



Casa abierta al tiempo

**UNIVERSIDAD AUTÓNOMA METROPOLITANA**  
**Unidad Iztapalapa**

---

---

Maestría en Ciencias Matemáticas Aplicadas e Industriales

---

**Modelación de datos composicionales vía mezclas  
de distribuciones normales multivariadas**

---

**T E S I S**

Para obtener el grado de:

**Maestro en Ciencias**  
**(Matemáticas Aplicadas e Industriales)**

Presenta:

**Arnoldo Daniel Miranda Fournier**

Asesor:

**Dr. Gabriel Núñez Antonio**

México, CDMX a 22 de mayo de 2020



# Índice general

Resumen	V
Introducción	VII
<b>1. Preliminares</b>	<b>1</b>
1.1. Datos Composicionales . . . . .	1
1.2. El espacio muestral y las operaciones asociadas a los datos composicionales	2
1.3. Problemas con el análisis clásico . . . . .	6
1.3.1. Transformaciones log-ratio . . . . .	7
1.4. Estadística Descriptiva . . . . .	10
1.4.1. Descripción Gráfica . . . . .	10
1.4.2. Medidas Descriptivas . . . . .	13
1.4.3. Distribuciones de Probabilidad en el Simplex . . . . .	15
1.5. Estadística Bayesiana . . . . .	16
1.6. Métodos de Simulación . . . . .	19
1.6.1. Breve introducción a las Cadenas de Markov . . . . .	20
1.6.2. Métodos Monte Carlo vía Cadenas de Markov . . . . .	22
<b>2. El Modelo Propuesto</b>	<b>27</b>
2.1. El modelo Normal Univariado . . . . .	27
2.2. El modelo Normal Multivariado . . . . .	29
2.3. Un modelo no-paramétrico . . . . .	30
<b>3. Inferencia Bayesiana del Modelo</b>	<b>35</b>
3.1. Inferencia para el Modelo Normal . . . . .	35
3.1.1. Inferencia para el Modelo Normal Univariado . . . . .	35
3.1.2. Inferencia para el Modelo Normal Multivariado . . . . .	37
3.2. Inferencia para el Modelo no-paramétrico . . . . .	38
<b>4. Aplicaciones</b>	<b>43</b>
4.1. Datos Simulados . . . . .	43
4.2. Datos Reales . . . . .	57
<b>Conclusiones y Perspectivas</b>	<b>67</b>
<b>Bibliografía</b>	<b>69</b>
<b>A. Códigos en R</b>	<b>73</b>
A.1. Códigos utilizados en R . . . . .	73



# Resumen

Desde siempre el ser humano se ha enfrentado a procesos que involucran DATOS. En consecuencia, la modelación de fenómenos reales se ha convertido en una tarea muy importante. Hoy en día esta tarea es cada vez más compleja debido al gran volumen de información con el que se cuenta, esto ha llevado a los científicos a buscar metodologías más eficientes para modelar de manera adecuada dicha información.

Un conjunto de datos muy particular son los que forman las variables composicionales, los datos composicionales son aquellos que describen en esencia las partes de un todo y su espacio muestral es el simplex. El problema del análisis estadístico de datos composicionales ha sido y es una fuente de preocupación para muchos científicos desde que en 1897 Karl Pearson en [Pearson, 1897] pusiera de manifiesto la problemática de aplicar métodos estadísticos clásicos para el estudio de los mismos. Es vasta la frecuencia con que aparecen medidas de esta índole en las ciencias aplicadas —ciencias de la tierra (geoquímica, petrología, ...), biología, química, ciencias ambientales, economía, medicina, sociología, ingeniería— y, por ende, la importancia de disponer de herramientas adecuadas para su análisis.

Aunque en la literatura se han propuesto enfoques para el tratamiento de este tipo de datos [Aitchison, 1982], estos se basan en la geometría derivada de log cocientes, donde se asumen ciertas distribuciones paramétricas particulares para los datos transformados. Sin embargo, existen datos con características de asimetría o de grupo que difícilmente pueden ser descritas por un modelo paramétrico. La metodología que se desarrolla en este trabajo de tesis se basa en la propuesta de un modelo no-paramétrico (mezcla infinita de densidades paramétricas) para describir datos composicionales.



# Introducción

En la actualidad, la modelación de fenómenos reales a través de modelos matemáticos es compleja. Aún más, cuando esta descripción se basa en el análisis de modelos donde las inferencias sobre los parámetros es dirigida por DATOS, la tarea se vuelve retadora.

Aunque inicialmente pareciera que la dimensionalidad de los datos no tiene un impacto en la modelación, no es lo mismo establecer y analizar modelos donde las variables (quizá varias) son univariadas, que bajo la situación donde las variables son de naturaleza multivariada. Lo anterior, ya que cuando se trabaja con datos multivariados, un problema crucial es la definición de funciones de probabilidad conjuntas que describan adecuadamente la correlación entre las diferentes componentes de los datos. Por otro lado, como lo menciona el profesor Aitchison en [Aitchison, 1982], un primer paso importante en un problema estadístico es la identificación del espacio muestral donde *viven* los datos. Por ejemplo, en el estudio de variables **composicionales**, el espacio muestral naturalmente asociado es el simplex unitario  $p$ -dimensional  $\mathbb{S}^p$ . Por otro lado, en el caso de los denominados **datos direccionales**, los datos son vectores definidos en la esfera unitaria  $k$  dimensional  $\mathbb{C}^k$ . En ambos casos, hoy en día, la definición de modelos de probabilidad que describan datos de esta naturaleza es un campo abierto en continuo crecimiento.

Este trabajo de tesis se centrará en el estudio y análisis de variables composicionales. Es decir, datos multivariados cuyas componentes representan las partes de un todo y están sujetas a la restricción de suma constante. Este tipo de datos se pueden encontrar de manera natural en distintas áreas del conocimiento. Por ejemplo en procesos electorales, donde la proporción de votos a favor de cada una de las fuerzas electorales es de relevancia política en un país. En Sedimentología, es de especial interés estudiar la relación que existe entre la composición de los sedimentos y las diferentes profundidades del agua. En Ingeniería de Alimentos, con la intención de mejorar la calidad de la leche de vaca se llevan a cabo estudios sobre la composición de la leche antes y después de someter a las vacas a un cierto régimen alimenticio. Así, los datos composicionales son realizaciones de vectores aleatorios de suma constante y es justo esta restricción, aunado al hecho de que el cambio en una de las partes provoca como mínimo el cambio en otra de las componentes, lo que se ha considerado como la fuente de todos los problemas.

Aunque en la literatura se han propuesto enfoques para el tratamiento de este tipo de datos [Aitchison, 1982], estos se basan en la geometría derivada de log-cocientes, donde se asumen distribuciones paramétricas para los datos transformados. Sin embargo, existen datos con características de asimetría o de grupo que difícilmente pueden ser

descritas por un modelo paramétrico. La metodología que se desarrolla en este trabajo de tesis se basa en la propuesta de un modelo no-paramétrico (mezcla infinita de densidades paramétricas) definido en  $\mathbb{R}^p$  para posteriormente mapearlo al simplex unitario.

Específicamente, la propuesta para analizar variables composicionales se basa en la inferencia Bayesiana del modelo propuesto a través de la obtención de muestras de la distribución predictiva posterior, las cuales se obtienen mediante un *slice sampling*. Se evaluará la eficacia del modelo con conjuntos de datos simulados y se aplicará a un conjunto de datos obtenidos en el ejercicio estadístico del Conteo Rápido para la Gubernatura del Estado de Morelos, organizado por el Instituto Nacional Electoral (INE) en 2018.

La estructura de esta tesis es la siguiente. En el Capítulo 1 se revisa la naturaleza y métodos descriptivos de las variables composicionales, así como los retos para especificar modelos que los describan adecuadamente, finalizando con la introducción de conceptos del enfoque Bayesiano de la estadística y de las herramientas de simulación que serán utilizadas.

En el Capítulo 2 se presenta el modelo Normal Univariado, el modelo Normal Multivariado y el modelo no-paramétrico propuesto para la modelación de las variables composicionales. Este último basado en los modelos de mezclas infinitas de normales univariadas y multivariadas respectivamente. Además se introducen conceptos y representaciones que serán fundamentales para llevar a cabo su inferencia, tales como el proceso Dirichlet en su representación Stick-Breaking y el modelo de mezclas de procesos Dirichlet (MDP).

En el Capítulo 3 se presentan las inferencias para el modelo Normal Univariado y Multivariado. Se desarrolla la inferencia para el modelo no-paramétrico, por lo que, se determinan todas las distribuciones condicionales completas asociadas al modelo. Por último, se muestra la manera de simular de la distribución predictiva final del modelo propuesto.

En el Capítulo 4 se realizan diversas aplicaciones para la metodología propuesta. Comenzando con la aplicación del modelo a conjuntos de datos simulados, con el objetivo de evaluar el desempeño del mismo. Terminando con la aplicación de la metodología propuesta a un conjunto de datos reales derivados de los ejercicios de conteos rápidos realizados por el Instituto Nacional Electoral (INE) en las elecciones estatales de 2018.

Finalmente, en el último Capítulo se presentan las conclusiones y perspectivas derivadas del presente trabajo de tesis.



# Capítulo 1

## Preliminares

Este capítulo comienza con un estudio introductorio a los datos composicionales, desde su naturaleza y ejemplificación en el mundo real hasta definir su espacio muestral así como las operaciones que se pueden efectuar sobre los mismos. Posteriormente se abarcan los principales problemas que competen al análisis estadístico de los datos composicionales con la finalidad de introducir la metodología adecuada para su estudio. Se introducen conceptos para su manipulación descriptiva así como para el estudio de distribuciones de probabilidad que pueden ser de gran ayuda para la modelación de los mismos. Posteriormente, se da una breve introducción al enfoque de la estadística bayesiana para finalizar con los métodos de simulación que serán utilizados en la inferencia del modelo propuesto más adelante.

