizan mixturas de modelos basados en distancias sobre candidatos políticos ordenados por preferencias. Turner (2000) utiliza la mixtura de distribuciones para modelar el cambio en la propagación de enfermedades en plantas. Por último, desarrollos teóricos y estudios más exhaustivos sobre mixturas de distribuciones finitas se pueden encontrar en McLachlan y Peel (2000).

Un tema de investigación frecuente es el estudio del tiempo ( $T$ ) hasta la ocurrencia de un determinado evento ( $Y$ ). Usualmente ( $Y$ ) hace referencia a la falla de un equipo o muerte de un organismo, el estudio de este tipo de eventos se extiende desde la ingeniería, ciencias médicas, geología y economía (Marshall y Olkin, 2007). Entre los modelos más usados destacan destacan la distribución exponencial, la distribución gamma, la distribución Weibull, entre otros (Leiva, 2016) (Marshall y Olkin, 2007).

El concepto de fatiga o tiempo de vida se puede extender más allá de términos cronológicos. Por ejemplo, la fatiga puede ser modelada en términos de fuerzas aplicadas cíclicamente, kilómetros recorridos, flexión y degradación Leiva (2016). Al respecto, Birnbaum y Saunders (1969) propusieron una distribución de dos parámetros para el estudio de variables no negativas. Estas variables aleatorias representan fuerzas cíclicas (cargas) aplicadas a objetos estáticos, los cuales pueden terminar en una ruptura por fatiga propio de las cargas aplicadas.

Birnbaum y Saunders (1969b) obtuvieron originalmente los estimadores de máxima verosimilitud de los parámetros. Mann, Schafer y Singpurwalla (1974) mostraron que la distribución Birnbaum-Saunders (BS) es unimodal. Engelhardt et al. (1981) propusieron intervalos de confianza para los parámetros. Además, presentaron un estudio de test de hipótesis. Desmond (1985) derivó la distribución BS en base a un modelo biológico, el autor también extendió la justificación física para el uso de esta distribución, relajando algunas suposicio-

nes hechas por Birnbaum y Saunders (1969). Desmond (1986) investigó la relación entre la distribución BS y la distribución gamma inversa (GI). Achcar (1993) presentó un método de estimación bayesiana, Lu y Chang (1997) utilizaron métodos bootstrap para construir intervalos de predicción para las observaciones futuras de una distribución BS. Dupuis y Mills (1997) utilizaron métodos robustos de estimación para los parámetros de la distribución BS en presencia de valores extremos.

La distribución Birbaum-Saunders surgió en el contexto de ingeniería de materiales, ya que la distribución es apropiada para describir procesos de degradación acumulativa, esto hace que la distribución BS sea un modelo que puede ser usado en otras áreas. En Leiva et al. (2008) y Leiva et al. (2010) pueden ser encontrados algunos ejemplos de aplicaciones en ciencias de la salud, ambiental, forestal, demográficas, actuarial, financiera, entre otras. Todos estos aspectos han permitido considerar a la distribución BS como un modelo de probabilidad, en lugar de un modelo que permite apenas modelar tiempos de vida (Marshall y Olkin, 2007).

Algunas generalizaciones y aplicaciones de la distribución BS fueron consideradas introduciendo nuevos parámetros o considerando distribuciones simétricas/asimétricas en lugar de la distribución normal, con la intención de obtener modelos más robustos o más flexibles para ser utilizados en problemas reales donde existen observaciones atípicas. Por ejemplo Owen y Padgett (1999) desarrollaron la distribución BS con tres parámetros. Barros, Paula y Leiva (2008) consideraron la distribución  $t$  de Student en lugar de la distribución normal estándar, donde los grados de libertad permiten controlar la curtosis de la distribución resultante, este nuevo modelo representa una extensión robusta de la distribución BS usual. Díaz-García y Leiva (2005) generalizaron la distribución BS basada en la clase de distribuciones elípticas que incluye las distribuciones Cauchy, Laplace, Logística, Normal,  $t$  de Student, entre otras.

La distribución normal asimétrica (NA) ha sido estudiada ampliamente en las últimas décadas: véase, por ejemplo, Azzalini (1985) y Azzalini (2005). Cuando los datos son asimétricos, la distribución NA podría ser utilizada en lugar de la distribución normal. En este caso, los percentiles ubicados en las colas (izquierda o derecha) deben ser acomodados de una mejor manera.

Las extensiones de la distribución BS basada en la familia de distribuciones asimétricas han sido consideradas en base a la clase de distribuciones asimétrico-elípticas. Algunos resultados pueden ser encontrados en Maehara (2018a) Vilca y Leiva (2006a) y Leiva, Vilca y Balakrishnan (2010). En estos trabajos fueron obtenidos resultados teóricos, extendiendo las propiedades de las distribuciones BS y log-BS. El desarrollo de métodos de estimación e inferencia utilizando el algoritmo de maximización condicional de la esperanza (ECM) para la distribución Normal Asimétrica - Birnbaum-Saunders (BS-NA) pueden ser encontrados en Vilca, Santana, Leiva y Balakrishnan (2011).

Balakrishnan et al. (2009) encuentra algunas extensiones de la distribución BS como

el modelo Birbaum-Saunders basado en mixturas de escala normal (BS-MEN). Si bien la BS-MEN puede predecir datos contaminados (outliers), esta tiene dificultades en predecir percentiles extremos en datos altamente concentrados en la izquierda de la distribución. Por otro lado, su análisis sobre la propuesta de Vilca, Santana, Leiva y Balakrishnan (2011) sobre una distribución BS-NA, indica que predice muy bien percentiles a la izquierda de la distribución pero tiene problemas en realizar ajustes de valores extremos en los datos. Al respecto, Maehara (2018a) propone una extensión de la distribución Birnbaum-Saunders basada en las distribuciones con mixturas en parámetros de escala normal asimétrica (BS-MENA), que muestra mejores estimaciones de percentiles extremos en la cola izquierda de la distribución.

Este trabajo propone extender el modelo propuesto por Maehara (2018a), considerando una mixtura finita de la distribución BS-MENA. La mixtura finita de distribuciones BS-MENA es una atractiva clase de distribuciones que contiene como casos especiales a la distribución BS normal asimétrica (BS-NA), BS  $t$  de Student asimétrica (BS-TA), BS normal contaminada asimétrica (BS-NCA) y BS Slash asimétrica (BS-SLA). Además, esta clase de distribuciones permitirá modelar los datos con gran flexibilidad, acomodando simultáneamente la multimodalidad y colas pesadas.

## 1.2. Objetivos

El objetivo general de la tesis es estudiar propiedades, estimación y aplicación de la distribución MF-BS-MENA a un conjunto de datos reales desde el punto de vista de la estadística clásica. Para ello se realizará lo siguiente:

- Desarrollar un tipo de algoritmo EM para obtener los estimadores de máxima verosimilitud, donde se usará diferentes algoritmos de clusterización para la obtención de los valores iniciales.
- Utilizar la matriz de información observada para obtener los errores estándar.
- Realizar estudios de simulación.
- Aplicar el modelo a un conjunto de datos reales.

## 1.3. Organización del Trabajo

En el Capítulo 2 presentamos los conceptos más importantes de la distribución Birnbaum-Saunders y sus principales propiedades. Además, se presenta la formulación de Mixturas Finitas y las distribuciones Mixtura de Escala Normal Asimétrica (MF-BS-MENA). En el Capítulo 3 se describe la estimación de los parámetros de la distribución MF-BS-MENA mediante un tipo de algoritmo EM. Asimismo, se plantea una aproximación a la matriz de información de Fisher desarrollado por Basford, Greenway, McLachlan y Peel (1993), la cual será utilizada para obtener los errores estándar. Asimismo, se plantea el uso de algoritmos de clusterización para la obtención de valores iniciales. En el Capítulo 4 se presentan tres estudios de simulación que buscan mostrar el desempeño del método de estimación propuesto. Por último, en el Capítulo 5 se muestra la aplicación del modelo MF-BS-MENA para el conjunto de datos sobre grasa corporal y tipo de deporte trabajados por Cook y Weisberg (1982).

## Capítulo 2

# Conceptos preliminares

### 2.1. Distribución Birnbaum-Saunders

Birnbaum y Saunders (1969) estudiaron el proceso de fatiga como resultado de aplicar fuerzas o cargas de manera cíclica sobre un objeto estático. Estas cargas son denotadas por  $l_i$ , donde  $i = 1, \dots, m$ , hace referencia a la  $i$ -ésima oscilación repetida  $m$  veces en cada ciclo. Las cargas  $l_i$  causan rupturas graduales de tamaño  $X_i$ , donde el tamaño total de la ruptura del objeto en el  $k$ -ésimo ciclo es  $Y_k$ . La ruptura total luego de  $n$  ciclos está dado por  $S_n = \sum_{k=1}^n Y_k$ . Dicho esto, se define el objeto de estudio  $C$  como el menor número de ciclos hasta la ruptura o falla del objeto, es decir, cuando  $S_n$  supere un límite crítico  $\omega$ . Teniendo en cuenta que  $Y_k$  es una variable aleatoria (i.i.d.) con media  $\mu$  y varianza  $\sigma^2$ , la distribución de  $C$  hasta el punto de ruptura  $n$  se expresa como:

$$P(C \leq n) = P\left(\sum_{i=1}^n \frac{Y_i - \mu}{\sigma\sqrt{n}} > \frac{\omega - n\mu}{\sigma\sqrt{n}}\right).$$

Considerando que  $Y_i$  es i.i.d., el teorema del límite central nos permite obtener

$$Z_n = \sum_{i=1}^n \frac{Y_i - \mu}{\sigma\sqrt{n}} \sim N(0, 1).$$

Además, si definimos el número de ciclos hasta la ruptura del objeto como el tiempo de falla o fracaso  $T$ , una variable aleatoria continua no negativa, entonces podemos decir que la variable aleatoria  $T$  tiene una distribución Birnbaum-Saunders con notación  $T \sim \text{BS}(\alpha, \beta)$ , si su función de distribución acumulada es:

$$F_T(t) = \Phi\left(\frac{1}{\alpha}\xi\left(\frac{t}{\beta}\right)\right), \quad t > 0, \alpha > 0, \beta > 0, \quad (2.1)$$

donde  $\alpha = \frac{\sigma}{\sqrt{\omega\mu}}$ ,  $\beta = \frac{\omega}{\mu}$  y  $\xi(x) = \sqrt{x} - \frac{1}{\sqrt{x}}$ , con  $x > 0$ . Además, tenemos que  $\beta$  es un parámetro de escala y  $\alpha$  es un parámetro de forma. Más aún, de acuerdo a Saunders (1976),  $\beta$  puede entenderse como la mediana de los datos y  $\alpha$  como un parámetro de forma asociado a la fatiga del objeto de estudio.

### 2.1.1. Función de densidad de probabilidad

De acuerdo a la f.d.a. en (2.1) tenemos que:

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

Aplicando la transformación de variables aleatorias la función de densidad de probabilidad (f.d.p.) de  $T \sim \text{BS}(\alpha, \beta)$  es:

$$f_T(t) = \phi \left( \frac{1}{\alpha} \xi \left( \frac{t}{\beta} \right) \right) \frac{\partial}{\partial t} \left( \frac{1}{\alpha} \xi \left( \frac{t}{\beta} \right) \right).$$

Podemos re-escribir la f.d.p. de la distribución BS como:

$$f_T(t) = \phi(a_t(\alpha, \beta)) A_t(\alpha, \beta), \quad t > 0,$$

donde  $a_t(\alpha, \beta) = \frac{1}{\alpha} \xi \left( \frac{t}{\beta} \right)$  y  $A_t(\alpha, \beta) = \frac{\partial}{\partial t} a_t(\alpha, \beta) = t^{3/2} (t + \beta) / (2\alpha\beta^{1/2})$ .

### 2.1.2. Propiedades

Una propiedad importante de la distribución BS es que la variable  $T$  puede ser representada como una función de otra variable aleatoria  $Z$  de la siguiente forma:

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

donde  $Z$  es una variable aleatoria que sigue una distribución normal estándar. La esperanza de  $T$  es  $E(T) = \beta \left( 1 + \frac{\alpha^2}{2} \right)$  y su varianza es  $\text{Var}(T) = \beta^2 \alpha^2 \left( 1 + \frac{5}{4} \alpha^2 \right)$ . En general, el momento  $r$ -ésimo de la distribución BS está dado por la siguiente expresión:

$$E \left( \left( \frac{T}{\beta} \right)^r \right) = \sum_{j=0}^r \binom{2r}{2j} \sum_{k=0}^r \binom{j}{k} \frac{(2(r-j+k))!}{2^{r-j+k} (r-j+k)!} \left( \frac{\alpha}{2} \right)^{2(r-j+k)}.$$

De la función de densidad de la distribución BS, Saunders (1974) presenta las siguientes propiedades:

1.  $bT \sim \text{BS}(\alpha, b\beta)$  para  $b > 0$ ;
2.  $1/T \sim \text{BS}(\alpha, \frac{1}{\beta})$ ;
3.  $V = Z^2 = \frac{1}{\alpha^2} \left( \frac{T}{\beta} + \frac{\beta}{T} - 2 \right) \sim \chi^2(1)$ , con  $E(V) = 1$  y  $\text{Var}(V) = 2$ ;
4.  $Y = \log(T) \sim \text{log-BS}(\alpha, \log(\beta))$ .

La Figura 2.1 muestra la forma de la distribución BS con parámetros  $\beta = 1$  y diferentes valores para  $\alpha$ . Se puede observar como a medida que  $\alpha$  se reduce, la forma de la distribución se vuelve más simétrica alrededor de  $\beta$ . Además, se observa que la función es asimétrica

positiva y unimodal, lo cual permite modelar datos altamente concentrados hacia la izquierda.

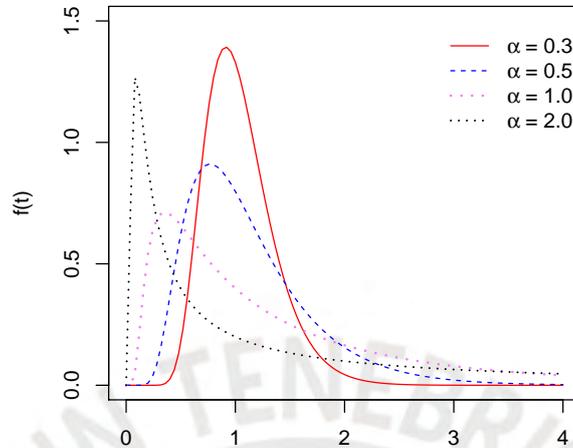


Figura 2.1: Gráfico de la densidad de la dsitribución BS con valores dados de  $\alpha$  y  $\beta = 1$

La distribución BS toma mucho interés porque la propuesta, basada en daños acumulativos en objetos estáticos, cumple ciertas propiedades matemáticas y físicas interesantes. Así, inicialmente la distribución BS fue utilizada en modelos de fatiga como el de Birnbaum y Saunders (1969b), quienes realizan una estimación de  $\beta$  y  $\alpha$  para el tiempo de vida de pequeñas muestras de aluminio *6061-T6* sometidas a distintos grados de estrés. Estudios posteriores tuvieron otras áreas de aplicación, por ejemplo, Jin y Kawczak (2003) utilizaron estimadores kernel basados en la distribución Birbaum-Saunders y la log-normal para estimar las funciones de densidad de tiempos de duración, específicamente, intervalos de tiempo de alta frecuencia en los que se negocian activos financieros. Otra aplicación en datos asimétricos es el de Paula, Leiva, Barros y Liu (2011), quienes realizan una regresión de montos coberturados por seguros en función al tiempo de operatividad.