### 1.1. Datos Composicionales

Los datos composicionales son datos que describen cuantitativamente las partes de un todo y aportan solo información relativa entre sus componentes. Los datos composicionales aparecen en forma de vectores de dos o más componentes no negativas, las cuales se expresan regularmente como proporciones, porcentajes o partes por millón de algún todo y cuya suma es un valor constante  $k$  (igual a 1, 100 o  $10^6$ , respectivamente).

Los datos composicionales surgen en diversas ciencias tales como, la geología (análisis de los minerales presentes en las rocas), economía (portafolios de inversión) y química (distribución de contaminantes en agua, aire y suelo), entre otros, a continuación se presentan algunos ejemplos:

- Estudio de la composición de la leche

Para mejorar la calidad de la leche de vaca, se estudia la composición de la leche que produce una de cada treinta vacas antes y después de una dieta estrictamente controlada y un régimen hormonal por un periodo de ocho semanas. Se decide tener el control de otras treinta vacas criadas en las mismas condiciones, pero sobre un régimen regular al establecido. El propósito del experimento es determinar si el nuevo régimen ha producido algún cambio significativo en la composición de las diferentes proteínas de la leche. La pregunta esencial para ese problema es ¿El nuevo régimen ha producido algún cambio en la composición de la leche?

- Sedimentos del lago Ártico en diferentes profundidades

En sedimentología, especímenes de sedimentos son separados en partes mutuamente exclusivos y exhaustivos (por ejemplo, arena, limo y arcilla) y las proporciones de estas partes por peso son llamadas composiciones. Algunas interrogantes que surgen al analizar estos datos son: ¿La composición de sedimentos depende de la profundidad del agua? Si es así, ¿cómo se puede modelar esa dependencia?

- Composición Geoquímica de las Rocas

En petrología, el análisis de las composiciones geoquímicas de las rocas es un aspecto fundamental, tales composiciones se expresan como porcentajes en peso de diez o más óxidos principales o como porcentajes de algunos minerales básicos. ¿De qué manera podría describirse la variabilidad en la composición de las rocas?

- Patrones de Presupuesto Familiar

Un aspecto significativo en el estudio de la demanda del consumidor es el análisis de encuestas sobre el presupuesto familiar, donde la atención suele centrarse en los gastos de una muestra de hogares en una serie de grupos de productos mutuamente excluyentes y exhaustivos y su relación con el gasto total, el ingreso y el tipo de vivienda, la composición del hogar, etc. En una encuesta de una muestra de personas solteras que viven solas en alojamientos alquilados, veinte hombres y veinte mujeres fueron seleccionados al azar y se les pidió que registraran, en un periodo de un mes, sus gastos en los siguientes cuatro grupos de productos excluyentes y exhaustivos:

1. Vivienda incluyendo luz y combustible
2. Productos Alimenticios
3. Otros bienes, incluyendo ropa, calzado y bienes duraderos
4. Servicios, incluidos transporte y vehículos

Algunas preguntas interesantes al observar los datos registrados podrían ser: ¿En qué medida el patrón de la distribución del presupuesto de los gastos para los hombres depende de la cantidad total gastada? ¿Hay diferencias entre hombres y mujeres en sus patrones de gasto? ¿Hay algunos grupos de productos que tienen prioridad en la asignación del gasto?

## 1.2. El espacio muestral y las operaciones asociadas a los datos composicionales

Los datos composicionales son realizaciones de vectores aleatorios de sumas constantes, es decir, cualquier vector  $\mathbf{y}$ , cuyas componentes representan las partes de un todo, está sujeto a la restricción de que la suma de sus componentes sea la unidad, o en el caso general, una constante. Formalmente se tiene la siguiente definición:

**Definición 1.** *Un dato composicional  $\mathbf{y} = (y_1, y_2, \dots, y_{p+1})$  con  $p + 1$  partes, es un vector con componentes estrictamente positivas, tal que la suma de todas ellas es igual a una constante  $k$ . Su espacio muestral es el simplex  $\mathbb{S}^{p+1}$ , definido por:*

$$\mathbb{S}^{p+1} = \left\{ \mathbf{y} = (y_1, y_2, \dots, y_{p+1}) \in \mathbb{R}^{p+1} : y_i > 0 \text{ (} i=1,2,\dots,p+1 \text{)} \quad y \quad \sum_{i=1}^{p+1} y_i = k \right\}$$

Cabe señalar que  $p \geq 1$ , pues un dato solo puede ser composicional si tiene al menos dos componentes, ya que de lo contrario no se puede hablar de las partes de un total, pues al informar una sola parte implícitamente estamos relacionando un total predefinido o una parte complementaria. En su formulación estadística, los datos composicionales son realizaciones de una composición, vector aleatorio cuyo recorrido está en  $\mathbb{S}^{p+1}$ , la notación  $(p + 1)$  cobrará sentido más adelante al introducir la metodología a utilizar la cual reducirá en uno la dimensión.

Antes de indicar las problemáticas específicas que surgen al querer aplicar el análisis estadístico clásico a los datos composicionales, se presentan definiciones de gran importancia para su manipulación.

A partir de un vector en  $\mathbb{R}^{p+1}$  con componentes positivas, es decir, en  $\mathbb{R}_+^{p+1}$  siempre se puede obtener un dato composicional en  $\mathbb{S}^{p+1}$ . Basta con dividir cada una de sus componentes por la suma de todas ellas y multiplicar cada una de ellas por una constante  $k$ , denominada la constante de clausura. Formalmente se tiene la siguiente definición:

**Definición 2.** *El operador clausura  $\mathcal{C} : \mathbb{R}_+^{p+1} \rightarrow \mathbb{S}^{p+1}$  es una transformación que hace corresponder a cada vector  $\mathbf{x} = (x_1, x_2, \dots, x_{p+1})$  en  $\mathbb{R}_+^{p+1}$  su dato composicional asociado  $\mathbf{y} = \mathcal{C}(\mathbf{x}) = \left( \frac{k * x_1}{\sum_{i=1}^{p+1} x_i}, \frac{k * x_2}{\sum_{i=1}^{p+1} x_i}, \dots, \frac{k * x_D}{\sum_{i=1}^{p+1} x_i} \right)$  en  $\mathbb{S}^{p+1}$ , con  $k$  la constante de clausura.*

Es necesario indicar que en adelante se manejará una constante de clausura  $k = 1$ , es decir, la restricción de la suma constante para un dato composicional  $\mathbf{y}$  en  $\mathbb{S}^{p+1}$  cualquiera, se reduce a  $\sum_{i=1}^{p+1} y_i = 1$  y se llamará *símplex unitario  $p + 1$ -dimensional*.

A continuación se presentan algunos ejemplos:

$$\mathbf{x}_1 = (17, 45) \Rightarrow \mathbf{y}_1 = \mathcal{C}(\mathbf{x}_1) = \left( \frac{17}{(17 + 45)}, \frac{45}{(17 + 45)} \right) = (0.27, 0.73)$$

$$\begin{aligned} \mathbf{x}_2 = (12, 3, 4) \Rightarrow \mathbf{y}_2 = \mathcal{C}(\mathbf{x}_2) &= \left( \frac{12}{(12 + 3 + 4)}, \frac{3}{(12 + 3 + 4)}, \frac{4}{(12 + 3 + 4)} \right) \\ &= (0.63, 0.16, 0.21) \end{aligned}$$

En algunos casos es de interés analizar únicamente el valor de las magnitudes relativas sobre un subconjunto de partes para un conjunto de datos composicionales, por lo que, resulta necesario disponer de un procedimiento para la formación de subcomposiciones.

**Definición 3.** *Si  $\mathcal{S}$  es un subconjunto cualquiera de las partes  $1, 2, \dots, p + 1$  de un dato composicional  $\mathbf{y}$  de  $\mathbb{S}^{p+1}$  y  $\mathbf{y}_\mathcal{S}$  simboliza el subvector formado por las correspondientes partes de  $\mathbf{y}$ , entonces  $\mathbf{s} = \mathcal{C}(\mathbf{y}_\mathcal{S})$  recibe el nombre de *subcomposición de las  $\mathcal{S}$  partes de  $\mathbf{y}$* .*

Si bien la formación de una subcomposición es en esencia una transformación de  $\mathbb{S}^{p+1}$  a un *símplex* de dimensión inferior, puede observarse que se tiene la propiedad de conservar la magnitud relativa entre las partes. Observemos ésto con un ejemplo:

$$\mathbf{y} = (0.46, 0.22, 0.32) \text{ tomando a } \mathbf{y}_s = (0.46, 0.32) \Rightarrow \mathbf{s} = \mathcal{C}(\mathbf{y}_s) = (0.59, 0.41)$$

Donde la magnitud relativa de las partes de  $\mathbf{y}$  de donde se tomo el subvector  $\mathbf{y}_s$ , es decir,  $y_1$  y  $y_3$  es  $\frac{y_1}{y_3} = \frac{0.46}{0.32} = 1.44$  que resulta ser la misma que la magnitud relativa de las partes de  $\mathbf{s}$ ,  $\frac{s_1}{s_2} = \frac{0.59}{0.41} = 1.44$ .

Ahora se definen en el Símplex  $\mathbb{S}^{p+1}$  las operaciones de perturbación y potenciación introducidas por Aitchison en [Aitchison, 1982], las cuales resultan ser equivalentes a la suma y producto en  $\mathbb{R}^{p+1}$ , respectivamente.

**Definición 4.** La operación perturbación  $\oplus : \mathbb{S}^{p+1} \times \mathbb{S}^{p+1} \rightarrow \mathbb{S}^{p+1}$  entre

$\mathbf{w} = (w_1, w_2, \dots, w_{p+1}) \in \mathbb{S}^{p+1}$  y  $\mathbf{u} = (u_1, u_2, \dots, u_{p+1}) \in \mathbb{S}^{p+1}$  esta definida por:

$$\mathbf{w} \oplus \mathbf{u} = \mathcal{C}(w_1 * u_1, w_2 * u_2, \dots, w_{p+1} * u_{p+1}) \quad (1.2.1)$$

donde  $\mathcal{C}$  es el operador clausura.

El conjunto de perturbaciones forma un grupo abeliano el cual tiene como elemento neutro la composición  $\mathbb{E} = \left( \frac{1}{p+1}, \frac{1}{p+1}, \dots, \frac{1}{p+1} \right)$ , es decir,  $\mathbf{w} \oplus \mathbb{E} = \mathbb{E} \oplus \mathbf{w} = \mathbf{w}$ . Este elemento neutral juega el papel del vector cero en los datos composicionales. La composición inversa de  $\mathbf{w}$ , denotada por  $\ominus \mathbf{w}$ , está dada por  $\ominus \mathbf{w} = \left( \frac{1}{w_1}, \frac{1}{w_2}, \dots, \frac{1}{w_{p+1}} \right)$ . Esta operación juega el rol de la resta en los datos composicionales, es decir,  $\mathbf{w} \oplus \ominus \mathbf{w} = \ominus \mathbf{w} \oplus \mathbf{w} = \mathbb{E}$ .

**Definición 5.** La perturbación con una composición inversa se suele denotar normalmente con el operador binario  $\ominus : \mathbb{S}^{p+1} \times \mathbb{S}^{p+1} \rightarrow \mathbb{S}^{p+1}$  definido por  $\mathbf{w} \ominus \mathbf{u} := \mathbf{w} \oplus \ominus \mathbf{u}$  para cuales quiera dos elementos  $\mathbf{w}, \mathbf{u}$  del símplex  $\mathbb{S}^{p+1}$ .

Veamos los siguientes ejemplos:

$$\mathbf{w}_1 = (0.38, 0.62), \mathbf{w}_2 = (0.24, 0.76), \mathbf{w}_3 = (0.11, 0.43, 0.46), \mathbf{w}_4 = (0.18, 0.36, 0.46)$$

$$\mathbf{w}_1 \oplus \mathbf{w}_2 = \mathcal{C}(0.38 * 0.24, 0.62 * 0.76) = (0.16, 0.84)$$

$$\mathbf{w}_3 \oplus \mathbb{E} = \mathcal{C}\left(0.11 * \frac{1}{3}, 0.43 * \frac{1}{3}, 0.46 * \frac{1}{3}\right) = (0.11, 0.43, 0.46)$$

$$\mathbf{w}_4 \ominus \mathbf{w}_4 := \mathbf{w}_4 \oplus \ominus \mathbf{w}_4 = \mathcal{C}\left(0.18 * \frac{1}{0.18}, 0.36 * \frac{1}{0.36}, 0.46 * \frac{1}{0.46}\right) = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$$

$$\mathbf{w}_3 \ominus \mathbf{w}_4 := \mathbf{w}_3 \oplus \ominus \mathbf{w}_4 = \mathcal{C}\left(0.11 * \frac{1}{0.18}, 0.43 * \frac{1}{0.36}, 0.46 * \frac{1}{0.46}\right) = (0.22, 0.42, 0.36)$$

**Definición 6.** La operación potenciación  $\odot : \mathbb{R}_+ \times \mathbb{S}^{p+1} \rightarrow \mathbb{S}^{p+1}$  entre  $\mathbf{w} = (w_1, w_2, \dots, w_D) \in \mathbb{S}^{p+1}$  y  $\alpha \in \mathbb{R}_+$  esta definida por:

$$\alpha \odot \mathbf{w} = \mathcal{C}(w_1^\alpha, w_2^\alpha, \dots, w_{p+1}^\alpha) \quad (1.2.2)$$

donde  $\mathcal{C}$  es el operador clausura.

Se dan algunos ejemplos:

$$1 \odot \mathbf{w}_1 = \mathcal{C}(0.38^1, 0.62^1) = (0.38, 0.62)$$

$$0.23 \odot \mathbf{w}_2 = \mathcal{C}(0.24^{0.23}, 0.76^{0.23}) = (0.43, 0.57)$$

$$3 \odot \mathbf{w}_3 = \mathcal{C}(0.11^3, 0.43^3, 0.46^3) = (0.01, 0.44, 0.55)$$

$$\frac{3}{4} \odot \mathbf{w}_4 = \mathcal{C}\left(0.18^{\frac{3}{4}}, 0.36^{\frac{3}{4}}, 0.46^{\frac{3}{4}}\right) = (0.21, 0.36, 0.43)$$

Se puede comprobar que con la operación de perturbación y potenciación por un escalar la terna  $(\mathbb{S}^{p+1}, \oplus, \odot)$  es un Espacio Vectorial, [Billheimer, 2001].

Es importante definir la norma, producto interior y distancia en el Símplex, estos elementos son de gran ayuda para definir líneas, ángulos, ortogonalidad, etc. Por lo tanto, a continuación se presentan algunas definiciones al respecto.

**Definición 7.** Sean  $\mathbf{w} = (w_1, w_2, \dots, w_{p+1})$  y  $\mathbf{u} = (u_1, u_2, \dots, u_{p+1})$  dos elementos de  $\mathbb{S}^{p+1}$ . El producto interior de Aitchison está definido como:

$$\langle \mathbf{w}, \mathbf{u} \rangle_A = \sum_{i=1}^{p+1} \ln \frac{w_i}{g(\mathbf{w})} \cdot \ln \frac{u_i}{g(\mathbf{u})} \quad (1.2.3)$$

donde  $g(\cdot)$  es la media geométrica, es decir,  $g(\mathbf{w}) = \sqrt[p+1]{w_1 \cdot w_2 \cdot \dots \cdot w_{p+1}}$ .

**Definición 8.** Sea  $\mathbf{w} = (w_1, w_2, \dots, w_{p+1})$  un elemento de  $\mathbb{S}^{p+1}$ . La norma de Aitchison de  $\mathbf{w}$  está definida como:

$$\|\mathbf{w}\|_A = \sqrt{\frac{1}{2(p+1)} \sum_{i=1}^{p+1} \sum_{j=1}^{p+1} \left( \ln \frac{w_i}{w_j} \right)^2} \quad (1.2.4)$$

**Definición 9.** La distancia de Aitchison entre  $\mathbf{w} = (w_1, w_2, \dots, w_{p+1})$  y  $\mathbf{u} = (u_1, u_2, \dots, u_{p+1})$  dos elementos de  $\mathbb{S}^{p+1}$  se define como:

$$d_A(\mathbf{w}, \mathbf{u}) = \sqrt{\frac{1}{2(p+1)} \sum_{i=1}^{p+1} \sum_{j=1}^{p+1} \left( \ln \frac{w_i}{w_j} - \ln \frac{u_i}{u_j} \right)^2} \quad (1.2.5)$$

Anteriormente se mencionó que  $\mathbb{S}^{p+1}$  dotado de las operaciones perturbación y potencia es un espacio vectorial, si anexamos el producto interior de Aitchison se puede comprobar que el conjunto  $(\mathbb{S}^{p+1}, \oplus, \odot, \langle \cdot, \cdot \rangle_A)$  conforma un Espacio Euclídeo.

### 1.3. Problemas con el análisis clásico

El problema de los datos composicionales ha sido y es una fuente de preocupación para muchos científicos desde que Karl Pearson en [Pearson, 1897] pusiera de manifiesto la inadecuación de los métodos estadísticos clásicos para el estudio de los mismos. Chayes en [Chayes, 1960] menciona que ignorar la restricción de suma constante o incorporarla indebidamente en un modelado estadístico, puede provocar tener resultados erróneos, es decir, análisis inadecuados que llevan a tener inferencias dudosas o distorsionadas. A continuación se revisan algunos de los problemas que se pueden presentar en el análisis de datos composicionales.

En primer lugar la mayoría de los análisis multivariados comienzan con un análisis univariado de las variables individuales, pero en el caso de datos composicionales cada variable marginal no tiene un significado en si misma aislada del resto.

Los componentes independientes muestran correlaciones negativas, ver más en [Chayes, 1960], este sesgo negativo contradice las interpretaciones habituales de correlación y covarianza, donde la independencia suele estar relacionada con correlación cero.

La covarianza entre dos componentes depende de qué componentes se incluyan en el conjunto original de datos. Esto descalifica las herramientas clásicas basadas en la varianza como descripciones objetivas de la dependencia entre solo dos variables (A esto se le conoce como el problema de la correlación espuria, ver más en [Aldrich et al., 1995]).

Las matrices de varianzas y covarianzas son siempre singulares debido a la restricción de suma constante y muchos de los métodos estadísticos multivariados se basan en matrices de rango completo.

Los componentes no pueden distribuirse normalmente debido al rango de valores acotados, por lo tanto, el modelo normal no es adecuado para las composiciones (no transformadas), por que es incapaz de describir datos acotados. Muchos métodos multivariados están al menos motivados por distribuciones normales multivariadas.

Entonces la mayoría de los métodos multivariados son engañosos o inaplicables para conjuntos de datos composicionales, todo esto pone de manifiesto una diferencia sustancial entre datos composicionales y otros conjuntos de datos multivariados.

A continuación se presentan una serie de principios básicos definidos por Aitchison, que todo análisis composicional debería cumplir para evitar caer en los problemas antes mencionados:

- **Invarianza de escala:** Establece que los datos solo llevan información relativa, de modo que cualquier cambio en su escala no tiene ningún efecto.
- **Invarianza bajo permutación:** Establece que los resultados no dependen del orden en que aparecen las partes de la composición, en otras palabras, los resultados de cualquier análisis no debe depender de la secuencia en la que se proporcionan las componentes de un conjunto de datos.