Otros estudios extienden la distribución BS en funciones de densidad más generales, por ejemplo, Vilca y Leiva (2006b) define una distribución BS generalizada (GBS) basándose en una distribución elíptica asimétrica (EA) definida como  $T \sim \text{GBS}(\alpha, \beta, g, \lambda)$ , donde  $\alpha$  y  $\beta$  son los parámetros definidos en (2.1),  $\lambda$  es el parámetro de asimetría y  $g$  es una función generadora de distribuciones. Agregando a lo anterior, un caso particular de la distribución EA es la distribución normal asimétrica y fue utilizada en la distribución BS por Vilca, Santana, Leiva y Balakrishnan (2011). En dicho estudio, muestran que para percentiles extremos en datos con distribución asimétrica la distribución BS-NA ajusta mejor los valores extremos en los datos, particularmente en datos de concentración de ozono. No obstante, se identificó mayor sesgo relativo en las simulaciones para estimar el parámetro de asimetría.

El trabajo de Maehara (2018a) extiende la distribución BS al considerar familias de distribuciones asimétricas mediante la mixtura de escala normal asimétrica, la cual contiene a las distribuciones normal asimétrica (NA),  $t$  de Student asimétrica (TA), Slash asimétrica (SLA) y normal contaminada asimétrica (NCA). Por otro lado, Benites, Maehara, Vilca y Marmolejo-Ramos (2017) propone la mixtura finita de distribuciones Birbaum-Saunders y un algoritmo de clusterización para la obtención de valores iniciales. Ambos estudios, permiten extender la distribución BS para modelar distribuciones multimodales y asimétricas.

## 2.2. Mixturas finitas

Siguiendo lo descrito por McLachlan y Peel (2000). Sea  $\mathbf{Y} = (Y_1, \dots, Y_n)^\top$  una muestra aleatoria de tamaño  $n$ , donde  $\mathbf{Y}_j$  es un vector  $p$ -dimensional con f.d.p.  $f(\mathbf{y}_j)$  en  $\mathbb{R}^p$ .

$$\mathbf{Y}_j = \begin{bmatrix} y_{j1} \\ \vdots \\ y_{jp} \end{bmatrix} \sim f(\mathbf{y}_j),$$

donde  $\mathbf{Y}_j$  contiene  $p$  variables aleatorias en el  $j$ -ésimo vector aleatorio de  $\mathbf{Y}$ . Además, denotando  $\mathbf{y}_j$  como un valor observado del vector aleatorio  $\mathbf{Y}_j$ , podemos expresar la densidad de  $\mathbf{Y}_j$  como:

$$f(\mathbf{y}_j) = \sum_{i=1}^g \pi_i f_i(\mathbf{y}_j), \quad (2.2)$$

donde  $0 < \pi_i < 1$ , ( $i = 1, \dots, g$ ) y  $\sum_{i=1}^g \pi_i = 1$ , las probabilidades  $\pi_1, \dots, \pi_g$  son los pesos correspondientes a los  $g$  componentes de la mixtura y  $f_i(\mathbf{y}_j)$  es la  $i$ -ésima densidad de la mixtura de distribuciones. Bajo esta definición, en la presente tesis, el número de componentes  $g$  se considera fijo; sin embargo, en algunas aplicaciones  $g$  puede ser estimado a partir de los datos.

Una manera de generar el vector aleatorio  $\mathbf{Y}_j$  es mediante la densidad condicional de  $\mathbf{Y}_j | \mathbf{Z}_j = f_i(\mathbf{y}_j)$ ,  $i = (1, \dots, g)$ , donde  $\mathbf{Z}_j$  es una variable con distribución multinomial, la cual indica qué valores muestrales pertenecen a cada componente  $g$ , con probabilidad  $\pi_1, \dots, \pi_g$ .

$$\mathbf{Z}_j \sim \text{Multinomial}(1, \boldsymbol{\pi}), \quad (2.3)$$

donde  $\boldsymbol{\pi} = (\pi_1, \dots, \pi_g)^T$ . Una interpretación de (2.3) es que  $\mathbf{Y}_j$  se obtiene de una población  $G$  compuesta por  $g$  grupos,  $G_1, \dots, G_g$ , con proporciones  $\pi_1, \dots, \pi_g$ . Si la densidad de  $\mathbf{Y}_j$ , respecto a la población del componente  $G_i$ , es  $f_i(\mathbf{y}_j)$  para  $i = 1, \dots, g$ , entonces la densidad de  $\mathbf{Y}_j$  es una mixtura con  $g$  componentes como en (2.2).

Por último, la estimación del vector de parámetros podría determinarse con los datos observados mediante el método de máxima verosimilitud. No obstante, el vector de componentes de mixturas  $\mathbf{Z}_j$  no se encuentra en la distribución observada de  $\mathbf{Y}_j$ . Por ello, bajo el enfoque del algoritmo EM, se aprovechará la relación entre la distribución conjunta con la densidad marginal y condicional  $f(y, z) = f(y|z)p(z)$  de modo que se obtenga una función de log-verosimilitud más sencilla de operar.

### 2.3. Distribución MENA (Mixtura de Escala Normal Asimétrica)

De acuerdo a Branco y Dey (2001), la distribución MENA es un caso particular de la distribución elíptica asimétrica (EA). Esta clase de distribuciones son una extensión de los modelos propuestos por O'Hagan y Leonard (1976) y Azzalini (2005), los cuales se utilizan para modelar datos asimétricos. Siguiendo a Cabral CRB (2012) y Basso, Lachos, Cabral y Ghosh (2010), una variable  $z$  tiene una distribución con mixtura en parámetros de escala (ME) si su función de densidad tiene la siguiente forma:

$$g(z) = \int_0^{\infty} f(z|u)dH(u),$$

donde  $f(\cdot|u)$  es la función de densidad de  $z$  condicionada a  $U = u$ , siendo  $U$  una variable aleatoria (v.a.) con distribución  $H$ . Además,  $u$  es un factor que afecta al parámetro de escala y  $H$  es la distribución acumulada de  $U$ . Particularmente, una variable aleatoria  $z$  tiene distribución MENA con un parámetro de localización  $\mu \in \mathbb{R}$ , un parámetro de dispersión  $\sigma > 0$  y un parámetro adicional  $\nu$  si su f.d.p. es dada por:

$$\phi_{\text{MENA}}(z) = 2 \int_0^{\infty} \phi(z; \mu, \kappa(u)\sigma^2) \Phi\left(\lambda\kappa^{-1/2}(u)\left(\frac{z-\mu}{\sigma}\right)\right) S(u; \nu), \quad (2.4)$$

cuya notación se expresa como  $z \sim \text{MENA}(\mu, \sigma^2, \lambda; S)$ , donde  $S(\cdot, \nu)$  es la f.d.a. de una variable aleatoria positiva  $U$ ,  $\phi(\cdot)$  y  $\Phi(\cdot)$  es la f.d.p. y f.d.a. respectivamente de una distribución  $N(0, 1)$  y  $\kappa(\cdot)$  es una función positiva de  $U$ . Cuando  $\kappa(U) = 1$  se obtiene la distribución normal asimétrica y si, además,  $\lambda = 0$  se obtiene la distribución Mixtura de Escala Normal (MEN). La distribución MEN, al igual que el caso asimétrico, tiene como casos particulares a la distribución  $t$  de Student, Slash y normal contaminada. Estos casos especiales, asociados a la distribución BS, serán presentados con más detalle en el siguiente capítulo.

### 2.4. Algoritmo EM

El método de estimación de máxima verosimilitud vía el algoritmo EM es uno de los más utilizados para ajustar mixturas finitas de distribuciones. Siguiendo a Tanner (1996), la formulación del algoritmo EM utiliza los datos observados ( $Y$ ) aumentados por variables no observadas o latentes ( $Z$ ), siendo el vector de datos completos definido como  $Y_c = (Y, Z)$ . Posteriormente, se maximiza iterativamente los valores esperados de la log-verosimilitud de los datos completos  $Y_c$ . En otras palabras, dados los datos observados  $Y$ , la función de verosimilitud  $L_c(\psi|Y, Z)$  es proporcional a  $p(Z|\psi^i, Y)$ , donde  $\psi$  es un vector de parámetros de  $\Psi$  en  $\mathbb{R}^d$  y  $p(Z|\psi^i, Y)$  es la distribución de la variable latente  $Z$  condicionado en la evaluación  $i$ -ésima del algoritmo y a los datos observados  $Y$ . Una manera de expresar los datos aumentados es mediante una representación jerárquica, donde tenemos.

En una primera etapa:

$$y|Z \sim f(y|Z).$$

Para una segunda etapa:

$$Z \sim g(\pi).$$

En general, el paso-E consiste en calcular el valor esperado de la función de log-verosimilitud completa  $L_c(\psi|Y, Z)$ :

$$Q(\psi, \psi^i) = E(\log [L_c(\psi|Y, Z)])$$

$$Q(\psi, \psi^i) = \int_Z \log [L(\psi|Z, Y)] p(Z|\psi^i, Y) dZ. \quad (2.5)$$

Continuando con el algoritmo, el paso-M consiste en maximizar la función  $Q$  respecto a  $\psi$  para obtener  $\psi^{i+1}$ . El algoritmo se repite iterativamente hasta que  $\|\psi^{i+1} - \psi^i\|$  sea suficientemente pequeño.

No obstante, una limitación del algoritmo EM es que, en algunas funciones de verosimilitud complejas, pueden requerirse pasos iterativos de maximización. Por ello, Meng y Rubin (1993) proponen extender la función  $Q$ , reemplazando el paso-M por una secuencia de maximizaciones condicionales formuladas como  $M = \{m_s(\psi); s = 1, \dots, S\}$ . Donde  $M$  es un conjunto de  $S$  vectores de funciones de  $\psi$ . Iniciando con  $\psi^{(0)}$  en la iteración  $(t+1)$ ,  $t = 0, 1, \dots$ , el algoritmo ECM realiza el paso-E descrito en (2.5) y posteriormente  $S$  pasos-CM, donde en cada paso de maximización  $s = 1, \dots, S$ , encontramos  $\psi^{(t+s/S)}$  de modo que maximice la función  $Q(\psi|\psi^{(t)})$  sobre  $\psi \in \Psi$  sujeto a la restricción  $m_s(\psi) = m_s(\psi^{\{t+(s-1)/S\}})$ , es decir:

$$Q(\psi^{t+s/S}|\psi^{(t)}) = \operatorname{argmax}_{\psi} Q(\psi|\psi^{(t)}). \quad (2.6)$$

El resultado del paso CM se denota por  $\psi^{(t+1/S)} = (\psi_1^{(t+1/S)}, \dots, \psi_s^{(t+1/S)})$ . Además, el valor de  $\psi$  para empezar la siguiente iteración del algoritmo ECM,  $\psi^{(t+1)}$ , se define como el resultado de (2.6), es decir cuando  $\psi^{(t+S/S)} \equiv \psi^{(t+1)}$ .

## Capítulo 3

# Modelo propuesto

### 3.1. Introducción

Maehara (2018a) realiza una extensión de la distribución MENA basándose en la distribución BS. Para lo cual parte de la siguiente representación estocástica:  $Z = \mu + \sigma\kappa^{1/2}(U)Z_0$ , donde  $Z_0 \sim \text{NA}(0, 1, \lambda)$  y  $\kappa(U) > 0$  es una función que asigna pesos a los parámetros de escala con f.d.a.  $S(u, \nu)$ , siendo la función de densidad de  $Z$  la descrita en (2.4). Siguiendo la misma idea del trabajo mostrado Vilca et al. (2011) y como es mostrado en Maehara (2018a) se considera que  $\mu = 0$ , y  $\sigma^2 = 1$ .

$$Z = \frac{1}{\alpha} \left( \sqrt{\frac{T}{\beta}} - \sqrt{\frac{\beta}{T}} \right) \sim \text{MENA}(0, 1, \lambda; S).$$

Luego, siguiendo lo propuesta de Birnbaum y Saunders (1969), obtenemos que  $T$  tiene una distribución BS-MENA si puede ser representada por

$$T = \frac{\beta}{4} \left[ \alpha\kappa^{1/2}(U)Z_0 + \sqrt{\alpha^2\kappa(U)Z_0^2 + 4} \right]. \quad (3.1)$$

De acuerdo a Maehara (2018a), cuando  $U = 1$  la distribución de (3.1) es BS-NA  $(\alpha\kappa^{1/2}(u), \beta, \alpha)$ . Algunos casos particulares que el modelo propuesto extiende son los siguientes:

- Para  $S = 1$  obtenemos el modelo de MF-BS-MENA propuesto por Maehara (2018a).
- Para  $\lambda = 0$ ,  $U = 1$  obtenemos el modelo MF-BS propuesto por Benites et al. (2017).
- Para  $\lambda = 0$ , obtenemos el modelo propuesto por Benites (2019).
- Para  $S = 1$ ,  $\lambda = 0$  y  $U = 1$  obtenemos el modelo propuesto por Birnbaum y Saunders (1969).
- Para  $S = 1$ ,  $U = 1$  obtenemos el modelo propuesto por Vilca et al. (2011).
- Para  $S = 1$ ,  $\lambda = 0$  obtenemos el modelo propuesto por Balakrishnan et al. (2009).
- Para  $S = 2$ ,  $\lambda = 0$ ,  $U = 1$  obtenemos el modelo propuesto por Balakrishnan et al. (2011).
- Para  $U = 1$  obtenemos el modelo propuesto por Maehara (2018b).

Los siguientes Teoremas fueron presentados en Maehara (2018a). El Teorema 1 muestra la función de densidad de la distribución BS-MENA. El Teorema 2 y 3 será utilizado en la representación jerárquica como parte del algoritmo de tipo EM.

**Teorema 1** Sea  $T \sim \text{BS-MENA}(\alpha, \beta, \lambda; S)$ , entonces, la f.d.p. de  $T$  puede ser expresada como

$$\begin{aligned} f_T(t; \alpha, \beta, \lambda, \nu) &= \phi_{\text{MENA}}(a_t(\alpha, \beta))A_t(\alpha, \beta), \quad t > 0 \\ &= 2 \int_0^\infty \phi(a_t(\alpha, \beta); 0, \kappa(u))\Phi(\lambda\kappa^{1/2}(u)a_t(\alpha, \beta))A_t(\alpha, \beta)dS(u; \nu), \end{aligned} \quad (3.2)$$

donde  $\phi_{\text{MENA}}(\cdot)$  es la f.d.p. de una distribución  $\text{MENA}(0, 1, \lambda; S)$ .

La media de esta distribución es  $E(T) = \frac{\beta}{2} [2 + \alpha^2\mu_2 + \alpha\omega_1]$  y su varianza  $\text{Var}[T] = \frac{\beta^2}{4} [(4\alpha^2 - 2\alpha^2\omega_1)\mu_2 - \alpha^4\mu_2^2 + 2\alpha^4\mu_4 - \alpha^2\omega_1^2 + 2\alpha^3\omega_3]$ , donde  $\mu_j = E(Z^j)$ ,  $j = 2, 4$  y  $\omega_k = E(Z^k\sqrt{\alpha^2Z^2 + 4})$ ,  $k = 1, 3$ .