- **Coherencia subcomposicional:** Establece que los resultados obtenidos para una subcomposición, debe permanecer igual que en la composición original.

Entonces para evitar problemas y erróneas conclusiones todo análisis composicional debería de satisfacer en medida de lo posible estos principios.

Se debe mencionar que la definición clásica de los coeficientes de covarianza y correlación no son coherentes subcomposicionalmente. Esto se conecta con dos problemas:

- **Correlación espuria:** La correlación entre cocientes con denominador común son arbitrarios en un grado incierto. Para más detalles sobre la correlación espuria el lector puede referirse a [Aldrich et al., 1995].
- **El sesgo negativo:** Aparece porque cada renglon o columna en la matriz de covarianza de una composición cerrada suman cero dado que la varianzas son positivas, esto implica que algunas covarianzas son forzadas a tomar valores negativos.

### 1.3.1. Transformaciones log-ratio

Esta demostrado que trabajar con las partes de una composición en sí mismas nos conduce a ser subcomposicionalmente incoherentes y a no ser invariantes bajo escala, por eso para ser coherentes subcomposicionalmente e invariantes bajo escala se trabaja con sus *cocientes (ratios)*.

Pero estos ratios son números positivos y pueden tener valores bastante grandes o bastante pequeños, por lo que sus desviaciones estándar suelen ser tan grandes o inclusive más grandes que las medias, lo que indica un alto grado de sesgo en sus distribuciones. Además los ratios  $\frac{x_1}{x_2} \neq \frac{x_2}{x_1}$ , entonces dependen del orden de las partes no son invariantes bajo permutación.

Por estas razones los ratios se transforman *logarítmicamente*, convirtiendo valores estrictamente positivos en números reales que pueden ser negativos o positivos. Otra razón es para hacer que los datos sean más simétricos y reducir el efecto de los valores atípicos, además de que la transformación logarítmica convierte datos en escala de Razón en datos en escala de Intervalo y la mayoría de los métodos estadísticos asumen que los datos están en una escala de Intervalo.

Y en general con las transformaciones *log-ratio* la inversión  $\ln\left(\frac{x_1}{x_2}\right) = -\ln\left(\frac{x_2}{x_1}\right)$ , es decir, solo producen un cambio de signo dando así una simetría con respecto al ordenamiento de las partes. Por ello se introducen a continuación tres de las transformaciones tipo log-ratio más utilizadas para mapear los datos composicionales al espacio real y poder llevar a cabo su análisis. Observe que estas transformaciones satisfacen dos de los tres principios básicos de Aitchison, por lo que podremos llevar a cabo un buen análisis sin caer en los problemas antes mencionados.

Ya se menciona que el conjunto  $(\mathbb{S}^{p+1}, \oplus, \odot, < \cdot, \cdot >_A)$  conforma un Espacio Euclídeo, esto significa que se puede trasladar prácticamente cualquier cosa definida para vectores reales a los composicionales, ya que un espacio euclidiano es siempre equivalente al espacio real, esta equivalencia se logra a través de una Isometría, es decir, una

transformación del Simplex al Espacio Real que mantiene los ángulos y las distancias.

Aitchison en [Aitchison, 1982] introduce las transformaciones log-ratio aditiva(alr) y centrada(clr), aplicando el análisis estadístico clásico a las observaciones transformadas, utilizando la transformación alr para el modelado y la transformación clr para aquellas técnicas basadas en una métrica. Aunque no es explícito, la razón subyacente fue que la transformación alr no conserva las distancias, mientras que la transformación clr preserva las distancias pero conduce a una matriz de covarianza singular. Debido a estos inconvenientes fue necesario introducir la transformación log-ratio isométrica(ilr) cuyo término significa en realidad la asociación de composiciones con coordenadas en un sistema ortonormal general.

### Transformación log-ratio aditiva

**Definición 10.** La transformación log-ratio aditiva(alr) es una transformación 1 - 1 de  $\mathbf{w} = (w_1, w_2, \dots, w_{p+1}) \in \mathbb{S}^{p+1}$  a  $Y \in \mathbb{R}^p$  definida por

$$Y = alr(\mathbf{w}) = \left( \ln \frac{w_1}{w_{p+1}}, \log \frac{w_2}{w_{p+1}}, \dots, \log \frac{w_p}{w_{p+1}} \right) \quad (1.3.1)$$

Desafortunadamente la transformación alr es asimétrica respecto a las partes de la composición, debido a que la componente utilizada como denominador  $w_{p+1}$  cobra especial protagonismo. Por otro lado, con estas coordenadas no es posible usar el producto interno habitual y la distancia, además no cumple con la invarianza bajo permutación y no debería utilizarse, por ejemplo, para el análisis de conglomerados de lo contrario se corre el riesgo de tener diferentes agrupaciones dependiendo de la última variable en el conjunto de datos.

### Transformación log-ratio centrada

**Definición 11.** La transformación log-ratio centrada(clr) de una composición  $\mathbf{w} = (w_1, w_2, \dots, w_{p+1}) \in \mathbb{S}^{p+1}$  de  $(p + 1)$ -partes a  $Z \in \mathbb{R}^{p+1}$  se define como

$$Z = clr(\mathbf{w}) = \left( \log \frac{w_1}{g(\mathbf{w})}, \log \frac{w_2}{g(\mathbf{w})}, \dots, \log \frac{w_{D-1}}{g(\mathbf{w})} \right) \quad (1.3.2)$$

donde  $g(\mathbf{w})$  es la media geométrica de  $\mathbf{w}$ .

La transformación clr es simétrica e isométrica en las partes, pero la imagen de  $\mathbb{S}^{p+1}$  está restringida a un subespacio de  $\mathbb{R}^{p+1}$  y las matrices de covarianza y correlación de los datos clr transformados son singulares ( $det = 0$ ).

### Transformación log-ratio isométrica

En el 2003 Egozcue y sus colaboradores proponen la transformación log-ratio isométrica, la cual evita los problemas existentes en las transformaciones anteriores.

**Definición 12.** La transformación log-ratio isométrica(ilr) de una composición  $\mathbf{w} = (w_1, w_2, \dots, w_{p+1}) \in \mathbb{S}^{p+1}$  de  $(p + 1)$ -partes se define como

$$ilr(\mathbf{w}) = (\langle \mathbf{w}, e_1 \rangle_A, \langle \mathbf{w}, e_2 \rangle_A, \dots, \langle \mathbf{w}, e_p \rangle_A) \quad (1.3.3)$$

donde  $(e_1, e_2, \dots, e_p)$  es una base ortonormal del simplex.



Existen tantas *ilr* como bases ortonormales se pueden definir, Egozcue se baso en la matriz de Helmert para formar la base ortonormal. Esta transformación (*ilr*) es isométrica, la ventaja es que transforma los datos composicionales en coordenadas en un sistema ortogonal, es decir, se puede usar cualquier técnica estadística multivariante para su estudio.

La propiedad más importante de esta transformación es que reduce la geometría de Aitchison en  $\mathbb{S}^{p+1}$  a la geometría euclidiana ordinaria en  $\mathbb{R}^p$ . Por ejemplo, si  $\mathbf{w}, \mathbf{u} \in \mathbb{S}^{p+1}$ ,

$$d_A(\mathbf{w}, \mathbf{u}) = d_e(ilr(\mathbf{w}), ilr(\mathbf{u}))$$

donde  $d_e(\cdot, \cdot)$  es la distancia euclidiana.

Por este motivo en el presente trabajo de tesis se propone como parte fundamental de la metodología, utilizar la transformación *ilr* con la finalidad de transformar las composiciones en datos reales en  $\mathbb{R}^p$ , para llevar a cabo su modelación a través del modelo propuesto en el siguiente capítulo. Los resultados obtenidos mediante la inferencia de dicho modelo serán devueltos a su espacio original  $\mathbb{S}^{p+1}$  mediante la transformación inversa  $ilr^{-1}$ , para su correcta interpretación.

### ¿Por qué tantas transformaciones log-ratio?

Existen diversas opiniones respecto a cuál de ellas es mejor; Sin embargo, algunas de las razones por las que existen diversas transformaciones log-ratio son:

- Porque ninguna es perfecta. Las tres respetan la perturbación y la potencia, sin embargo la transformación *alr* no preserva el producto escalar.
- La transformación *clr* produce matrices de covarianza singulares y esto podría ser una fuente de problemas si el método estadístico utilizado necesita invertirla.
- La ventaja de *clr* es que representa un enlace uno a uno entre las partes originales y las transformadas, lo que parece ser útil para su interpretación.
- La transformación *ilr* con cualquier base ortonormal puede incluir muchas partes lo que hace prácticamente imposible llevar a cabo una interpretación.
- *ilr* es isométrica y sus valores transformados producen matrices de covarianza de rango completo por lo que pueden analizarse sus datos de forma adecuada.

A continuación, se muestran algunas propiedades de las tres transformaciones mencionadas.

Transformación log-ratio aditiva

$$alr(\mathbf{w} \oplus \mathbf{u}) = alr(\mathbf{w}) + alr(\mathbf{u}) \quad (1.3.4)$$

$$alr(\alpha \odot \mathbf{w}) = \alpha \odot alr(\mathbf{w}) \quad (1.3.5)$$

$$\langle \mathbf{w}, \mathbf{u} \rangle_A \neq alr(\mathbf{w}) \cdot alr^t(\mathbf{u}) \quad (1.3.6)$$

Transformación log-ratio centrada

$$clr(\mathbf{w} \oplus \mathbf{u}) = clr(\mathbf{w}) + clr(\mathbf{u}) \quad (1.3.7)$$

$$clr(\alpha \odot \mathbf{w}) = \alpha \odot clr(\mathbf{w}) \quad (1.3.8)$$

$$\langle \mathbf{w}, \mathbf{u} \rangle_A = \langle clr(\mathbf{w}), clr(\mathbf{u}) \rangle \quad (1.3.9)$$

Transformación log-ratio isométrica

$$ilr(\mathbf{w} \oplus \mathbf{u}) = ilr(\mathbf{w}) + ilr(\mathbf{u}) \quad (1.3.10)$$

$$ilr(\alpha \odot \mathbf{w}) = \alpha \odot ilr(\mathbf{w}) \quad (1.3.11)$$

$$\langle \mathbf{w}, \mathbf{u} \rangle_A = ilr(\mathbf{w}) \cdot ilr^t(\mathbf{u}) \quad (1.3.12)$$

$$\langle \mathbf{w}, \mathbf{u} \rangle_A = \langle ilr(\mathbf{w}), ilr(\mathbf{u}) \rangle \quad (1.3.13)$$

donde  $(\cdot)$  representa el producto en los reales,  $(\langle \cdot, \cdot \rangle)$  el producto punto en los reales y  $(^t)$  representa la transpuesta de ese vector.

## 1.4. Estadística Descriptiva

### 1.4.1. Descripción Gráfica

Tal y como ocurre con el espacio real, el simplex también tiene representaciones gráficas donde será posible visualizar a los datos composicionales y como se muestra en la Figura 1.1 estas representaciones del simplex están contenidas de manera natural en espacios reales.

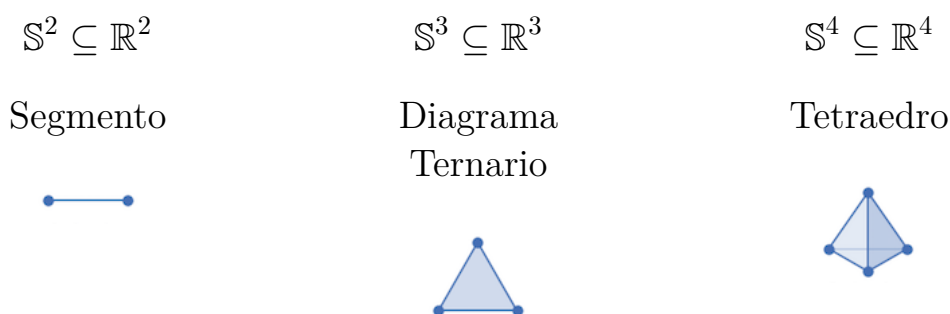


Figura 1.1: Representaciones gráficas del simplex.

Entonces los datos composicionales pueden ser representados gráficamente para llevar a cabo un primer análisis descriptivo de los mismos. La representación gráfica más utilizada son los diagramas ternarios, pero también existen otros tipos de representaciones como por ejemplo, visualizarlos como secuencias en diagramas de barras o como puntos en diagramas de dispersión convencionales, entre otros. A continuación se presentan los dos primeros, el lector puede consultar más representaciones en [Van den Boogaart and Tolosana-Delgado, 2013] y [Pawlowsky-Glahn et al., 2015].

#### Diagramas Ternarios

Los datos composicionales en  $\mathbb{S}^3$  suelen representarse mediante *diagramas ternarios*, triángulos equiláteros de altura  $k$ , con  $k$  la constante de clausura del Simplex. Existe

una correspondencia biunívoca entre los datos composiciones con 3 partes y los puntos de un diagrama ternario. Un dato composicional  $\mathbf{x} = (x_1, x_2, x_3)$  se corresponde con el punto que dista  $x_1, x_2$  y  $x_3$ , respectivamente, de los lados opuestos a los vértices 1, 2 y 3 como se observa en la Figura 1.2.

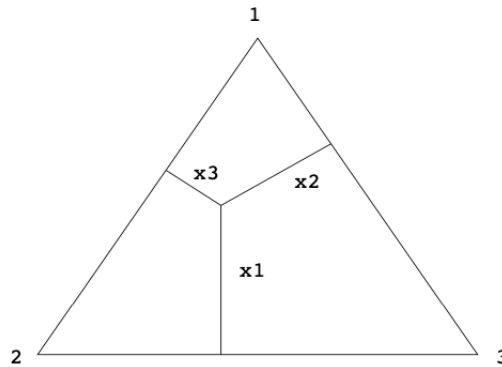


Figura 1.2: Representación de un dato composicional  $\mathbf{x} = (x_1, x_2, x_3)$  en el Símplex  $\mathbb{S}^3$ .

En el diagrama ternario cada uno de los vértices del triángulo representa una componente. Cada vértice equivale al 100 % del componente que le pertenece del dato composicional y este porcentaje decrece conforme el punto avanza hacia el lado opuesto. A manera de ejemplo a continuación se presentan tres diagramas ternarios correspondientes a tres conjuntos de datos composicionales.

En la Figura 1.3 se representa la composición mineral de 25 rocas tipo Hongite, en los minerales (A, B, C), este conjunto de datos composicionales fue obtenido de [Aitchison, 2003].

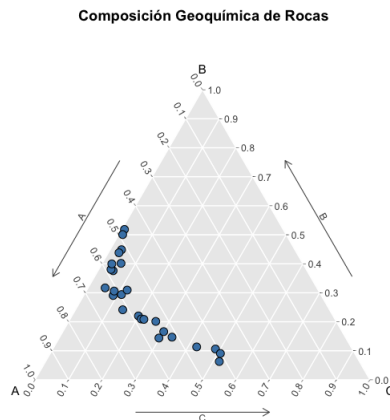


Figura 1.3: Diagrama Ternario de 25 rocas tipo Hongite.

En la Figura 1.4 se puede visualizar la subcomposición de los tres partidos electorales más fuertes del conjunto de datos utilizados para el cálculo del ejercicio estadístico del conteo rápido, realizado por el Instituto Nacional Electoral (INE) para la Gubernatura del estado de Puebla en 2019, dicho conjunto de datos composicionales conformado por 429 registros fue extraído de [Muestra de Puebla, 2019].

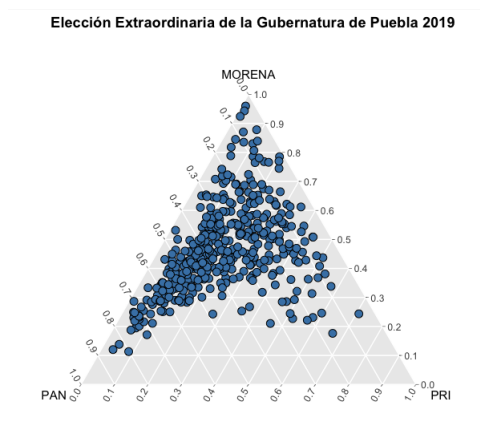


Figura 1.4: Diagrama Ternario de 429 datos electorales de la Gubernatura de Puebla 2019.

Se puede ver que existen diferentes comportamientos dependiendo de la naturaleza y el área de donde provengan, por lo tanto, no es de extrañarse que en algún campo se presenten patrones con características de agrupamiento.

En la Figura 1.5 se muestra un conjunto de 250 datos simulados a través de una mezcla de normales bivariadas y llevada al simplex mediante la transformación  $ilr^{-1}$ , es evidente que los datos fueron diseñados para poseer la característica de agrupamiento antes mencionada.

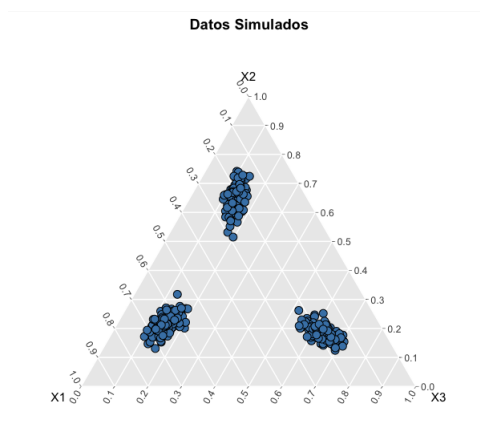


Figura 1.5: Diagrama Ternario de 250 datos simulados.

Los diagramas ternarios son de gran ayuda para observar de manera gráfica patrones de variabilidad que puedan existir en los datos composicionales, realizar pruebas de significancia, análisis en modelos de regresión, etc.

### Diagrama de barras

Un diagrama de barras es una representación de todas las partes de la composición. En un diagrama de barras, uno representa la cantidad de cada parte de un individuo como una barra dentro de un conjunto. Es decir, las barras se apilan con la altura correspondiente de cada una de las partes de la composición del individuo hasta añadir

el total de la composición. En la Figura 1.6 se muestra el diagrama de barras para un conjunto de datos composicionales de 5 tipos de leguminosas, conformadas por 5 componentes que son (Grasa, Sodio, Carbohidratos, Proteínas, Otros), dicho conjunto fue extraído de [Van den Boogaart and Tolosana-Delgado, 2013].

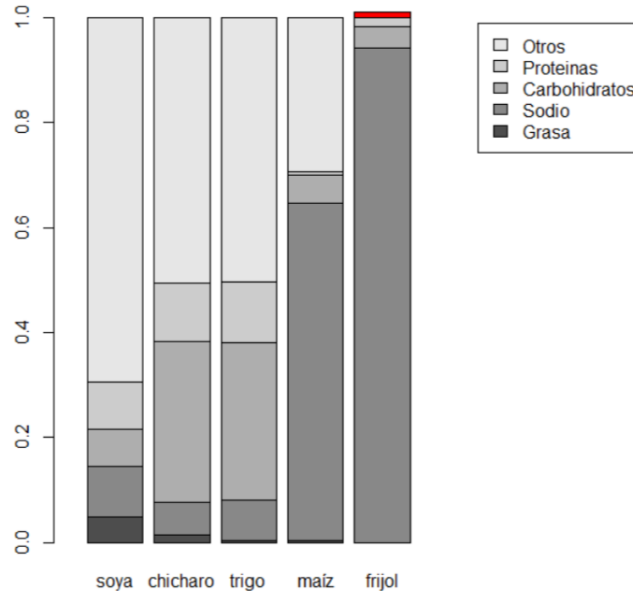


Figura 1.6: Diagrama de barras asociado a 5 tipos de leguminosas cuyas componentes son  $\mathbf{w} = (\text{Grasa}, \text{Sodio}, \text{Carbohidratos}, \text{Proteínas}, \text{Otros})$ .