**Teorema 2** Sea  $T \sim \text{BS-MENA}(\alpha, \beta, \lambda, S)$ , y  $H|(U = u) \sim \text{HN}(0, \kappa(u))$ . Entonces,  $T|(H = h, U = u)$  tiene la f.d.p.

$$f_{T|H,U}(t|h, u) = \phi(\lambda_{h,u} + a_t(\alpha_{\delta,u}, \beta))A_t(\alpha_{\delta,u}, \beta),$$

donde  $\alpha_{\delta,u} = \alpha\kappa^{1/2}(u)\sqrt{1 - \delta^2}$ ,  $\lambda_{h,u} = -\delta h(\kappa^{1/2}(u)\sqrt{1 - \delta^2})$ ,  $a_t(\alpha_{\delta,u}, \beta) = \frac{1}{\alpha_{\delta,u}}\xi\left(\frac{t}{\beta}\right)$ ,  $A_t(\alpha_{\delta,u}, \beta) = \frac{\partial}{\partial t}a_t(\alpha_{\delta,u}, \beta)$  y  $\delta = \lambda/\sqrt{1 + \lambda^2}$ .

La ecuación anterior también es conocida como distribución Birbaum-Saunders extendida, y se denota por  $\text{BSE}(\alpha_{\delta,u}, \beta, \sigma = 2, \lambda_{h,u})$ .

**Teorema 3** Sea  $T \sim \text{BS-MENA}(\alpha, \beta, \lambda, S)$ , y  $H|(U = u) \sim \text{HN}(0, \kappa(u))$ . Entonces,  $H|(T = t, U = u)$  tiene f.d.p.

$$f_{H|T,U}(h|t, u) = \frac{\phi(h|\delta\kappa^{1/2}(u)a_t(\alpha_u, \beta), \kappa(u)(1 - \delta^2))}{\Phi(\lambda a_t(\alpha_u, \beta))},$$

donde  $\alpha_u = \alpha\sqrt{\kappa(u)}$  y  $\phi(\cdot|\mu, \sigma^2)$  es una distribución normal con media  $\mu$  y varianza  $\sigma^2$ . Con lo cual,

$$E[\kappa^{-1}(U)H|T = t] = \delta a_t(\alpha, \beta)\kappa + \tau\sqrt{1 - \delta^2}$$

y

$$E[\kappa^{-1}(U)H^2|T = t] = \delta^2 a_t^2(\alpha, \beta)\kappa + [1 - \delta^2] + \delta a_t(\alpha, \beta)\tau\sqrt{1 - \delta^2},$$

donde  $\kappa = E[\kappa^{-1}(U)|T = t]$ ,  $\tau = E\left[\kappa^{-1/2}(U)W_\Phi\left(\frac{\delta a_t(\alpha, \beta)}{\kappa^{1/2}(u)\sqrt{1 - \delta^2}}\right)|T = t\right]$  y  $W_\phi = \phi(u)/\Phi(u)$ , para  $u \in \mathbb{R}$ .

### 3.1.1. Casos especiales

Esta sección describe algunos casos especiales de la distribución BS-MENA descritas en Maehara (2018a), basadas en la distribución  $t$  de Student asimétrica (TA), normal contami-

nada asimétrica (NCA) y Slash asimétrica (SLA) y  $\kappa(U) = 1/U$ . Las cuales pueden obtenerse de la representación estocástica en (3.1) y del Teorema 1.

i *Distribución BS t de student asimétrica* (BS-TA)

Un caso especial de la distribución BS-MENA es la t-Student asimétrica Birbaum-Saunders (ST-BS). Para este caso se tiene que  $U \sim \text{Gamma}(\nu/2, \nu/2)$ ,  $\nu > 0$ , y la distribución estaría dada por:

$$f_T(t) = 2t (a_t(\alpha, \beta) | 0, 1; \nu) T \left( \sqrt{\frac{\nu + 1}{a_t^2(\alpha, \beta) + \nu}} B_t; \nu + 1 \right) A_t(\alpha, \beta), t > 0,$$

donde  $B_t = \lambda a_t(\alpha, \beta)$ ,  $t(\cdot | 0, 1; \nu)$  y  $T(\cdot; \nu)$  son las f.d.p. y f.d.a. de una distribución t-student. La notación para esta distribución es  $T \sim \text{BS-TA}(\alpha, \beta, \lambda, \nu)$ .

ii *Distribución BS normal contaminada asimétrica* (BS-NCA)

La distribución normal contaminada asimétrica puede ser utilizada para generar la distribución Birnbaum-Saunders normal contaminada asimétrica (BS-NCA).  $U$  es una variable aleatoria discreta con la siguiente f.d.p.:

$$h(u; \nu, \gamma) = \nu \mathbb{I}_\gamma(u) + (1 - \nu) \mathbb{I}_1(u), \quad 0 < \nu < 1, \quad 0 < \gamma < 1,$$

donde  $\mathbb{I}_A(\cdot)$  es una función indicadora de A. La función de distribución de T, a partir de (3.2) es:

$$f_T(t) = 2 \left[ \nu \phi \left( a_t(\alpha, \beta); 0, \frac{1}{\gamma} \right) \Phi \left( \gamma^{1/2} B_t \right) + (1 - \nu) \phi(a_t(\alpha, \beta)) \Phi(B_t) \right] A_t(\alpha, \beta).$$

La notación para esta distribución es  $T \sim \text{BS-NCA}(\alpha, \beta, \lambda, \nu, \gamma)$ .

iii *Distribución BS Slash asimétrica* (BS-SLA)

Para este caso se considera  $U \sim \text{Beta}(\nu, 1)$ ,  $\nu > 0$ , decimos que  $T$  tiene una distribución BS Slash asimétrica (BS-SLA) con f.d.p.:

$$f_T(t) = 2\nu \int_0^1 u^{\nu-1} \phi(a_t(\alpha, \beta); 0, 1/u) \Phi(u^{1/2} B_t) du A_t(\alpha, \beta).$$

La notación para esta distribución es  $T \sim \text{BS-SLA}(\alpha, \beta, \lambda, \nu)$ .

## 3.2. MF-BS-MENA

### 3.2.1. Descripción del modelo

Siguiendo a Balakrishnan et al. (2011), definimos la mixtura de distribuciones BS-MENA de la siguiente manera:

$$f(t; \mathbf{p}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\lambda}, \nu) = \sum_{j=1}^G p_j f_{T_j}(t; \alpha_j, \beta_j, \lambda_j, \nu_j), \quad y \in \mathbb{R}_+, \quad (3.3)$$

donde  $p_j$  es el parámetro de mezcla del  $j$ -ésimo componente restringido a  $\sum_{j=1}^G p_j = 1$ , y  $f_{T_j}(t; \alpha_j, \beta_j, \lambda_j, \nu_j)$  es la f.d.p. BS-MENA  $(\alpha_j, \beta_j, \lambda_j, \nu_j)$  del  $j$ -ésimo componente de la mezcla, con  $\alpha_j > 0$ ,  $\beta_j > 0$ ,  $\lambda_j \in \mathbb{R}$ ,  $j = 1, \dots, G$ . Además,  $\mathbf{p} = (p_1, \dots, p_G)^\top$ ,  $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_G)^\top$ ,  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_G)^\top$ ,  $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_G)^\top$ . Respecto a los parámetros de mezcla de escala  $\nu$ , se asumirá que  $\nu_1 = \nu_2 = \dots = \nu_G = \nu$  ya que ha mostrado que es más conveniente desde un punto de vista computacional, además es importante mencionar que esta condición se ha realizado en Zeller et al. (2019), Basso et al. (2010). Dicho esto, decimos que una v.a. positiva  $T$ , con función de densidad (3.3), es una mezcla finita de distribuciones BS-MENA y será denotada por:

$$T \sim \text{MF-BS-MENA}(\mathbf{p}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\lambda}, \nu).$$

Siguiendo la misma metodología de Birnbaum y Saunders (1969), podemos encontrar algunas propiedades de la distribución MF-BS-MENA:

**Teorema 4** Si  $T \sim \text{MF-BS-MENA}(\mathbf{p}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\lambda}, \nu)$ , entonces:

- (i)  $cT \sim \text{MF-BS-MENA}(\mathbf{p}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\lambda}, \nu)$ , donde  $c \in \mathbb{R}_+$ ;
- (ii)  $T^{-1} \sim \text{MF-BS-MENA}(\mathbf{p}, \boldsymbol{\alpha}, \boldsymbol{\beta}^{-1}, \boldsymbol{\lambda}, \nu)$ , donde  $\boldsymbol{\beta}^{-1} = \left(\frac{1}{\beta_1}, \dots, \frac{1}{\beta_G}\right)^\top$ ;
- (iii) Para  $\beta_1 = \beta_2 = \dots = \beta_G$ ,  $\beta/T$  y  $T/\beta$  tienen la misma distribución

### 3.2.2. Estimación de máxima verosimilitud

En esta sección, se presenta el problema de estimación para el modelo MF-BS-MENA usando el algoritmo ECM (Maximización de Esperanza Condicional). Este algoritmo busca las estimaciones de máxima verosimilitud (MV) ante la presencia de datos faltantes. Además, se presenta la metodología para la obtención de los valores iniciales, la regla de parada del algoritmo ECM y la aproximación de los errores estándar.

### 3.2.3. Estimación de parámetros mediante el algoritmo ECM

Discutiremos brevemente la implementación del algoritmo Maximización de Esperanza condicional (Expectation Conditional Mean, ECM) propuesta por Meng y Rubin (1993) para la estimación de MV de los parámetros de modelo de MF-BS-MENA. La idea básica del algoritmo ECM es que el paso de maximización (M) del algoritmo EM es reemplazado por varios pasos de maximización condicional (CM) computacionalmente simples. Por conveniencia notacional, tenemos que  $\mathbf{t} = (t_1, \dots, t_n)^\top$  el vector de datos observados,  $\mathbf{h} = (h_1, \dots, h_n)^\top$  y  $\mathbf{u} = (u_1, \dots, u_n)^\top$  son vectores de datos no observados y  $\mathbf{z} = (\mathbf{Z}_1, \dots, \mathbf{Z}_n)^\top$  el vector de componentes latentes donde  $\mathbf{Z}_i = (Z_{i1}, \dots, Z_{iG})$ ,  $i = 1, \dots, n$ . Se tendrá que  $Z_{ij} = 1$  si  $t_i$  pertenece al grupo  $j$  y  $Z_{ij} = 0$  en caso contrario, donde  $\sum_{j=1}^G Z_{ij} = 1$ . Dado los pesos  $p_1, \dots, p_G$ , los componentes de indicadores  $\mathbf{Z}_1, \dots, \mathbf{Z}_n$  son independientes, con función de densidad multinomial  $f(\mathbf{z}_1) = p_1^{z_{11}} p_2^{z_{12}} \dots p_G^{z_{1G}}$ , que es denotado por  $\mathbf{Z}_i \sim \text{Multinomial}(1, p_1 \dots p_G)$  donde  $p_j = P(Z_{ij} = 1) = 1 - P(Z_{ij} = 0)$ . Siguiendo el Teorema 2, expresamos estas distribuciones

de manera jerárquica:

$$T_i|h_i, u_i, Z_{ij} = 1 \stackrel{\text{ind}}{\sim} \text{BSE} \left( \alpha_{\delta_j, u_i}, \beta_j, \sigma = 2, \lambda_{h_j, u_i} \right) \quad (3.4)$$

$$H_i|u_i, Z_{ij} = 1 \stackrel{\text{ind}}{\sim} \text{HN} (0, \kappa(u_i)) \quad (3.5)$$

$$U_i|Z_{ij} = 1 \stackrel{\text{ind}}{\sim} \text{H} (u_i, \nu) \quad (3.6)$$

$$Z_i \stackrel{\text{iid}}{\sim} \text{Multinomial} (1, p_1, \dots, p_G), \quad i = 1, \dots, n, \quad j = 1, \dots, G, \quad (3.7)$$

donde BSE es la distribución Birbaum-Saunders Extendida propuesta por Leiva et al. (2010) y tiene la siguiente forma:

$$f_{BSE}(t|h, u) = \phi(\lambda_{h,u} + a_t(\alpha_{\delta,u}, \beta)) A_t(\alpha_{\delta,u}, \beta),$$

donde  $\alpha_{\delta,u} = \alpha \kappa^{1/2}(u) \sqrt{1 - \delta^2}$ ,  $a_t(\alpha_{\delta,u}, \beta) = \frac{1}{\alpha_{\delta,u}} \xi\left(\frac{t}{\beta}\right)$ ,  $A_t(\alpha, \beta) = \frac{\partial}{\partial t} a_t(\alpha_{\delta,u}, \beta)$ ,  $\lambda_{h,u} = -\delta h / \kappa^{1/2}(u) \sqrt{1 - \delta^2}$  y  $\delta = \lambda / \sqrt{1 + \lambda^2}$ .

Resolviendo para  $A_t(\alpha, \beta)$ :

$$\begin{aligned} A_t(\alpha_{\delta,u}, \beta) &= \frac{\partial}{\partial t} a_t(\alpha_{\delta,u}, \beta) \\ &= \frac{1}{\alpha_{\delta,u}} \left[ \frac{1}{2} \left( \frac{t}{\beta} \right)^{1/2-1} \frac{1}{\beta} \right] - \left[ \frac{1}{2} \left( \frac{t}{\beta} \right)^{1/2-1} \frac{\beta}{t^2} (-1) \right] \\ &= \frac{1}{2\alpha_{\delta,u}\beta} \left[ \left( \frac{t}{\beta} \right)^{-1/2} + \left( \frac{t}{\beta} \right)^{-3/2} \right]. \end{aligned}$$

Resolviendo para BSE:

$$f_{BSE}(t|h, u) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} (\lambda_{h,u} + a_t(\alpha_{\delta,u}, \beta))^2\right) A_t(\alpha_{\delta,u}, \beta)$$

en la que tendremos que

$$\begin{aligned} (\lambda_{h,u} + a_t(\alpha_{\delta,u}, \beta))^2 &= \lambda_{h,u}^2 + 2\lambda_{h,u}a_t(\alpha_{\delta,u}, \beta) + a_t(\alpha_{\delta,u}, \beta)^2 \\ &= \left( \frac{-\delta h}{\kappa^{1/2}(u)\sqrt{1 - \delta^2}} \right)^2 \\ &\quad + 2 \left( \frac{-\delta h}{\kappa^{1/2}(u)\sqrt{1 - \delta^2}} \right) \frac{1}{\alpha \kappa^{1/2}(u)\sqrt{1 - \delta^2}} \xi\left(\frac{t}{\beta}\right) \\ &\quad + \left( \frac{1}{\alpha \kappa^{1/2}(u)\sqrt{1 - \delta^2}} \xi\left(\frac{t}{\beta}\right) \right)^2. \end{aligned}$$

Factorizando  $1/(\kappa^{1/2}(u)\sqrt{1 - \delta^2})$  tenemos:

$$\begin{aligned} (\lambda_{h,u} + a_t(\alpha_{\delta,u}, \beta))^2 &= \frac{1}{\kappa(u)(1 - \delta^2)} \left( \frac{1}{\alpha} \xi\left(\frac{t}{\beta}\right) - \delta h \right)^2 \\ &= \frac{1}{\kappa(u)(1 - \delta^2)} (a_t(\alpha, \beta) - \delta h)^2. \end{aligned}$$

Remplazando en BSE:

$$f_{\text{BSE}} = \frac{1}{\sqrt{2\pi}} \exp \left( -\frac{1}{2} \left( \frac{1}{\kappa(u)(1-\delta^2)} (a_t(\alpha, \beta) - \delta h)^2 \right) \right) A_t(\alpha_{\delta, u}, \beta).$$

De acuerdo con (3.4)-(3.7) la función de log-verosimilitud completa de  $\boldsymbol{\theta} = (p_1, \dots, p_G, \alpha_1, \dots, \alpha_G, \beta_1, \dots, \beta_G, \delta_1, \dots, \delta_G, \nu)^\top$  dado  $(\mathbf{t}, \mathbf{h}, \mathbf{u}, \mathbf{Z})$  puede ser expresada como:

$$\begin{aligned} \ell(\boldsymbol{\theta}|\mathbf{t}, \mathbf{h}, \mathbf{u}, \mathbf{Z}) = c + \sum_{i=1}^n \sum_{j=1}^G z_{ij} \left\{ \log(p_j) - \frac{(a_{t_i}(\alpha_j, \beta_j) - \delta_j h)^2}{2\kappa(u_i)(1-\delta_j^2)} + \right. \\ \left. \log(A_{t_i}(\alpha_{\delta_j, u_i}, \beta)) + \right. \\ \left. \log(\phi_{\text{HN}}(0, \kappa(u_i))) + \log(\text{H}(u_i, \nu)) \right\}, \end{aligned} \quad (3.8)$$

donde  $c$  es una constante que no depende de  $\boldsymbol{\theta}$ .

Tomando esperanza condicional de (3.8), podemos expresar la esperanza de la función de log-verosimilitud completa como  $Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}) = E \left[ \ell(\boldsymbol{\theta}|\mathbf{t}, \mathbf{h}, \mathbf{u}, \mathbf{Z})|\mathbf{t}, \hat{\boldsymbol{\theta}} \right]$ , donde

$$\begin{aligned} Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}) = \sum_{i=1}^n \sum_{j=1}^G \left\{ \log(p_j) - \frac{\hat{\kappa}\xi^2(t_i/\beta_j)}{2\alpha_j^2(1-\delta_j^2)} + \frac{\delta_j \hat{h}_{ij}\xi(t_i/\beta_j)}{\alpha_j(1-\delta_j^2)} - \frac{\delta_j^2 \hat{h}_{ij}^2}{2(1-\delta_j^2)} - \hat{z}_{ij} \log(\alpha_j) - \right. \\ \left. \hat{z}_{ij} \log \left( E \left[ \kappa^{1/2}(U) | \hat{\boldsymbol{\theta}} \right] \right) - \hat{z}_{ij} \log(1-\delta_j^2)/2 + \hat{z}_{ij} \log \left( \frac{t_i + \beta_j}{\sqrt{\beta_j}} \right) + \right. \\ \left. E \left[ z_{ij} \log(\phi_{\text{HN}}(0, \kappa(u_i))) | \hat{\boldsymbol{\theta}} \right] + E \left[ z_{ij} \log(\text{H}(u_i, \nu)) | \hat{\boldsymbol{\theta}} \right] \right\}, \end{aligned}$$

$$\text{donde: } \hat{z}_{ij} = \frac{\hat{p}_j^{(k)} f_j(t_i | \hat{\boldsymbol{\theta}}^{(k)})}{\sum_{j=1}^G \hat{p}_j^{(k)} f_j(t_i | \hat{\boldsymbol{\theta}}^{(k)})}, \quad \hat{\kappa}_{ij} = E \left[ z_{ij} \kappa^{-1}(U_i) | \hat{\boldsymbol{\theta}}^{(k)}, t_i \right],$$

$$\hat{h}_{ij} = E \left[ z_{ij} \kappa^{-1}(U_i) H_i | \hat{\boldsymbol{\theta}}^{(k)}, t_i \right], \quad \hat{h}_{ij}^2 = E \left[ z_{ij} \kappa^{-1}(U_i) H_i^2 | \hat{\boldsymbol{\theta}}^{(k)}, t_i \right], \quad i = 1, \dots, n.$$

De acuerdo al Teorema 3:

$$\begin{aligned} \hat{h}_{ij} &= \hat{\delta}_j a_t(\hat{\alpha}_j, \hat{\beta}_j) \hat{\kappa}_{ij} + \hat{\tau}_{ij} \sqrt{1 - \hat{\delta}_j^2} \\ \hat{h}_{ij}^2 &= \hat{\delta}_j^2 a_t^2(\hat{\alpha}_j, \hat{\beta}_j) \hat{\kappa}_{ij} + [1 - \hat{\delta}_j^2] + \hat{\delta}_j a_t(\hat{\alpha}_j, \hat{\beta}_j) \hat{\tau}_{ij} \sqrt{1 - \hat{\delta}_j^2}. \end{aligned}$$

### 3.2.4. Algoritmo EM para el MF-BS-MENA

El algoritmo EM se expresa la siguiente manera.

**Paso-E.** Dado  $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}^{(k)}$ , calcular  $\hat{\kappa}_{ij}$ ,  $\hat{h}_{ij}$ , y  $\hat{h}_{ij}^2$ , para  $i = 1, \dots, n$ .

Maximización:

$$\begin{aligned} \frac{\partial Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial \alpha_j} &= \sum_{i=1}^n \left\{ \frac{\widehat{\kappa}_{ij} \xi^2(t_i, \beta_j)}{\alpha_j^3 (1 - \delta_j^2)} - \frac{\delta_j \widehat{h}_{ij} \xi(t_i, \beta_j)}{\alpha_j^2 (1 - \delta_j^2)} - \frac{\widehat{z}_{ij}}{\alpha_j} \right\} = 0, \\ &= \sum_{i=1}^n \left\{ -\frac{(-2\delta_j) \widehat{\kappa}_{ij} \xi^2(t_i/\beta_j)}{2\alpha_j^2 (1 - \delta_j^2)} - \frac{\widehat{h}_{ij}^2}{2} \left( \frac{2\delta_j (1 - \delta_j^2) - \delta_j^2 (-2\delta_j)}{(1 - \delta_j^2)^2} \right) + \right. \\ &\quad \left. \frac{\widehat{h}_{ij} \xi(t_i/\beta_j)}{\alpha_j} \left( \frac{(1 - \delta_j^2) - \delta_j (-2\delta_j)}{(1 - \delta_j^2)^2} \right) - \frac{2\delta_j}{(1 - \delta_j^2)} \right\} = 0, \end{aligned}$$

$$\begin{aligned} \frac{\partial Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})}{\partial \delta_j} &= \sum_{i=1}^n \left\{ \frac{\delta_j \widehat{\kappa}_{ij} \xi^2(t_i/\beta_j)}{\alpha_j^2 (1 - \delta_j^2)} - \frac{\delta_j \widehat{h}_{ij}^2}{(1 - \delta_j^2)} - \frac{\delta_j^3 \widehat{h}_{ij}^2}{(1 - \delta_j^2)^2} + \frac{\widehat{h}_{ij} \xi(t_i/\beta_j)}{\alpha_j (1 - \delta_j^2)} + \right. \\ &\quad \left. \frac{2\delta_j^2 \widehat{h}_{ij} \xi(t_i/\beta_j)}{(1 - \delta_j^2)^2} - \frac{2\delta_j}{(1 - \delta_j^2)} \right\} = 0. \end{aligned}$$

**Paso 1-CM.** Fijar  $\widehat{\boldsymbol{\beta}}^{(k)}$  y actualizar  $\widehat{\alpha}_j^{(k)}$ ,  $\widehat{\delta}_j^{(k)}$  y  $\widehat{p}_j^{(k)}$ .

$$\widehat{\alpha}_j^{(k+1)} = \frac{\sum_{i=1}^n \widehat{\kappa}_{ij} \xi^2(t_i/\beta_j^{(k)})}{\sum_{i=1}^n \widehat{z}_{ij}} + \left[ 1 - \frac{\sum_{i=1}^n \widehat{h}_{ij}^2(t_i/\beta_j^{(k)})}{\sum_{i=1}^n \widehat{z}_{ij}} \right] \left[ \frac{\sum_{i=1}^n \widehat{h}_{ij}^{(k)} \xi(t_i/\beta_j^{(k)})}{\sum_{i=1}^n \widehat{h}_{ij}^2(t_i/\beta_j^{(k)})} \right]^2,$$

$$\widehat{\delta}_j^{(k+1)} = \frac{1}{\widehat{\alpha}_j} \frac{\sum_{i=1}^n \widehat{h}_{ij}^{(k)} \xi(t_i/\beta_j^{(k)})}{\sum_{i=1}^n \widehat{h}_{ij}^2(t_i/\beta_j^{(k)})},$$

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

**Paso 2-CM.** Fijar  $\widehat{\alpha}_j^{(k)}$ ,  $\widehat{\delta}_j^{(k)}$ ,  $\widehat{p}_j^{(k)}$  y actualizar  $\widehat{\boldsymbol{\beta}}^{(k)}$ .

$$\widehat{\boldsymbol{\beta}}^{(k+1)} = \underset{\boldsymbol{\beta}}{\operatorname{argmax}} Q\left(\widehat{\alpha}^{(k+1)}, \boldsymbol{\beta}, \widehat{\delta}_j^{(k+1)} \mid \widehat{\boldsymbol{\theta}}^{(k)}\right).$$

Actualizar  $\widehat{\nu}^{(k)}$  maximizando la función de log-verosimilitud, obteniendo.

$$\widehat{\nu}^{(k+1)} = \underset{\nu}{\operatorname{argmax}} \sum_{i=1}^n \log \left( \sum_{j=1}^G \widehat{p}_j^{(k+1)} f_{T_j} \left( t; \widehat{\alpha}_j^{(k+1)}, \widehat{\beta}_j^{(k+1)}, \widehat{\lambda}_j^{(k+1)}, \nu \right) \right).$$

Las iteraciones se repiten hasta lograr la convergencia del algoritmo ECM de acuerdo al siguiente criterio de parada

$$\|l^{(k+1)} - l^{(k)}\| < 10^{-6},$$

donde  $l^{(k)}$  es la log-verosimilitud observada en  $\hat{\boldsymbol{\theta}}^{(k)}$ .

### 3.2.5. Especificación de los valores iniciales

Es bien conocido que los modelos de mixtura puede proveer una función de log-verosimilitud multimodal. Por lo que el método de estimación de MV a través del algoritmo ECM puede no tener una solución máxima global (es decir múltiples máximos locales) si los valores iniciales están distantes a los valores reales de los parámetros. Para garantizar que se identifiquen las estimaciones verdaderas de MV, recomendamos ejecutar el algoritmo ECM utilizando un rango de valores iniciales diferentes. Los valores iniciales del algoritmo ECM propuesto fueron obtenidos siguiendo los siguientes pasos:

1. Particionar la muestra en  $G$  grupos usando el algoritmo de clusterización  $k$ -means o  $k$ -medoids.
2. Calcular la proporción de puntos de datos que pertenecen al mismo grupo  $j$ , osea  $p_j^{(0)}$ ,  $j = 1, \dots, G$ . Esto es el valor inicial de  $p_j$ .
3. Para cada componente calculamos los valores iniciales de  $\alpha_j^{(0)}$  y  $\beta_j^{(0)}$  utilizando el método de momentos modificado *mmeth* propuesto por Ng, Kundu y Balakrishnan (2003) e implementado en el paquete *bssn* (Maehara y Benites, 2015). Para el parámetro  $\lambda_j^{(0)}$  se utilizó usando el coeficiente de correlación de Fisher implementado en el paquete *sn*.
4. Para la estimación de los valores iniciales de  $\nu^{(0)}$  calculamos la log-verosimilitud perfilada  $\ell(\hat{\boldsymbol{\theta}}_1^{(0)}(\nu), \dots, \hat{\boldsymbol{\theta}}_G^{(0)}(\nu))$ , ver (3.8), para la red de valores de  $\nu$ . Así, para las distribuciones MF-BS-TA (mixtura finita de una distribución BS t-student asimétrica), MF-BS-SLA (mixtura finita de una distribución BS Slash asimétrica) se utilizaron números enteros entre 1 y 20, mientras que para la distribución MF-BS-NCA (mixtura finita de una distribución BS Normal contaminada asimétrica) se utilizaron números entre 0.1 y 0.9. Para cada caso, se elije  $\nu$  que maximice la log-verosimilitud perfilada.

### 3.2.6. Matriz de información observada

Una manera de aproximarse a la matriz de covarianzas de las estimaciones de máxima verosimilitud  $\hat{\boldsymbol{\theta}}$  es a través de la inversa de la esperanza de la matriz de información. No obstante, esta metodología puede ser computacionalmente costosa. Por ello, Basford, Greenway, McLachlan y Peel (1993) proponen una aproximación a la matriz de covarianzas de máxima verosimilitud de la siguiente forma.

$$\mathbf{J}_o(\hat{\boldsymbol{\theta}}|t) = \sum_{i=1}^n \hat{\mathbf{s}}_i \hat{\mathbf{s}}_i^\top,$$

donde  $\hat{\mathbf{s}}_i = \partial \log f(t_i; \boldsymbol{\theta}) / \partial \boldsymbol{\theta} |_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$  es el vector gradiente de la función de verosimilitud evaluado en  $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$  basado en la  $i$ -ésima observación. Cabe mencionar que para la obtención de los

errores estándar de los parámetros, obtendremos el error estándar del parámetro  $\lambda_j$  y no del parámetro  $\delta_j = \lambda_j / \sqrt{1 + \lambda_j^2}$  pues resulta algebraicamente más simple. Considerando a  $\widehat{\mathbf{s}}_i$  como un vector particionado en todos los componentes de  $\boldsymbol{\theta}$  dado por

$$\widehat{\mathbf{s}}_i = (\widehat{s}_{i,p_1}, \dots, \widehat{s}_{i,p_{G-1}}, \widehat{s}_{i,\alpha_1}, \dots, \widehat{s}_{i,\alpha_G}, \widehat{s}_{i,\beta_1}, \dots, \widehat{s}_{i,\beta_G}, \widehat{s}_{i,\lambda_1}, \dots, \widehat{s}_{i,\lambda_G}),$$

donde los elementos de  $\widehat{\mathbf{s}}_i$  están dados por:

$$\begin{aligned} \widehat{s}_{i,p_j} &= \frac{f_{T_j}(t_i, \alpha_j, \beta_j, \lambda_j) - f_{T_G}(t_i, \alpha_G, \beta_G, \lambda_G)}{f(t_i, \mathbf{p}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\lambda})}, & \widehat{s}_{i,\alpha_j} &= \frac{p_j D_{\alpha_j}(f_{T_j}(t_i, \alpha_j, \beta_j, \lambda_j))}{f(t_i, \mathbf{p}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\lambda})}, \\ \widehat{s}_{i,\beta_j} &= \frac{p_j D_{\beta_j}(f_{T_j}(t_i, \alpha_j, \beta_j, \lambda_j))}{f(t_i, \mathbf{p}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\lambda})}, & \widehat{s}_{i,\lambda_j} &= \frac{p_j D_{\lambda_j}(f_{T_j}(t_i, \alpha_j, \beta_j, \lambda_j))}{f(t_i, \mathbf{p}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\lambda})}, \quad j = 1, \dots, G, \end{aligned}$$

donde

$$\begin{aligned} D_{\alpha_j}(f_{T_j}(t_i, \alpha_j, \beta_j, \lambda_j)) &= \frac{\partial}{\partial \alpha_j} f_{T_j}(t_i, \alpha_j, \beta_j, \lambda_j), \\ D_{\beta_j}(f_{T_j}(t_i, \alpha_j, \beta_j, \lambda_j)) &= \frac{\partial}{\partial \beta_j} f_{T_j}(t_i, \alpha_j, \beta_j, \lambda_j) \\ \text{y } D_{\lambda_j}(f_{T_j}(t_i, \alpha_j, \beta_j, \lambda_j)) &= \frac{\partial}{\partial \lambda_j} f_{T_j}(t_i, \alpha_j, \beta_j, \lambda_j). \end{aligned}$$