### 1.4.2. Medidas Descriptivas

Los datos composicionales son realizaciones de vectores aleatorios cuyo soporte está en  $\mathbb{S}^{p+1}$ , por lo tanto, se plantea la necesidad de aplicar técnicas estadísticas adecuadas para el estudio e interpretación de los mismos contemplando su topología. A continuación se presentan algunas de las medidas estadísticas numéricas más útiles para la descripción de los datos composicionales.

Una medida de ubicación para datos composicionales se define a través de la media geométrica.

**Definición 13.** La media geométrica composicional o centro  $\bar{\mathbf{x}}$  de un conjunto de datos composicionales  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n \in \mathbb{S}^{p+1}$  se define como:

$$\bar{\mathbf{x}} = \mathcal{C} \left[ \exp \left( \frac{1}{n} \sum_{i=1}^n \ln(x_{ij}) \right) \right] = \mathcal{C} [g_1, g_2, \dots, g_{p+1}] \quad (1.4.1)$$

donde

$$g_j = \left( \prod_{i=1}^n x_{ij} \right)^{\frac{1}{n}}$$

con  $j = 1, 2, \dots, p+1$  y  $\mathcal{C}$  el operador clausura.

El centro de una composición aleatoria debe considerarse como el valor esperado o media de una composición. La estimación del centro puede hacerse en coordenadas

ilr, y las propiedades de su estimador corresponde a las del estimador de medias reales .

Por otro lado, como una medida global de dispersión se puede usar la *varianza total* definida como

$$\begin{aligned} \text{Vartot}[\mathbf{x}] &= \frac{1}{p+1} \sum_{i=1}^p \sum_{j=i+1}^{p+1} \text{Var} \left[ \ln \left( \frac{x_i}{x_j} \right) \right] \\ &= \sum_{j=1}^p \text{Var}[\text{ilr}(\mathbf{x})] \end{aligned} \quad (1.4.2)$$

La estimación del centro, de la varianza total y sus componentes, puede hacerse en coordenadas ilr y las propiedades de estos estimadores corresponderán a las de los estimadores de medias y varianzas reales ver más en [Pawłowsky-Glahn and Egozcue, 2001].

Una medida que explica la relación existente entre dos variables es la covarianza.

**Definición 14.** *La estructura de covarianza de una composición  $\mathbf{x}$  de  $(p+1)$ -partes es el conjunto*

$$\sigma_{ij,kl} = \text{Cov} \left( \log \left( \frac{x_i}{x_k} \right), \log \left( \frac{x_j}{x_l} \right) \right). \quad (1.4.3)$$

Cuando  $i, j, k, l$  toman solo dos valores, se obtiene la definición de varianza, es decir,

**Definición 15.** *Para dos elementos  $i$  y  $j$  de una composición  $\mathbf{x}$  de  $(p+1)$ -partes, la varianza log-cociente se define como*

$$\text{var} \left( \log \left( \frac{x_i}{x_j} \right) \right). \quad (1.4.4)$$

Existen diferentes formas útiles y equivalentes para describir los patrones de variabilidad, cada una de estas covarianzas posee diferentes propiedades como se menciona a continuación.

### Especificación de la estructura de covarianza

En esta sección se presentan algunas de las estructuras de covarianza como la matriz de covarianza log-cociente y la matriz de covarianza log-cociente centrada, así como algunas de sus características.

**Definición 16.** *(Matriz de covarianza log-cociente). Para una composición  $\mathbf{x}$  de  $(p+1)$ -partes la matriz de  $(p+1) \times (p+1)$*

$$\Sigma = \text{Cov} \left( \log \left( \frac{x_i}{x_k} \right), \log \left( \frac{x_j}{x_l} \right) \right), \quad (1.4.5)$$

*es denominada la matriz de covarianza log-cociente.*

Esta matriz de covarianza, generalmente es una matriz no singular y es asimétrica en el tratamiento de las partes.

Una manera de conservar la forma especificada anteriormente de la matriz de covarianza y al mismo tiempo obtener un comportamiento simétrico de todas las  $(p+1)$ -partes es reemplazando el denominador por la media geométrica  $g(\mathbf{x})$ .

**Definición 17.** (*Matriz de Covarianza log-cociente Centrada*). Para una composición  $\mathbf{x}$  de  $(p + 1)$ -partes la matriz  $(p + 1) \times (p + 1)$

$$\Gamma := Cov \left( \log \left( \frac{x_i}{g(\mathbf{x})} \right), \log \left( \frac{x_j}{g(\mathbf{x})} \right) \right), \quad (1.4.6)$$

con  $i, j = 1, 2, \dots, p + 1$  es llamada la matriz de covarianza log-cociente centrada.

La matriz de covarianza log-cociente centrada tiene una estructura de matriz de covarianza, generalmente es una matriz no singular y es simétrica en el tratamiento de las partes.

Como ya se ha mencionado, los datos composicionales aparecen frecuentemente en diversas disciplinas, por lo que resulta necesario disponer de herramientas adecuadas para su análisis estadístico. Así, además de contar con medidas descriptivas numéricas y métodos gráficos adecuados, se dispone de modelos de probabilidad en  $\mathbb{S}^{p+1}$  para describir patrones de variabilidad en el Símplex. A continuación se revisan algunos de estos modelos.

### 1.4.3. Distribuciones de Probabilidad en el Símplex

Por su propia naturaleza las componentes de una composición toman sus valores en el intervalo  $[0, 1]$ . Esta naturaleza hace evidente que las distribuciones multivariantes tradicionales más usuales, como la distribución normal, pueden producir resultados erróneos si son aplicadas directamente a los datos, por ello a continuación damos dos distribuciones cuyo soporte cumple con la naturaleza composicional [Martín-Fernández et al., 2004].

#### La Distribución de Dirichlet

La *Distribución de Dirichlet*,  $Dir(\boldsymbol{\alpha})$  es una familia de distribuciones de probabilidad multivariable, continua y parametrizada por un vector  $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_{p+1})$  real de términos positivos. Esta distribución es la generalización multivariable de la distribución beta. La distribución de Dirichlet de orden  $p \geq 1$  con parámetros  $\alpha_1, \alpha_2, \dots, \alpha_{p+1} > 0$  tiene una función de densidad de probabilidad en el espacio euclidiano  $\mathbb{R}^p$  dada por:

$$f(x_1, x_2, \dots, x_{p+1} | \alpha_1, \alpha_2, \dots, \alpha_{p+1}) = \frac{1}{B(\boldsymbol{\alpha})} \prod_{i=1}^{p+1} x_i^{\alpha_i - 1} \quad (1.4.7)$$

Recordando que la distribución Beta es una distribución de probabilidad continua, cuya función de densidad tiene soporte en  $x \in [0, 1]$ , de esta forma la distribución Dirichlet toma el valor dado por la ecuación 1.4.7 en el Símplex abierto  $(p + 1)$ -dimensional definido por:

$$\begin{aligned} x_1, x_2, \dots, x_p &> 0 \\ x_1 + x_2 + \dots + x_p &< 1 \\ x_{p+1} = 1 - x_1 - x_2 - \dots - x_p &> 0 \end{aligned}$$

y cero en otro caso.

La constante de normalización en la ecuación 1.4.7 es la función Beta multinomial  $B(\boldsymbol{\alpha})$ , la cual se puede expresar en términos de la función Gamma. Es decir,

$$B(\boldsymbol{\alpha}) = \frac{\prod_{i=1}^{p+1} \Gamma(\alpha_i)}{\Gamma(\sum_{i=1}^k \alpha_i)}$$

con  $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_{p+1})$ .

Desafortunadamente, esta familia paramétrica no es adecuada para la descripción de la variabilidad de datos composicionales, principalmente cuando en los datos composicionales se tienen patrones cóncavos, debido a que los contornos de isoprobabilidad de  $Dir(\boldsymbol{\alpha})$  son convexos. Además, la clase de Dirichlet no soporta un grado suficiente de dependencia composicional.

### La Distribución Logística Normal

En busca de alternativas a la clase Dirichlet, McAlister en [McAlister, 1879] se percató de que los datos composicionales adoptaban patrones similares a los de una normal  $N(\boldsymbol{\mu}, \sigma^2)$  mediante transformaciones tipo log-cociente, en particular con la transformación  $ilr$  se puede obtener la *Distribución Logística Normal* (o *Distribución Normal en el Simplex*).

Una composición aleatoria  $\mathbf{X}$  tiene una distribución Logística Normal con un vector de medias  $\boldsymbol{\mu}$  y matriz de precisión  $\boldsymbol{\lambda}$ , denotada como  $\mathcal{N}^{p+1}(\boldsymbol{\mu}, \boldsymbol{\lambda})$ , si al proyectar en cualquier dirección arbitraria  $u$  con el producto escalar de Aitchison se obtiene una variable aleatoria con una distribución normal univariada, con vector de medias  $\langle \boldsymbol{\mu}, u \rangle_A$  y precisión  $clr(u) \cdot \boldsymbol{\lambda} \cdot clr^t(u)$ . En particular, si se toma una base ortonormal en el Simplex,  $\mathbf{V}$ , entonces las coordenadas  $ilr(\mathbf{X})$  de la composición aleatoria tienen una distribución normal multivariada, es decir, la densidad conjunta es

$$N_p(\mathbf{x} \mid \boldsymbol{\mu}_V, \boldsymbol{\lambda}_V) = |\boldsymbol{\lambda}_V|^{\frac{1}{2}} (2\pi)^{-\frac{p}{2}} \exp \left\{ -\frac{1}{2} (ilr(\mathbf{x}) - \boldsymbol{\mu}_V)^t \boldsymbol{\lambda}_V (ilr(\mathbf{x}) - \boldsymbol{\mu}_V) \right\}$$

con  $\boldsymbol{\mu}_V$  y  $\boldsymbol{\lambda}_V$ , respectivamente, sus vectores de medias y matriz de precisión.