Definiendo

$$I_i^\Phi = \int_0^\infty \kappa^{-w(u)} \exp\left\{-\frac{1}{2}\kappa^{-1}(u)a_{t_i}^2(\boldsymbol{\alpha}, \boldsymbol{\beta})\right\} \Phi\left(\kappa^{-1/2}(u)B_{t_i}\right) dH(u)$$

y

$$I_i^\phi = \int_0^\infty \kappa^{-w(u)} \exp\left\{-\frac{1}{2}\kappa^{-1}(u)a_{t_i}^2(\boldsymbol{\alpha}, \boldsymbol{\beta})\right\} \phi\left(\kappa^{-1/2}(u)B_{t_i}\right) dH(u),$$

donde  $B_{t_i} = \boldsymbol{\lambda}a_{t_i}(\boldsymbol{\alpha}, \boldsymbol{\beta})$ ,  $i = 1, \dots, n$ , se puede obtener

$$\begin{aligned} D_{\alpha_j}(f_{T_j}(t_i, \alpha_j, \beta_j, \lambda_j)) &= \frac{2}{\sqrt{2\pi}} \left[ \frac{\partial A_{t_i}}{\partial \boldsymbol{\alpha}} I_i^\Phi(1/2) - a_{t_i} A_{t_i} \frac{\partial a_{t_i}}{\partial \boldsymbol{\alpha}} I_i^\Phi(3/2) + \boldsymbol{\lambda} A_{t_i} \frac{\partial a_{t_i}}{\partial \boldsymbol{\alpha}} I_i^\phi(1) \right], \\ D_{\beta_j}(f_{T_j}(t_i, \alpha_j, \beta_j, \lambda_j)) &= \frac{2}{\sqrt{2\pi}} \left[ \frac{\partial A_{t_i}}{\partial \boldsymbol{\beta}} I_i^\Phi(1/2) - a_{t_i} A_{t_i} \frac{\partial a_{t_i}}{\partial \boldsymbol{\beta}} I_i^\Phi(3/2) + \boldsymbol{\lambda} A_{t_i} \frac{\partial a_{t_i}}{\partial \boldsymbol{\beta}} I_i^\phi(1) \right], \\ D_{\lambda_j}(f_{T_j}(t_i, \alpha_j, \beta_j, \lambda_j)) &= \frac{2}{\sqrt{2\pi}} a_{t_i} A_{t_i} I_i^\phi(1). \end{aligned}$$

Calcular los errores estándar de  $\nu$  usando la aproximación presentada en esta sección es analíticamente compleja, debido a ello varios trabajos, vea por ejemplo Zeller et al. (2019), Basso et al. (2010), no han considerado la obtención del error estándar por el método de aproximación.

## Capítulo 4

### Estudio de Simulación

El objetivo de esta sección es mostrar la precisión de las estimaciones de  $\boldsymbol{\theta} = (\mathbf{p}^\top, \boldsymbol{\alpha}^\top, \boldsymbol{\beta}^\top, \boldsymbol{\lambda}^\top)$  con  $\mathbf{p} = (p_1, p_2)^\top$ ,  $\boldsymbol{\alpha} = (\alpha_1, \alpha_2)^\top$ ,  $\boldsymbol{\beta} = (\beta_1, \beta_2)^\top$  y  $\boldsymbol{\lambda} = (\lambda_1, \lambda_2)^\top$  mediante el algoritmo ECM propuesto en el Capítulo 3. Los resultados promedio para 1000 réplicas y tamaños de muestra  $n = 200, 300, 500, 800$  y  $1000$  para las distribuciones MF-BS-TA, MF-BS-SLA y MF-BS-NCA. Las simulaciones que serán consideradas en esta sección serán solamente para  $G = 2$  componentes ya que al considerar el algoritmo de  $k$ -medias como método de particionamiento para generar los valores iniciales, en cada ejecución del algoritmo se obtendrá valores iniciales distintos y permutados, lo cual dificulta en gran medida realizar dichas simulaciones, sin embargo esto es algo que se ha venido realizando en diferentes trabajos tales como Zeller et al. (2019), Basso et al. (2010). Para todos estos escenarios, se calculó la media de los parámetros estimados, la aproximación de los errores estándar (IM-SE), las desviaciones estándar de Monte Carlo (MC Sd) y el porcentaje de veces que el intervalo de confianza (CI) cubre el verdadero valor de los parámetros (COB MC). Los casos especiales analizados tienen las siguientes consideraciones:

- El Cuadro 4.1 muestra las simulaciones realizadas para la distribución MF-BS-TA con parámetros  $\boldsymbol{\alpha} = (0.25, 0.50)^\top$ ,  $\boldsymbol{\beta} = (5.0, 1.0)^\top$ ,  $\boldsymbol{\lambda} = (1.0, 3.0)^\top$ ,  $\mathbf{p} = (0.3, 0.7)^\top$  y  $\nu = 4$ . Se puede observar que a medida que aumenta el tamaño de muestra las estimaciones se aproximan a los verdaderos parámetros, principalmente para el parámetro de asimetría  $\boldsymbol{\lambda}$  y el parámetro  $\nu$ . Los parámetros  $\boldsymbol{\alpha}$ ,  $\boldsymbol{\beta}$  y  $\mathbf{p}$  muestran estimaciones bastante precisas incluso para el tamaño de muestra  $n=200$ .
- El Cuadro 4.2 muestra las simulaciones realizadas para la distribución MF-BS-NCA con parámetros  $\boldsymbol{\alpha} = (0.25, 0.50)^\top$ ,  $\boldsymbol{\beta} = (5.0, 1.0)^\top$ ,  $\boldsymbol{\lambda} = (1.0, 3.0)^\top$ ,  $\mathbf{p} = (0.3, 0.7)^\top$ ,  $\nu = 0.1$  y  $\gamma = 0.1$ . Al igual que la distribución anterior, conforme se incrementa el tamaño de muestra las estimaciones se aproximan a los verdaderos parámetros. Además, se observa que las estimaciones son muy aproximadas a los verdaderos parámetros incluso para un tamaño de muestra  $n=200$ .
- El Cuadro 4.3 muestra las simulaciones de la distribución MF-BS-SLA con parámetros  $\boldsymbol{\alpha} = (0.10, 0.50)^\top$ ,  $\boldsymbol{\beta} = (5.0, 1.0)^\top$ ,  $\boldsymbol{\lambda} = (-1.0, 3.0)^\top$ ,  $\mathbf{p} = (0.3, 0.7)^\top$  y  $\nu = 4$ . Al igual que las dos distribuciones anteriores, conforme se incrementa el tamaño de muestra las estimaciones se aproximan a los verdaderos parámetros.

Los cuadros muestran los errores estándar de la matriz de información (IM SE), las desviaciones estándar de Monte Carlo (MC Sd) y el porcentaje de cobertura como resultado de los intervalos de confianza al 95 %, asumiendo normalidad asintótica. Se observa que los IM SE son bastante próximos a las desviaciones obtenidas por Monte Carlo, lo cual indica que la aproximación a los errores estándar es confiable. La cobertura (COB) se define como  $\text{COB}(\hat{\theta}) = (1/m) \sum_{j=1}^m I(\theta \in [\hat{\theta}_{L_j}, \hat{\theta}_{U_j}])$ , donde  $I$  es una función indicadora que toma valor igual a 1 si  $\theta$  cae en el intervalo  $[\hat{\theta}_{L_j}, \hat{\theta}_{U_j}]$ ,  $\hat{\theta}_{L_j}$  y  $\hat{\theta}_{U_j}$  son los  $j$ -ésimos valores estimados de los límites inferiores y superiores del IC respectivamente.

Cuadro 4.1: Estimación de parámetros de la distribución MF-BS  $t$  de student asimétrica (MF-BS-TA), para 1000 réplicas de diferentes tamaño de muestra sin considerar valores iniciales. Con  $\alpha = (0.25, 0.50)^\top$ ,  $\beta = (5.0, 1.0)^\top$ ,  $\lambda = (1.0, 3.0)^\top$ ,  $\mathbf{p} = (0.3, 0.7)^\top$  y  $\nu = 4$

n	Medida	$\alpha_1$	$\alpha_2$	$\beta_1$	$\beta_2$	$\lambda_1$	$\lambda_2$	$p_1$	$\nu$
1000	Media	0.2532	0.5000	4.9927	1.0007	1.0357	3.0128	0.3012	4.2663
	IM SE	0.0320	0.0402	0.2821	0.0247	0.6113	0.5503	0.0194	
	MC Sd	0.0274	0.0199	0.1438	0.0129	0.2801	0.2555	0.0180	0.9312
	COB MC	95.7 %	99.4 %	97.8 %	99.5 %	99.5 %	99.7 %	96.8 %	
800	Media	0.2538	0.4991	4.9929	1.0015	1.0563	3.0060	0.3010	4.3297
	IM SE	0.0356	0.0449	0.3150	0.0277	0.6944	0.6162	0.0217	
	MC Sd	0.0328	0.0223	0.1789	0.0159	0.3781	0.3057	0.0205	1.0939
	COB MC	93.8 %	99.6 %	96.9 %	98.9 %	99.7 %	99.9 %	95.5 %	
500	Media	0.2548	0.5008	5.0034	1.0004	1.0814	3.0491	0.3008	4.4845
	IM SE	0.0452	0.0571	0.4071	0.0351	0.9139	0.8006	0.0276	
	MC Sd	0.0417	0.0334	0.2494	0.0220	0.5772	0.4391	0.0259	1.3775
	COB MC	92.8 %	98.7 %	96.0 %	99.4 %	99.5 %	99.0 %	95.9 %	
300	Media	0.2557	0.4975	5.0075	1.0011	1.1236	3.0682	0.3038	4.5527
	IM SE	0.0579	0.0739	0.5299	0.0460	1.2181	1.0664	0.0357	
	MC Sd	0.0495	0.0424	0.3126	0.0274	0.7806	0.6141	0.0314	1.5496
	COB MC	94.6 %	98.1 %	94.4 %	99.4 %	99.3 %	99.1 %	96.7 %	
200	Media	0.2507	0.4930	5.0423	1.0042	1.1801	3.1053	0.3000	4.4446
	IM SE	0.0684	0.0877	0.6264	0.0562	1.6308	1.3228	0.0430	
	MC Sd	0.0597	0.0523	0.4184	0.0379	1.2408	0.7436	0.0403	1.7133
	COB MC	92.8 %	97.3 %	93.2 %	99.0 %	98.6 %	98.5 %	95.2 %	

Cuadro 4.2: Estimación de parámetros de la distribución MF-BS normal contaminada asimétrica (MF-BS-NCA), para 1000 réplicas de diferentes tamaño de muestra sin considerar valores iniciales. Con  $\alpha = (0.25, 0.50)^\top$ ,  $\beta = (5.0, 1.0)^\top$ ,  $\lambda = (1.0, 3.0)^\top$ ,  $\mathbf{p} = (0.3, 0.7)^\top$ ,  $\nu = 0, 1$  y  $\gamma = 0, 1$

n	Medida	$\alpha_1$	$\alpha_2$	$\beta_1$	$\beta_2$	$\lambda_1$	$\lambda_2$	$p_1$	$\nu$	$\gamma$
1000	Media	0.2520	0.4942	4.9952	1.0033	1.0963	2.9662	0.3042	0.1076	0.1026
	IM SE	0.0351	0.0301	0.2924	0.0235	0.6152	0.4879	0.0166		
	MC Sd	0.0292	0.0274	0.1603	0.0168	0.4254	0.2531	0.0196	0.0482	0.0322
	COB MC	96.0 %	92.0 %	98.0 %	98.0 %	98.0 %	98.0 %	88.0 %		
800	Media	0.2505	0.4923	5.0197	1.0025	1.0767	3.0101	0.3003	0.1126	0.1079
	IM SE	0.0390	0.0332	0.3389	0.0261	0.6954	0.5545	0.0184		
	MC Sd	0.0327	0.0331	0.2555	0.0212	0.5235	0.4222	0.0196	0.0551	0.0416
	COB MC	92.8 %	90.9 %	94.1 %	97.8 %	97.9 %	97.1 %	93.6 %		
500	Media	0.2505	0.4884	5.0359	1.0040	1.1267	3.0139	0.3000	0.1221	0.1139
	IM SE	0.0494	0.0418	0.4399	0.0333	0.9260	0.7099	0.0233		
	MC Sd	0.0418	0.0437	0.3367	0.0281	0.7944	0.5593	0.0254	0.0724	0.0558
	COB MC	91.6 %	88.8 %	91.6 %	97.2 %	97.7 %	96.8 %	92.3 %		
300	Media	0.2467	0.4810	5.0989	1.0067	1.1698	3.0945	0.3008	0.1377	0.1161
	IM SE	0.0601	0.0529	0.5744	0.0431	1.2731	0.9512	0.0300		
	MC Sd	0.0507	0.0581	0.4600	0.0390	1.2347	0.8579	0.0331	0.0938	0.0706
	COB MC	89.9 %	87.1 %	89.5 %	96.5 %	97.3 %	95.9 %	91.2 %		
200	Media	0.2429	0.4754	5.1298	1.0114	1.2546	3.1523	0.3001	0.1479	0.1165
	IM SE	0.0738	0.0650	0.7405	0.0555	1.7847	1.2192	0.0367		
	MC Sd	0.0590	0.0711	0.5343	0.0570	1.6862	1.1212	0.0401	0.1120	0.0796
	COB MC	90.6 %	85.7 %	87.6 %	95.5 %	97.7 %	94.6 %	92.0 %		

Cuadro 4.3: Estimación de parámetros de la distribución MF-BS slash asimétrica (MF-BS-SLA), para 1000 réplicas de diferentes tamaño de muestra sin considerar valores iniciales. Con  $\alpha = (0.10, 0.50)^\top$ ,  $\beta = (5.0, 1.0)^\top$ ,  $\lambda = (-1.0, 3.0)^\top$ ,  $\mathbf{p} = (0.3, 0.7)^\top$  y  $\nu = 4$

n	Medida	$\alpha_1$	$\alpha_2$	$\beta_1$	$\beta_2$	$\lambda_1$	$\lambda_2$	$p_1$	$\nu$
1000	Media	0.0993	0.4958	4.996	1.0009	-1.0094	3.0276	0.2997	4.1986
	IM SE	0.0179	0.0290	0.1637	0.0251	0.6771	0.4892	0.0156	
	MC Sd	0.0097	0.0246	0.0632	0.0164	0.2723	0.3086	0.0150	1.4604
	COB MC	99.5 %	93.8 %	99.2 %	99.1 %	99.9 %	98.4 %	96.2 %	
800	Media	0.1003	0.4956	4.9994	1.0012	-1.0463	3.0159	0.2994	4.3174
	IM SE	0.0199	0.0327	0.1816	0.0284	0.7606	0.5507	0.0175	
	MC Sd	0.0112	0.0262	0.0709	0.0189	0.3342	0.3581	0.0175	1.6068
	COB MC	98.7 %	93.7 %	98.2 %	98.8 %	99.3 %	98.7 %	95.7 %	
500	Media	0.0994	0.4886	4.9983	1.0024	-1.0627	3.0144	0.2998	4.1617
	IM SE	0.0247	0.0402	0.2385	0.0357	0.9955	0.6929	0.0220	
	MC Sd	0.0139	0.0346	0.0913	0.0254	0.4526	0.465	0.0218	1.7062
	COB MC	98.4 %	90.5 %	97.1 %	98.0 %	99.0 %	97.3 %	94.4 %	
300	Media	0.0988	0.4872	4.9943	1.0013	-1.1181	3.0861	0.2997	4.1481
	IM SE	0.0307	0.0514	0.3167	0.0457	1.3618	0.9436	0.0284	
	MC Sd	0.0168	0.0388	0.1136	0.0304	0.6475	0.5944	0.0289	1.7999
	COB MC	97.0 %	90.1 %	96.8 %	98.4 %	99.4 %	99.0 %	93.3 %	
200	Media	0.0987	0.4866	4.9975	1.0012	-1.223	3.1517	0.2994	4.2105
	IM SE	0.0379	0.0628	0.4089	0.0560	1.9118	1.2005	0.0350	
	MC Sd	0.0202	0.0497	0.1337	0.0398	1.2227	0.772	0.0339	1.9762
	COB MC	97.4 %	89.6 %	95.8 %	97.7 %	99.5 %	97.3 %	95.3 %	

Las Figuras 4.1, 4.2 y 4.3 muestran los diagramas de cajas para las distribuciones MF-BS-TA, MF-BS-SLA y MF-BS-NCA respectivamente, para los distintos tamaños de muestra simulados y cada uno de los parámetros estimados. Se puede observar una mejora en la precisión de las estimaciones a medida que aumenta el tamaño de muestra. En la Figura 4.1

notamos que el parámetro de asimetría  $\lambda$  del primer componente presenta poca variabilidad a diferencia de los demás parámetros. Similarmente, el parámetro  $\beta$  muestra gráficamente mayor presencia de valores atípicos para el primer componente, principalmente en la MF-BS-NCA como muestra la Figura 4.3.

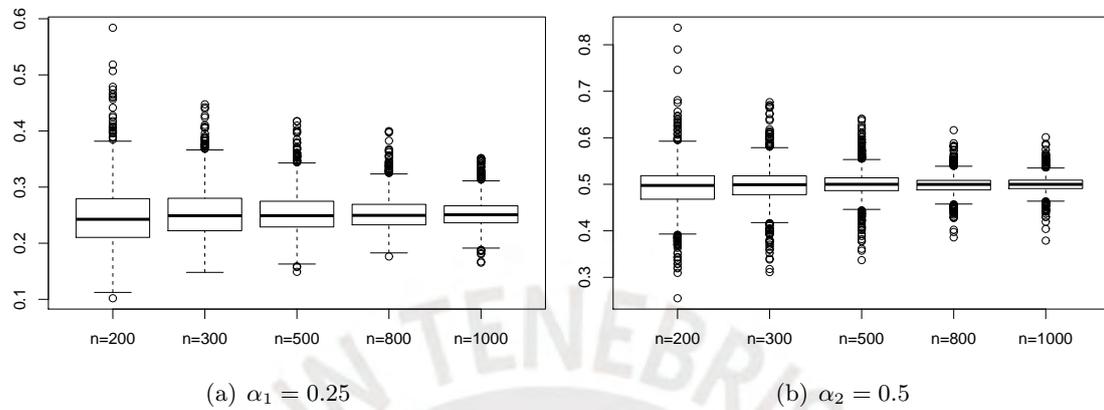
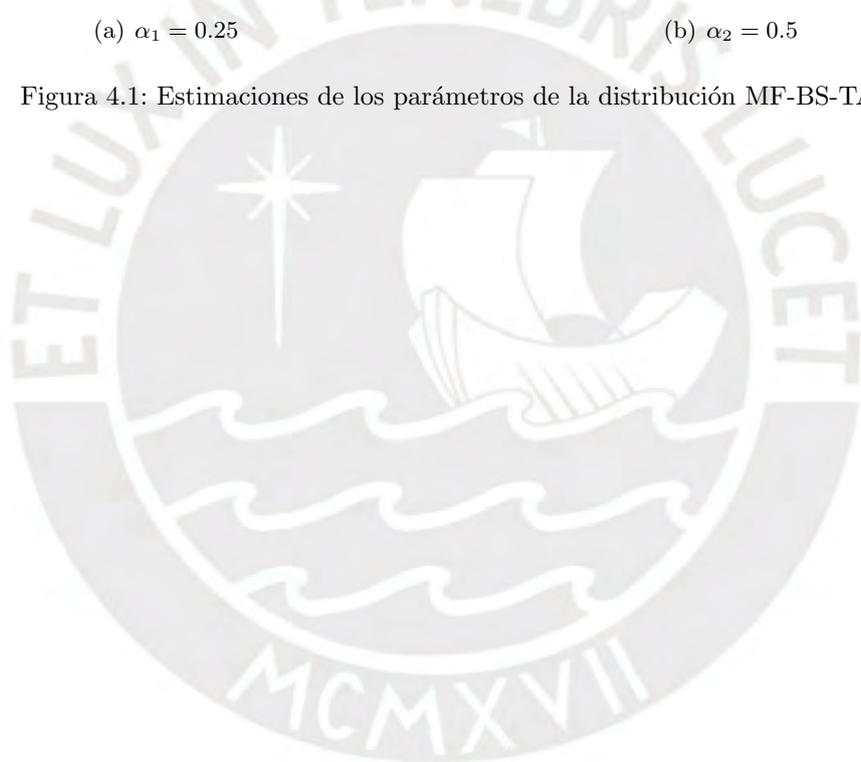