## 1.5. Estadística Bayesiana

La metodología propuesta en este trabajo de tesis nos conducirá a realizar inferencias sobre un modelo no-parámétrico, mismo que será propuesto más adelante, dicha inferencia se llevará cabo desde el enfoque Bayesiano de la estadística. Por este motivo a continuación se presentan algunos de los conceptos más importantes de la estadística Bayesiana.

En la estadística existen dos tipos de enfoques, el enfoque *clásico* o *frecuentista* y el enfoque *Bayesiano*. Ambos enfoques se forma a partir de los conceptos de probabilidad pero el enfoque clásico o frecuentista se centra principalmente en el cálculo de probabilidades y lo contrastes de hipótesis. En este enfoque los parámetros permanecen estáticos durante todo el estudio, por lo que, únicamente se utiliza la información obtenida en el ensayo del experimento estudiado. Esto último presenta problemas cuando



se dispone poca información muestral puesto que muchos de sus métodos se apoyan en resultados asintóticos, como la *Ley de los Grandes Números* y el *Teorema Central del Límite*.

Por otro lado, el enfoque Bayesiano está basado en el *Teorema de Bayes* y se diferencia del enfoque frecuentista básicamente por la posibilidad de poder incorporar información externa al estudio que se está realizando. Es decir, este enfoque hace posible la incorporación de hipótesis científicas que el experimentador o la literatura tengan acerca del estudio realizado.

El enfoque Bayesiano contiene como casos particulares muchos de los procedimientos frecuentistas, resolviendo muchas de las dificultades que enfrentan los métodos clásicos. Los métodos Bayesianos se derivan de un sistema axiomático, por lo tanto proporcionan una metodología general y ofrecen además un paradigma completo, tanto para la inferencia como para la toma de decisiones bajo escenarios donde la incertidumbre está presente. La aparición y el desarrollo a profundidad de este enfoque se puede consultar en [Bernardo and Smith, 2009], [Hoff, 2009], [Rossi et al., 2012], [Lynch, 2007], [Correa Morales and Barrera Causil, 2018], [Mendoza and Regueiro, 2011], entre otros.

El *Teorema de Bayes* es la obra más conocida del matemático británico Thomas Bayes, su trabajo publicado de forma póstuma en 1763 impactó de manera importante el modo en que se hace inferencia estadística. A continuación se presenta el teorema para el caso de eventos:

**Teorema 1.** Sea  $(\Omega, \mathcal{A}, \mathbb{P})$  un espacio de probabilidad. Sea  $\{A_1, A_2, \dots, A_n\}$  una partición de  $\Omega$ , tales que  $\mathbb{P}(A_i) > 0$  para cada  $i = \{1, 2, \dots, n\}$ . Entonces para cualquier  $B \in \Omega$  tal que  $\mathbb{P}(B) > 0$  se tiene que:

$$\begin{aligned} \mathbb{P}(A_i | B) &= \frac{\mathbb{P}(B | A_i)\mathbb{P}(A_i)}{\mathbb{P}(B)} \\ &= \frac{\mathbb{P}(B | A_i)\mathbb{P}(A_i)}{\sum_{k=1}^n \mathbb{P}(B | A_k)\mathbb{P}(A_k)} \end{aligned} \tag{1.5.1}$$

esto para cada  $i = \{1, 2, \dots, n\}$ .

De manera general, sea  $f(x | \theta)$  un modelo probabilístico paramétrico donde la función  $f(\cdot | \theta)$  puede ser la función masa de probabilidad (o de densidad) de una variable (o vector) aleatoria discreta (o continua), se desea llevar a cabo inferencias para el parámetro asociado  $\theta$  sobre el espacio paramétrico  $\Theta$ , en otras palabras se desea hacer inferencia sobre la familia paramétrica asociada que se define como:

$$\mathcal{F} = \{f(x | \theta) : \theta \in \Theta\}$$

Entonces bajo este enfoque la incertidumbre que se tiene sobre  $\theta$  se puede describir a través de una medida de probabilidad  $\pi(\theta)$  llamada distribución a “*priori*” o “*inicial*”, la cual representa la información externa o subjetiva sobre dicho parámetro.

Es importante aclarar que los parámetros no son variables aleatorias pero desde el enfoque Bayesiano se les trata así para calcular la incertidumbre de su valor. Dicha incertidumbre se va actualizando conforme a la información muestral disponible

mediante el Teorema de Bayes, obteniendo una distribución  $\pi(\theta | x)$  llamada a “*posteriori*” o “*final*”.

Así para una muestra aleatoria de datos  $\mathbf{x} = \{x_1, x_2, \dots, x_n\}$  se tiene que:

$$\begin{aligned}\pi(\theta | \mathbf{x}) &= \frac{f(\mathbf{x} | \theta)\pi(\theta)}{f(\mathbf{x})} \\ &= \frac{f(\mathbf{x} | \theta)\pi(\theta)}{\int_{\Theta} f(\mathbf{x} | \theta)\pi(\theta)d\theta}\end{aligned}\tag{1.5.2}$$

El cual también se puede escribir como:

$$\pi(\theta | \mathbf{x}) \propto f(\mathbf{x} | \theta)\pi(\theta)\tag{1.5.3}$$

dado que  $f(\mathbf{x})$  es una constante con respecto a  $\theta$  y  $f(\mathbf{x} | \theta)$  es la probabilidad conjunta de la muestra condicionada a  $\theta$  llamada también *verosimilitud*. Así, en el enfoque de la estadística Bayesiana un modelo paramétrico queda especificado por:

$$\{f(\mathbf{x} | \theta), \pi(\theta)\}$$

A  $f(\mathbf{x})$  se le llama distribución predictiva y juega un papel muy importante en las predicciones del fenómeno real, es decir, en el comportamiento de observaciones futuras. La distribución predictiva que describe el conocimiento acerca de una observación futura “ $x$ ” unicamente con la información contenida en  $\pi(\theta)$  se le denomina distribución predictiva a priori o inicial y proviene de la marginalización de la distribución conjunta para el vector aleatorio  $(X, \Theta)$  inducida por el modelo  $f(x | \theta)$  y la distribución inicial  $\pi(\theta)$  a través del concepto de probabilidad condicional.

$$f(x, \theta) = f(x | \theta)\pi(\theta)\tag{1.5.4}$$

Por lo tanto, al marginalizar esta distribución de probabilidad conjunta se obtiene:

$$\begin{aligned}f(x) &= \int_{\Theta} f(x, \theta)d\theta \\ &= \int_{\Theta} f(x | \theta)\pi(\theta)d\theta\end{aligned}\tag{1.5.5}$$

Una vez obtenida la muestra, el modelo  $f(x | \theta)$  y la distribución final  $f(\theta | \mathbf{x})$  inducen nuevamente una distribución conjunta para el vector aleatorio  $(X, \Theta)$  condicional a los valores observados  $\mathbf{x} = \{x_1, x_2, \dots, x_n\}$ :

$$\begin{aligned}f(x, \theta | \mathbf{x}) &= \frac{f(x, \theta, \mathbf{x})}{f(\mathbf{x})} \\ &= \frac{f(x | \theta, \mathbf{x})f(\theta, \mathbf{x})}{f(\mathbf{x})} \\ &= f(x | \theta, \mathbf{x})f(\theta | \mathbf{x}) \\ &= f(x | \theta)\pi(\theta | \mathbf{x})\end{aligned}\tag{1.5.6}$$

La última igualdad se da puesto que  $f(x | \theta, \mathbf{x}) = f(x | \theta)$  por la independencia condicional entre  $x$  y  $\mathbf{x}$  dado  $\theta$ . Así, marginalizando la distribución conjunta obtenemos:

$$\begin{aligned} f(x | \mathbf{x}) &= \int_{\Theta} f(x, \theta | \mathbf{x}) d\theta \\ &= \int_{\Theta} f(x | \theta) \pi(\theta | \mathbf{x}) d\theta \end{aligned} \tag{1.5.7}$$

la cual se denomina distribución predictiva a “posteriori” o “final” y describe el conocimiento acerca de una observación futura  $x$  basado tanto en la información contenida en la distribución inicial  $\pi(\theta)$  como en la información muestral  $\mathbf{x}$ .

## 1.6. Métodos de Simulación

Para llevar a cabo la implementación computacional de las inferencias para el modelo que se propondrá más adelante, se utilizarán métodos de simulación, por ello a continuación se presentan los puntos más esenciales de la simulación estocástica así como los métodos utilizados.

En el contexto estadístico, entendemos por simulación a la técnica de muestreo estadístico controlado, que se utiliza conjuntamente con un modelo, para obtener respuestas aproximadas a preguntas que surgen en problemas complejos de tipo probabilístico. La simulación estocástica es de gran utilidad cuando no tenemos expresiones cerradas para calcular medidas de incertidumbre y permite obtener estimaciones con menores supuestos que los métodos analíticos, a cambio de un trabajo computacional más intenso [Diharce, 2008].