Figura 4.1: Estimaciones de los parámetros de la distribución MF-BS-TA



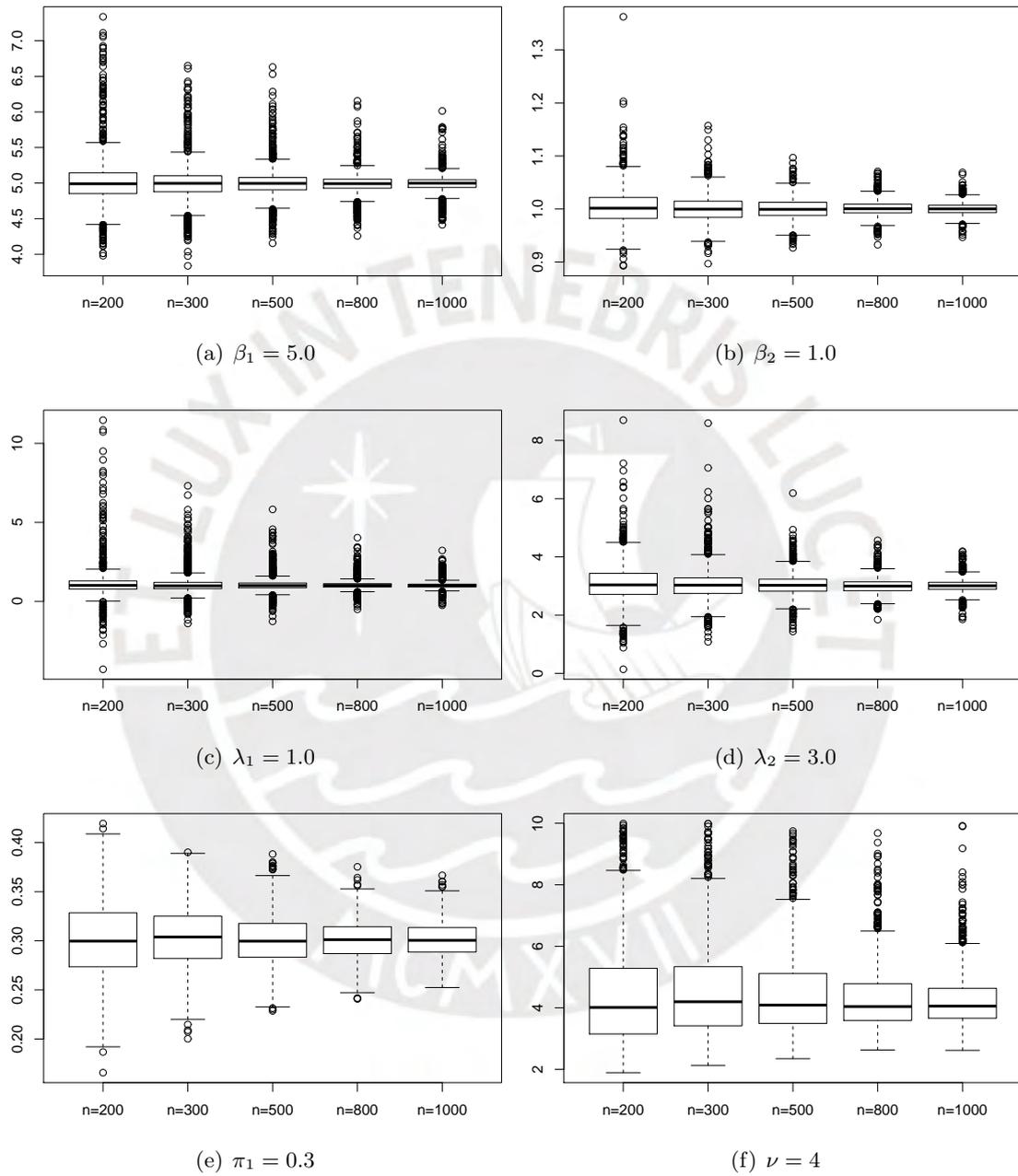


Figura 4.1: Estimaciones de los parámetros de la distribución MF-BS-TA

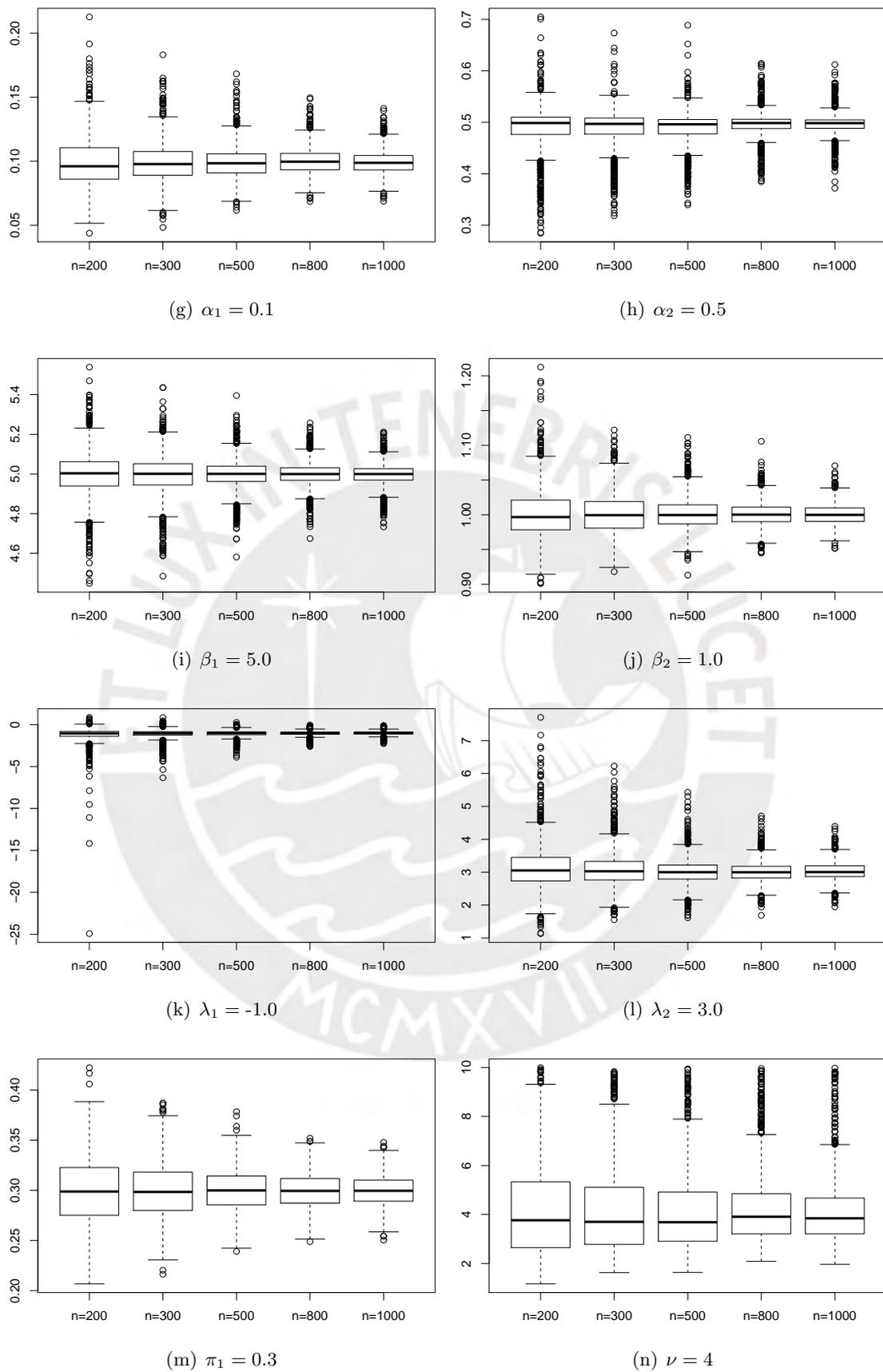


Figura 4.2: Estimaciones de los parámetros de la distribución MF-BS-SLA

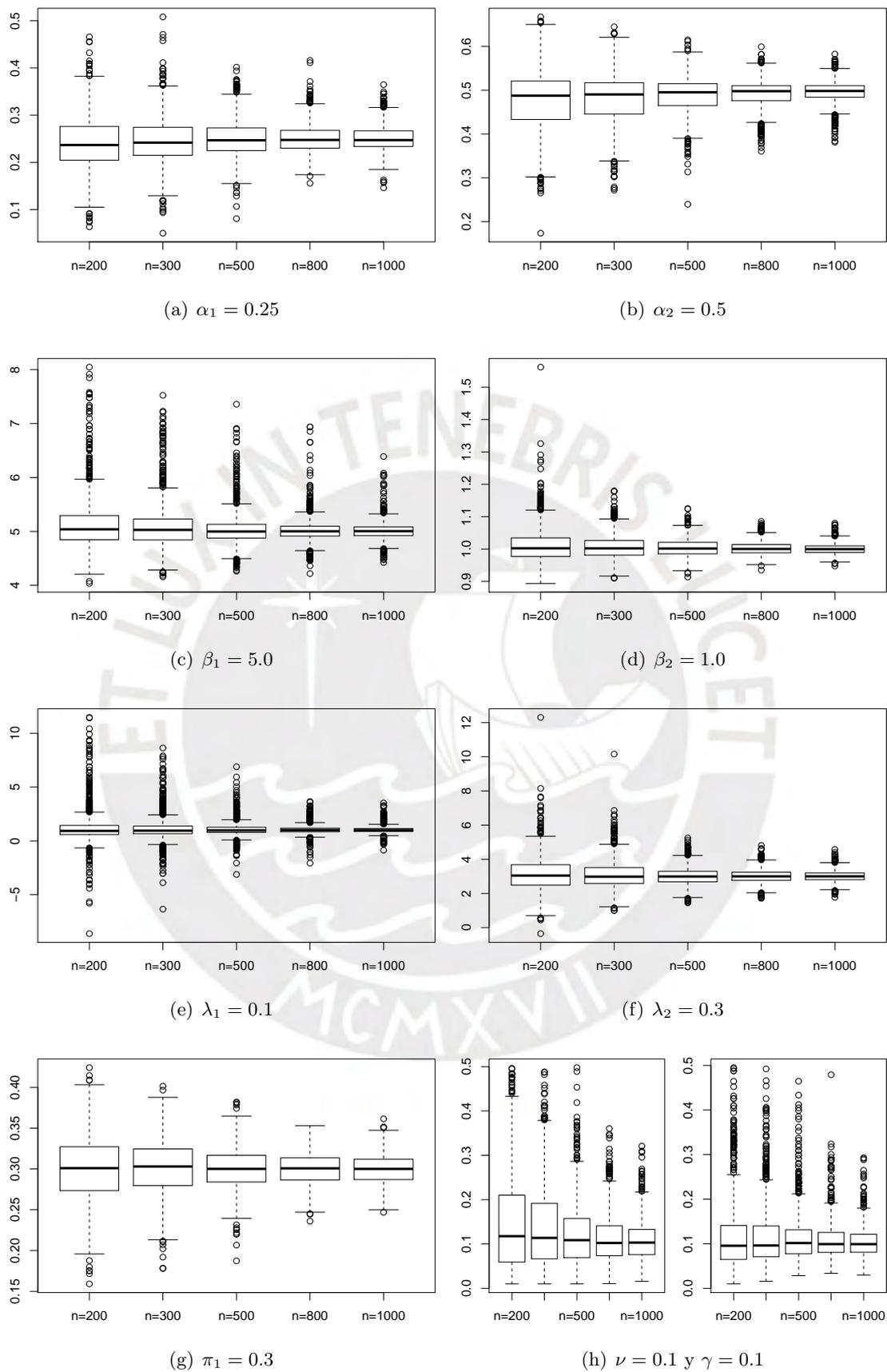


Figura 4.3: Estimaciones de los parámetros de la distribución MF-BS-NCA

## Capítulo 5

### Aplicación

En la presente aplicación, consideramos el índice de masa corporal de 2107 hombres de 18 a 80 años. El conjunto de datos proviene de la Encuesta Nacional de Examen de Salud y Nutrición, realizada por el Centro Nacional de Estadísticas de Salud (NCHS) de los Centros para el Control de Enfermedades (CDC) en los EE. UU. Está disponible en el paquete `mixsmsn` (Prates et al., 2013). Estos datos han sido analizados por Basso et al. (2010), quienes ajustaron a una mixtura finita de una distribución normal asimétrica (FM-NA). El algoritmo de estimación se implementa en el paquete R `mixsmsn` donde se usa el algoritmo de agrupamiento de  $k$ -medias para obtener los valores iniciales. Aplicamos el algoritmo EM para realizar la estimación de MV para el modelo FM-BS-MENA para diferentes número de componentes. La Figura 5.1 (paneles a y b) muestran el histograma con el kernel de una f.d.p. estimada superpuesta y el gráfico de cajas del índice de masa corporal respectivamente.

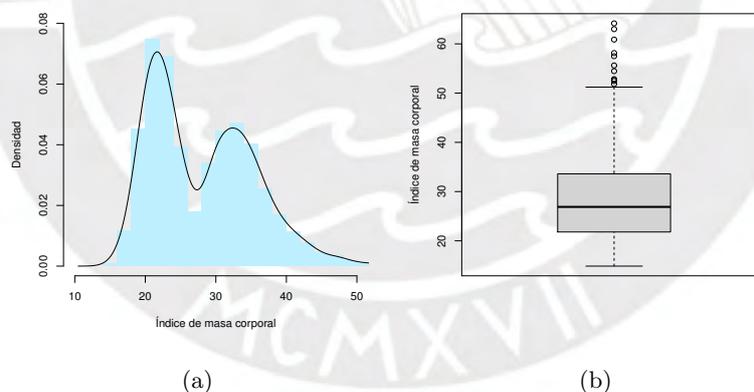


Figura 5.1: (a) Histograma con la superposición de la estimación de la f.d.p. por un kernel y (b) diagrama de cajas del índice de masa corporal

La Tabla 5.1 compara la estimación de varios modelos de mixtura para  $G = 1$  hasta  $G = 4$  componentes a través del criterio de Información de Akaike (AIC) y el Criterio de Información Bayesiano (BIC). Es importante mencionar que hay ciertos modelos que no llegan a converger tal como es el caso de modelo de MF-BS-SLA para  $G = 2, 3$  y 4 componentes (esto tal vez se deba a que la integral presente en la f.d.p no llega a ser resuelta numéricamente, situación que se da en algunos conjuntos de datos) y es por ello que no son considerados en dicha tabla. Observamos que el mejor modelo según el AIC es para el modelo MF-BS-TA con  $G = 3$  componentes y según el BIC el mejor modelo es también la MF-BS-TA pero con

$G = 2$  componentes.

La Figura 5.2, panel (a), muestra la log-verosimilitud perfilada de  $\nu$  en dos dimensiones, para el caso de MF-BS-TA y MF-BS-NCA con  $G = 2$  componentes, mientras que la Figura 5.2, panel (b), muestra la superficie de la log-verosimilitud perfilada de  $\nu$  para la distribución MF-BS-NCA en la que la etiqueta “gamma” representa el parámetro  $\gamma$  y la etiqueta “nu” representa el parámetro  $\nu$ , mientras que la etiqueta “z” representa el valor de la log-verosimilitud perfilada considerando la variación de  $\gamma$  y  $\nu$ . La Figura 5.2, panel (c), es la representación en dos dimensiones de log-verosimilitud perfilada de la distribución MF-BS-NCA. Según la Subsección 3.2.5 los valores iniciales del modelo con MF-BS-TA con  $G = 2$  componentes son  $\hat{\alpha}_1^{(0)} = 0,1320$ ,  $\hat{\alpha}_2^{(0)} = 0,1187$ ,  $\hat{\beta}_1^{(0)} = 34,6779$ ,  $\hat{\beta}_2^{(0)} = 22,0139$ ,  $\hat{\lambda}_1^{(0)} = 1$ ,  $\hat{\lambda}_2^{(0)} = 1$ ,  $\hat{p}_1^{(0)} = 0,4699$  y  $\hat{\nu}^{(0)} = 0,1320$ . La Tabla 5.2 presenta los estimadores de máxima verosimilitud de  $\alpha_1$ ,  $\alpha_2$ ,  $\beta_1$ ,  $\beta_2$ ,  $\lambda_1$ ,  $\lambda_2$ ,  $p_1$  y  $\nu$  con sus respectivos errores estándar (EE) para el modelo de MF-BS-NA, MF-BS-TA y MF-BS-NCA con  $G = 2$  componentes donde podemos observar que los errores estándar del modelo MF-BS-TA tienen los menores EE respecto al modelo MF-BS-NA y MF-BS-NCA, notar que modelo MF-BS-SLA no fue adicionado en la tabla por problemas de convergencia. En la Figura 5.3 muestra el histograma del índice de masa corporal con la superposición de la estimación de la f.d.p. estimada del modelo de MF-BS-TA con diferentes componentes. El gráfico muestra que los modelos de MF-BS-TA ( $G = 2$ ,  $G = 3$  y  $G = 4$  componentes) se adaptan a la forma del histograma con precisión.

Cuadro 5.1: Comparación de la máxima log-verosimilitud, BIC para el ajuste del modelo MF-BS-MENA usando el conjunto de datos del índice de masa corporal. El número de parámetros  $m$ , el Ranking representa la clasificación de los modelos MF-BS-MENA, iter representa el número de iteraciones hasta la convergencia y tiempo (segundos) de convergencia CPU

	G	m	log-verosimilitud	AIC	Ranking	BIC	Ranking	iter	tiempo
MF-BS-NA	4	1	-7490.760	14189	8	14229	8	13	0.460
MF-BS-TA	5	1	-7060.800	14130	7	14152	6	396	7.289
MF-BS-SLA	5	1	-7130.349	14269	11	14291	11	1	1.791
MF-BS-NCA	6	1	-7054.124	14118	4	14147	4	127	4.678
MF-BS-NA	7	2	-7490.760	14190	9	14230	9	15	0.530
MF-BS-TA	9	2	-6867.914	13752	2	13797	1	19	1.732
MF-BS-NCA	10	2	-7042.379	14103	3	14154	7	519	57.567
MF-BS-NA	11	2	-7478.850	14338	12	14400	12	12	1.672
MF-BS-TA	13	2	-6858.164	13740	1	13808	2	18	3.072
MF-BS-NCA	6	3	-7054.124	14118	4	14147	4	127	4.762
MF-BS-NA	15	4	-7478.850	14338	12	14400	12	12	2.258
MF-BS-TA	5	4	-7107.757	14224	10	14246	10	4	0.107
MF-BS-NCA	6	4	-7055.216	14120	6	14149	5	74	2.710

Notamos también que en la Figura 5.3 se muestra las estimaciones de los parámetros del modelo MF-BS-TA para diferentes componentes ( $G = 2$ ,  $G = 3$  y  $G = 4$ ) y observamos que para el caso  $G = 3$  y  $G = 4$  componentes tenemos que  $\hat{\nu} = 21,6781$  y  $\hat{\nu} = 43,9687$  con

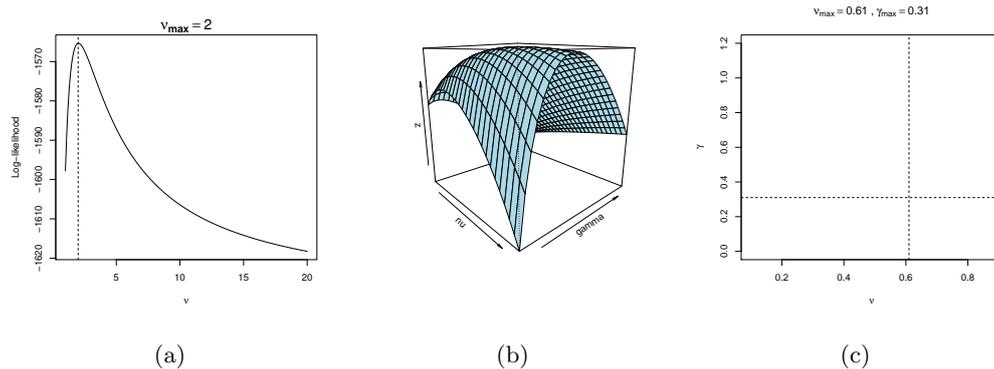


Figura 5.2: Log-verosimilitud perfilada de  $\nu$  para modelos con  $G = 2$  componentes para los modelos de (a) MF-BS-TA, (b) y (c) MF-BS-NCA

Cuadro 5.2: Estimación de parámetros vía el algoritmo EM con sus correspondientes errores estándar (EE)

	MF-BS-NA	EE	MF-BS-TA	EE	MF-BS-NCA	EE
$\alpha_1$	0.2729	1.0193	0.1326	0.0065	0.0928	0.3777
$\alpha_2$	0.1633	0.0478	0.1064	0.0181	0.0889	0.1483
$\beta_1$	29.4095	137.1913	32.5842	1.7642	19.2017	7.2154
$\beta_2$	21.1842	1.0485	20.9615	0.8143	20.0000	2.3194
$\lambda_1$	0.2779	23.0439	0.2743	0.5962	5.8306	5.2202
$\lambda_2$	1.0675	0.9181	0.4371	0.5543	2.8818	1.7881
$p_1$	0.5347	0.0655	0.5125	0.0261	0.3710	8.2382
$\nu$	-	-	6.7486	-	0.9219	-
$\gamma$	-	-	-	-	0.0425	-

lo que podemos afirmar que el modelo MF-BS-TA es la misma que un modelo MF-BS-NA. Finalmente concluimos que el mejor modelo que se ajusta a los datos de índice de masa corporal es el MF-BS-TA con  $G = 2$  componentes.

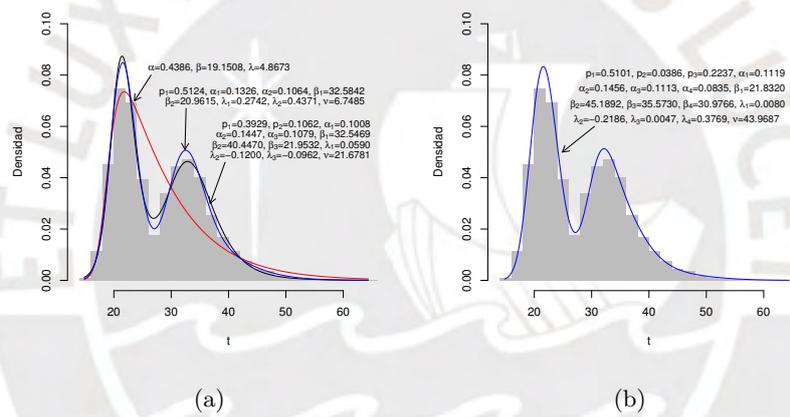


Figura 5.3: (a) y (b) Histograma del índice de masa corporal con la superposición de la estimación de la f.d.p. estimada del modelo de MF-BS-TA con diferentes componentes

## Capítulo 6

# Conclusiones

### 6.1. Conclusiones

- Se propuso el modelo MF-BS-MENA, como resultado de una extensión al modelo de los modelos propuestos por Maehara (2018a) y Benites, Maehara, Vilca y Marmolejo-Ramos (2017) que permite modelar multimodalidad y colas pesadas con tres casos especiales.
- Se realizaron tres estudios de simulación para los casos especiales de la distribución MF-BS-MENA. Se demostró la precisión de las estimaciones por medio del algoritmo ECM, y la proximidad entre los errores aproximados y las desviaciones por Monte Carlo.
- Se realizaron aplicaciones con datos reales sobre el porcentaje corporal. El modelo MF-BS-TA con dos componentes fue el que mejor se ajustó a los datos, seguido por el modelo MF-BS-NCA y MF-BS-NC con dos componentes y el MF-BS-N como cuarto mejor modelo según el AIC.

### 6.2. Sugerencias para investigaciones futuras

- Un ámbito de investigación es el desarrollo y comparación de propuestas de mejora de múltiples técnicas de obtención de valores iniciales, principalmente para distribuciones MF-BS-MENA con distintos componentes ( $G$ ).
- También podemos desarrollar la estimación del modelo MF-BS-MENA bajo un enfoque de inferencia bayesiana a través de métodos Monte Carlo basados en cadenas de Markov.
- Otro aspecto a desarrollar son modelos de regresión cuyos términos de error se distribuyan como los casos especiales de la MF-BS-MENA.

## Apéndice A

### Resultados teóricos

Derivación de la distribución Birbaum-Saunders

De acuerdo a Birnbaum y Saunders (1969).

1. Sea  $l_i$  una secuencia de fuerzas aplicadas sobre un material en específico de manera cíclica  $\forall m > 1, i = 1, \dots, m$ .
2. Sea  $X_i$  la extensión de un incremento gradual de la ruptura en el  $i$ -ésimo ciclo.
3. La extensión de la ruptura en el ciclo  $(j + 1)$  es:

$$Y_{j+1} = X_{jm+1} + \dots + X_{jm+m},$$

donde  $Y_j$  es una variable aleatoria independiente para cada ciclo con media  $\mu$  y varianza  $\sigma^2$

El modelo Birnbaum-Saunders busca el (menor) número de ciclos en el que  $W_n$  excede un valor  $\omega$

$$W_n = \sum_{j=1}^n Y_j.$$

Teniendo en cuenta que  $Y_j \sim N(\mu, \sigma^2)$ :

$$\begin{aligned} P(C \leq n) &= 1 - H_n(\omega) \\ &= P\left(\sum_{i=1}^n \frac{Y_i - \mu}{\sigma\sqrt{n}} > \frac{\omega - n\mu}{\sigma\sqrt{n}}\right) \\ &= 1 - \Phi\left(\frac{\omega - n\mu}{\sigma\sqrt{n}}\right). \end{aligned}$$

Propiedad de la normal estándar:  $1 - \Phi(X) = \Phi(-X)$

$$\begin{aligned}
 &= \Phi\left(-\left(\frac{\omega - n\mu}{\sigma\sqrt{n}}\right)\right) \\
 &= \Phi\left(\frac{n\mu}{\sigma\sqrt{n}} - \frac{\omega}{\sigma\sqrt{n}}\right) \\
 &= \Phi\left(\frac{\sqrt{n}\sqrt{n}\sqrt{\mu}\sqrt{\mu}}{\sigma\sqrt{n}} \times \frac{\sqrt{\omega}}{\sqrt{\omega}} - \frac{\sqrt{\omega}\sqrt{\omega}}{\sigma\sqrt{n}} \times \frac{\sqrt{\mu}}{\sqrt{\mu}}\right) \\
 &= \Phi\left(\frac{\sqrt{\mu\omega}}{\sigma} \left(\sqrt{\frac{n\mu}{\omega}} - \sqrt{\frac{\omega}{\mu}}\right)\right) \\
 &= \Phi\left(\frac{1}{\alpha}\xi\left(\frac{Y}{\beta}\right)\right),
 \end{aligned}$$

donde:  $\alpha = \frac{\sigma}{\sqrt{\mu\omega}}$ ,  $\beta = \frac{\omega}{\mu}$ ,  $\xi(t) = \sqrt{t} - \frac{1}{\sqrt{t}}$ ,  $t = \frac{Y}{\beta}$ ,  $Y = n$ .

Representación estocástica

Sea  $Z_0 \sim N(0, 1)$

$$\begin{aligned}
 Z_0 &= \frac{1}{\alpha}\xi\left(\frac{Y}{\beta}\right) \\
 Z_0 &= \frac{1}{\alpha} \left( \sqrt{\frac{Y}{\beta}} - \sqrt{\frac{\beta}{Y}} \right).
 \end{aligned}$$

Definimos:  $g = \sqrt{\frac{Y}{\beta}}$

$$\begin{aligned}
 Z_0 &= \frac{1}{\alpha} \left( g - \frac{1}{g} \right) \\
 \alpha Z_0 &= \left( \frac{g^2 - 1}{g} \right) \\
 g\alpha Z_0 &= g^2 - 1 \\
 0 &= g^2 - g\alpha Z_0 - 1 \\
 g &= \frac{-(-\alpha Z_0) \pm \sqrt{(\alpha Z_0)^2 - 4(1)(-1)}}{2(1)}.
 \end{aligned}$$

De la ecuación anterior tenemos las siguientes raíces:

$$g_1 = \frac{\alpha}{2}Z_0 + \sqrt{\left(\frac{\alpha Z_0}{2}\right)^2 + 1} \quad \text{y} \quad g_2 = \frac{\alpha}{2}Z_0 - \sqrt{\left(\frac{\alpha Z_0}{2}\right)^2 + 1}.$$

Por propiedad de raíces de polinomios Cardano Vieta:

$$g_1 \times g_2 = \frac{c}{a} = \frac{-1}{1}.$$

Luego

$$-g_2 = \frac{1}{g_1} = -\frac{\alpha}{2}Z_0 + \sqrt{\left(\frac{\alpha Z_0}{2}\right)^2 + 1}.$$

Entonces

$$\sqrt{\frac{Y}{\beta}} = \frac{\alpha}{2}Z_0 + \sqrt{\left(\frac{\alpha Z_0}{2}\right)^2 + 1}.$$

Por tanto

$$Y = \beta \left( \frac{\alpha}{2}Z_0 + \sqrt{\left(\frac{\alpha Z_0}{2}\right)^2 + 1} \right)^2.$$

De acuerdo a Saunders (1974) la expresión anterior permite definir las propiedades de reciprocidad de la distribución Birbaum-Saunders. Principalmente debido a que:

$$\xi(t) = -\xi\left(\frac{1}{t}\right), \quad \text{y} \quad \frac{1}{g(Z_0)} = g(-Z_0).$$

El Teorema 3.1 en Birnbaum y Saunders (1969) establece que si  $Y$  tiene una distribución  $BS(\alpha, \beta)$ . Entonces  $Y^{-1} \sim BS(\alpha, \beta^{-1})$  y que para todo  $c > 0$ ,  $cY \sim BS(\alpha, c\beta)$ .

La segunda parte de este teorema se puede demostrar este teorema a partir de la representación estocástica donde para todo  $a > 0$

$$cY = c\beta \left( \frac{\alpha}{2}Z_0 + \sqrt{\left(\frac{\alpha Z_0}{2}\right)^2 + 1} \right)^2.$$

La primera parte del teorema puede demostrarse considerando que  $\xi(t) = -\xi(\frac{1}{t})$ . Así

$$Z_0 = -\frac{1}{\alpha} \left( \sqrt{\frac{\beta}{Y}} - \sqrt{\frac{Y}{\beta}} \right).$$

Sea:  $k = \sqrt{\frac{\beta}{Y}}$

$$Z_0 = -\frac{1}{\alpha} \left( k - \frac{1}{k} \right)$$

$$\alpha k Z_0 = -k^2 + 1$$

$$0 = k^2 + \alpha k Z_0 - 1$$

$$k = \frac{-(\alpha Z_0) \pm \sqrt{(\alpha Z_0)^2 - 4(1)(-1)}}{2(1)}.$$

De la ecuación anterior obtenemos:

$$\sqrt{\frac{\beta}{Y}} = \frac{\alpha}{2}Z_0 + \sqrt{\left(\frac{\alpha Z_0}{2}\right)^2 + 1}$$

$$\frac{1}{Y} = \frac{1}{\beta} \left( \frac{\alpha}{2}Z_0 + \sqrt{\left(\frac{\alpha Z_0}{2}\right)^2 + 1} \right)^2.$$

La metodología utilizada por Birnbaum y Saunders (1969) es replicada por Vilca y Leiva (2006b) considerando una variable  $Z_1$  que se distribuye como una normal asimétrica con parámetros  $(0, 1, \lambda)$ .

$$Z_1 = \frac{1}{\alpha} \left( \sqrt{\frac{Y}{\beta}} - \sqrt{\frac{\beta}{Y}} \right) \sim \text{NA}(0, 1, \lambda).$$

Siguiendo la misma metodología que Birnbaum y Saunders (1969), tenemos que:

$$Y = \beta \left( \frac{\alpha}{2}Z_1 + \sqrt{\left(\frac{\alpha Z_1}{2}\right)^2 + 1} \right)^2.$$

Por las propiedades de reciprocidad, se cumple que  $Y^{-1} \sim \text{BS-NA}(\alpha, \beta^{-1}, \lambda)$  y para todo  $c > 0$ ,  $cY \sim \text{BS-NA}(\alpha, c\beta, \lambda)$ .

Similarmente, Maehara (2018a) considera una variable  $Z_1$  que se distribuye como una normal asimétrica con parámetros  $(0, 1, \lambda)$ .

$$Z_2 = \frac{1}{\alpha} \left( \sqrt{\frac{Y}{\beta}} - \sqrt{\frac{\beta}{Y}} \right) \sim \text{MENA}(0, 1, \lambda; S).$$

donde  $Y$  tiene la siguiente representación estocástica.

$$Y = \beta \left( \frac{\alpha}{2}Z_2 + \sqrt{\left(\frac{\alpha Z_2}{2}\right)^2 + 1} \right)^2,$$

donde  $Z_2 = \kappa^{1/2}(U)Z_1$ ,  $Z_1 \sim \text{NA}(0, 1, \lambda)$ ,  $\kappa(U) > 0$  es una función de mixturas de la variable aleatoria  $U$ .

## Bibliografía

- Achcar, J. (1993). Inference for the Birnbaum-Saunders fatigue life model using Bayesian methods, *Computational Statistics and Data Analysis* **15**: 367–380.
- Azzalini, A. (1985). A class of distributions which includes the normal ones, *Scandinavian Journal of Statistics* **12**: 171–178.
- Azzalini, A. (2005). The skew-normal distribution and related multivariate families, *Scandinavian Journal of Statistics* **32**: 159–188.
- Balakrishnan, N., Gupta, R., Kundu, D., Leiva, V. y Sanhueza, A. (2011). On some mixture of models based on Birbaum-Saunders distributions and associated inference, *Journal of Statistical Planning and Inference* **55**: 2175–2190.
- Balakrishnan, N., Leiva, V., Sanhueza, A. y Vilca, F. (2009). Estimation in the Birnbaum-Saunders distribution based on scale-mixture of normals and the EM-algorithm, *Statistics and Operations Research Transactions* **33**: 171–192.
- Barros, M., Paula, G. y Leiva, V. (2008). A new class of survival regression models with heavy-tailed errors: robustness and diagnostics, *Lifetime Data Analysis* **14**: 316–332.
- Basford, K., Greenway, D., McLachlan, G. y Peel, D. (1993). Standard errors of fitted component means of normal mixtures, *Computational Statistics* **12**: 1–17.
- Basso, R., Lachos, V., Cabral, C. y Ghosh, P. (2010). Robust mixture modeling based on scale-mixture of skew-normal distributions, *Computational Statistics and Data Analysis* **54**: 2926–2941.
- Benites, L. (2019). *Mixtura finita de la distribución Birnbaum-Saunders basada en la distribución de mixtura de escala normal*, Master's thesis, National University of San Marcos.
- Benites, L., Maehara, R., Vilca, F. y Marmolejo-Ramos, F. (2017). Finite Mixture of Birnbaum-Saunders distributions using the k-bumps algorithm, *arXiv:1708.00476* .
- Birnbaum, Z. y Saunders, S. (1969b). Estimation for a family of life distributions with applications to fatigue, *John Wiley and Sons* **6**: 328–347.
- Birnbaum, Z. W. y Saunders, S. C. (1969). A new family of life distributions, *Journal of Applied Probability* **6**: 319–327.
- Branco, M. y Dey, D. (2001). A general class of multivariate skew-elliptical distributions, *Journal of Multivariate Analysis* **79**: 99–113.
- Cabral CRB, Lachos VH, P. M. (2012). Multivariate mixture modeling using skew-normal independent distributions, *Computational Statistics and Data Analysis* **56**: 126–142.
- Cook, R. y Weisberg, S. (1982). *Residual and Influence in Regression*, Chapman and Hall, New York.

- Desmond, A. F. (1985). Stochastic models of failure in random environments, *Canadian Journal of Statistics* **13**: 171–183.
- Desmond, A. F. (1986). On the relationship between two fatigue-life models, *IEEE Transactions on Reliability* **35**: 167–169.
- Díaz-García, J. y Leiva, V. (2005). A new family of life distributions based on the contoured elliptically distributions, *Journal of Statistical Planning and Inference* **128**: 88–95.
- Dupuis, D. y Mills, J. (1997). Robust estimation of the Birnbaum-Saunders distribution, *IEEE Transactions on Reliability* **47**: 88–95.
- Engelhardt, M., Bain, L. y Wright, F. (1981). Inferences on the parameters of the Birnbaum-Saunders fatigue life distribution based on maximum likelihood estimation, *Technometrics* **23**: 251–256.
- Jin, X. y Kawczak, J. (2003). Birnbaum-Saunders and Lognormal Kernel Estimators for Modelling Duration in High Frequency Financial Data, *Annals of Economics and Finance* **4**: 103–124.
- Leiva, V. (2016). *The Birnbaum-Saunders Distribution*, Elsevier.
- 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. y Balakrishnan, N. 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.
- Lu, M. y Chang, D. (1997). Bootstrap prediction intervals for the birnbaum-saunders distribution.
- Maehara, R. (2018a). *An extension of Birnbaum-Saunders distributions based on scale mixtures of skew-normal distributions with applications to regression models*, PhD thesis, University of Sao Paulo.
- Maehara, R. (2018b). *Mixtura finita basada en la distribución Birnbaum-Saunders normal asimétrica*, Master's thesis, National Agrarian University La Molina.
- Maehara, R. y Benites, L. (2015). *bssn: Birnbaum-Saunders Model Based on Skew-Normal Distribution*. R package version 0.6.  
**URL:** <http://CRAN.R-project.org/package=bssn>
- Mann, N. R., Schafer, R. E. y Singpurwalla, N. (1974). *Methods for Statistical Analysis of Reliability and Life Data*, John Wiley and Sons, New York.
- Marshall, A. y Olkin, I. (2007). *Life Distributions-Structure of Nonparametric, Semiparametric, and Parametric Families*, Springer, New York.
- McLachlan, A. y Peel, D. (2000). *Finite Mixture Models*, Jhon Wiley and Sons Inc.
- Meng, X. y Rubin, D. (1993). Maximum likelihood estimation via the ecm algorithm: A general framework, *Biometrika* **80**: 267–278.
- Murphy, T. y Martin, D. (2003). Mixtures of distance-based models for ranking data, *Computational Statistics and Data Analysis* **41**: 645–655.

- Ng, H., Kundu, D. y Balakrishnan, N. (2003). Modified moment estimation for the two-parameter Birnbaum-Saunders distribution, *Computational Statistics and Data Analysis* **43**: 283–298.
- O'Hagan, A. y Leonard, T. (1976). Bayes estimation subject to uncertainty about constraints, *Biometrika* **63**: 201–203.
- Owen, W. y Padgett, W. (1999). Accelerated test models for system strength based on Birnbaum-Saunders distribution, *Lifetime Data Analysis* **5**: 133–147.
- Paula, G., Leiva, V., Barros, M. y Liu, S. (2011). Robust statistical modeling using the Birnbaum-Saunders-t distribution applied to insurance, *Applied Stochastic Models in Business and Industry* **19**: 235–249.
- Saunders, S. C. (1974). A family of random variables closed under reciprocation, *Journal of American Statistical Association* **6**: 533–539.
- Saunders, S. C. (1976). The problems of estimating fatigue with low probability of failure, *Engineering Fracture Mechanics* **8**: 208–215.
- Tanner, M. A. (1996). *Tools for Statistical Inference Methods for the Exploration of Posterior Distributions and Likelihood Functions*, 3 edn, Springer, New York.
- Turner, R. (2000). Estimating the propagation rate of viral infection of potato plants via mixtures of regressions, *Journal of Royal Statistical Society* **49**: 371–384.
- Vilca, F. y Leiva, V. (2006a). A new fatigue life model based on family of skew-elliptical distributions, *Communications in Statistics - Theory and Methods* **35**: 229–244.
- Vilca, F. y Leiva, V. (2006b). A new fatigue life model based on the family of skew-elliptical distributions, *Communications in Statistics* **35**: 229–244.
- Vilca, F., Santana, L., Leiva, V. y Balakrishnan, N. (2011). Estimation of extreme percentiles in Birnbaum-Saunders distributions, *Computational Statistics and Data Analysis* **55**: 1665–1678.
- Yau, K., Lee, A., Kundu, D. y Ng, A. (2003). Finite mixture regression with random effects: application to neonatal hospital length of stay, *Computational Statistics and Data Analysis* **41**: 359–366.
- Zeller, C. B., Cabral, C. R., Lachos, V. H. y Benites, L. (2019). Finite mixture of regression models for censored data based on scale mixtures of normal distributions, *Adv. Data Anal. Classif.* **13**(1): 89–116.