A continuación se presentan los puntos más esenciales de como funciona la simulación estocástica:

- Recolectar y organizar información del problema
- Plantear un modelo matemático con sus supuestos estocásticos
- Implementación computacional
- Validación del código
- Análisis de resultados
- Validación de la simulación

Los métodos de simulación a utilizar son los métodos *MCMC* (*Markov Chain Monte Carlo*), estos ofrecen la opción de muestrear densidades multivariadas que no son fáciles de simular, a menudo “rompiendo” estas densidades en densidades univariadas o multivariadas más manejables. Dada una muestra de datos los métodos MCMC permiten muestrear de la distribución final  $\pi(\theta | \mathbf{x})$  moviéndose a través de todo el soporte  $\Theta$ , permitiendo así aproximar esencialmente cualquier característica de la distribución que la generó.

Entonces los métodos MCMC solucionan muchos de los problemas en la inferencia Bayesiana, a través de una aproximación indirecta de la distribución final en la que se requiere simular cadenas de Markov, por ello antes de presentar directamente los métodos utilizados en este trabajo de tesis se da una breve introducción a las cadenas de Markov.

### 1.6.1. Breve introducción a las Cadenas de Markov

Una cadena de Markov es un tipo particular de un proceso estocástico, el cual trata de la caracterización de series de variables aleatorias. El estudio se basa principalmente en el comportamiento dinámico de la cadena. Un proceso estocástico se define como una sucesión de variables aleatorias  $\{X_t : t \in T\}$  para algún conjunto  $T$ . Al conjunto de todos los posibles valores se le denomina *espacio de estados* y  $T$  es el *conjunto de índices*.

**Definición 18.** Una cadena de Markov es un proceso estocástico  $\{X_t\}$  tal que para  $t \in \mathbb{N}$  y para cualquier sucesión de estados  $x_0, x_1, x_2, \dots, x_t, x$  se tiene que:

$$P(X_{t+1} = x | X_t = x_t, X_{t-1} = x_{t-1}, \dots, X_0 = x_0) = P(X_{t+1} = x | X_t = x_t)$$

A la probabilidad condicional  $P(X_{t+1} = x_j | X_t = x_i)$  se le denomina *probabilidad de transición* al tiempo  $t$  del estado  $x_i$  al  $x_j$ . Como simplificación de notación se denotará a los estados  $x_i$  y  $x_j$  simplemente como estados  $i$  y  $j$ .

Para comprender mejor el funcionamiento de las Cadenas de Markov en los métodos de Monte Carlo se considerarán cadenas homogéneas, es decir, aquellas cuya probabilidad de transición de un estado a otro no depende del tiempo en el que se encuentre. Es decir

$$P(X_{t+1} = j | X_t = i) = P(X_1 = j | X_0 = i) = p_{ij}$$

y por lo tanto, para una cadena homogénea se tiene que la probabilidad de transición de  $n$  pasos entre dos estados  $i$  y  $j$  está dada por:

$$p_{ij}^n = P(X_{k+n} = j | X_k = i) = P(X_n = j | X_0 = i)$$

Algunas propiedades relevantes de las Cadenas de Markov son las siguientes.

**Definición 19.** Una cadena de Markov  $X_t$  es irreducible si para cualesquiera dos estados  $i$  y  $j$  existe un  $n_{ij} > 0$  tal que

$$p_{ij}^{n_{ij}}(t) > 0$$

En otras palabras esta propiedad describe la posibilidad de pasar de cualquier estado actual a otro que pertenezca al espacio de estados. A esto también se le conoce como comunicación entre estados y es deseable que exista comunicación entre todos los posibles estados de la cadena.

Por otro lado se tiene la propiedad de recurrencia, la cual indica que la cadena regresará al estado actual después de un cierto tiempo  $t$ .

**Definición 20.** La cadena irreducible  $X_t$  será recurrente si para todo estado  $i$ ,

$$P(\min\{t > 0 : X_t = i | X_0 = i\} < \infty) = 1$$

La siguiente propiedad es una condición necesaria que garantiza que una cadena irreducible tenga una distribución estacionaria.

**Definición 21.** Dada  $X_t$  irreducible y recurrente, se denominará positiva recurrente si para todo estado  $i$  se cumple que

$$E[\min\{t > 0 : X_t = i | X_0 = i\}] < \infty$$

La cadena  $X_t$  es estacionaria si la distribución de  $X_t$  es la misma para todo  $t$ .

**Definición 22.** Si  $T$  es el espacio de estados de la cadena  $X_t$ , se dice que

**Definición 23.** Se dice que una distribución  $\pi$  es una distribución estacionaria de una cadena de Markov con probabilidad de transición  $p_{ij}$  si

$$\sum_{i=0}^{\infty} \pi(i)p_{ij} = \pi(j)$$

donde  $\pi(i)$  y  $\pi(j)$  es la distribución en el estado  $i$  y  $j$ , respectivamente.

Si bien para una cadena irreducible y positiva recurrente existe una distribución estacionaria, es importante garantizar que no sea posible regresar al estado actual, a través de la transición a estados distintos, únicamente en múltiplos de un entero mayor que 1. De existir esta posibilidad de regresar al estado actual de esa forma, la cadena podría entrar en un ciclo que no le permitiría alcanzar una convergencia lo que haría imposible obtener resultados relacionados con la distribución objetivo. Dicha característica se define como *aperiódica*.

**Definición 24.** Una cadena de Markov  $X_t$  es aperiódica si para todo estado  $i$  e índice  $t$  se cumple que

$$m.c.d.\{n > 0 : p_{ii}^n > 0\} = 1,$$

donde *m.c.d.* es el máximo común divisor.

Retomando el contexto de la inferencia bayesiana, las distribuciones estacionarias resultan de gran utilidad ya que serán las distribuciones finales de los parámetros de interés. De esta forma, con las propiedades establecidas para una cadena de Markov, se puede enunciar el siguiente teorema importante y de gran utilidad en el desarrollo práctico de la estadística bayesiana.

**Proposición 1.6.1.** Si  $X$  es una cadena irreducible, aperiódica y positiva recurrente, entonces posee una única distribución estacionaria  $\pi(\cdot)$ . En ese caso se dirá que  $X$  es irreducible y ergódica, y cumple con:

1.  $p_{ij}^n \rightarrow \pi(j)$ , cuando  $n \rightarrow \infty$  para toda  $i$  y  $j$ .
2. Dada  $g$  una función real y  $E_{\pi}\{|g(X)|\} < \infty$ , entonces

$$P\left(\frac{\sum_{t=1}^n g(X_t)}{n} \rightarrow E_{\pi}[g(X)]\right) = 1$$

donde  $E_{\pi}[g(X)] = \sum_i g(X_i)\pi(X_i)$ , la esperanza de  $g(X)$  con respecto a  $\pi(\cdot)$ .

### 1.6.2. Métodos Monte Carlo vía Cadenas de Markov

**Definición 25.** *Un Método de Monte Carlo vía una Cadena de Markov para la simulación de una distribución  $f$  es cualquier método que produce una cadena de Markov ergódica  $(X^{(t)})$  cuya distribución estacionaria es  $f$ .*

Dada esta definición entonces *Monte Carlo* se refiere al proceso de simulación estocástica y *Cadenas de Markov* se refiere al proceso de muestrear un nuevo valor en la cadena, dado el valor previo y se pueden proponer una infinidad de implementaciones prácticas. Algunos métodos MCMC son el Metrópolis Hastings, el Slice Sampling, el Gibbs Sampling, entre otros; De hecho, se cuenta con versiones de Slice dentro de Gibbs dentro de Metrópolis Hastings, es decir, se cuenta con diferentes versiones de acuerdo a las necesidades de cada estudio o análisis. A continuación se describen dos de ellos, los cuales serán importantes en el desarrollo de esta tesis.

#### Gibbs Sampling

El Gibbs Sampling fue propuesto por Geman y Geman en [Geman and Geman, 1984] dentro del contexto del procesamiento de imágenes. El nombre se debe a que Geman y Geman tenían como objetivo el interés de estudiar la distribución de Gibbs. Este trabajo pasó prácticamente inadvertido para la comunidad estadística hasta que Gelfand y Smith en [Gelfand et al., 1990] señalaron que el Gibbs Sampling podría servir para estudiar otras distribuciones como las distribuciones finales en estadística bayesiana.

El algoritmo del Gibbs Sampling se puede entender de la siguiente manera: Supóngase que se tiene un vector de parámetros  $d$ -dimensional  $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_d)$ , las distribuciones condicionales completas  $f_i(\theta_i | \boldsymbol{\theta}_{-i}, \mathbf{x})$ ,  $i = 1, \dots, d$ , son totalmente conocidas y es posible simular de ellas. Entonces, el Gibbs Sampling permite obtener una muestra de la distribución final (conjunta) de los parámetros dados los datos,  $\pi(\boldsymbol{\theta} | \mathbf{x})$ , a partir de las  $f_i(\theta_i | \boldsymbol{\theta}_{-i}, \mathbf{x})$ . Esto se logra creando una cadena de Markov multivariada  $\boldsymbol{\theta}^{(t)} = (\theta_1^{(t)}, \theta_2^{(t)}, \dots, \theta_d^{(t)})$  mediante la simulación secuencial de las condicionales completas para cada componente.

Así, el Gibbs Sampling genera la cadena de Markov  $\boldsymbol{\theta}^{(t)} = (\theta_1^{(t)}, \theta_2^{(t)}, \dots, \theta_d^{(t)})$  y con ello se obtiene una muestra de la densidad conjunta  $\pi(\boldsymbol{\theta} | \mathbf{x})$ , lo anterior puede describirse de la siguiente manera:

1. Asignar adecuadamente un valor inicial  $\boldsymbol{\theta}^{(0)} = (\theta_1^{(0)}, \theta_2^{(0)}, \dots, \theta_d^{(0)})$  y hacer el contador  $t = 0$ .
2. Generar  $\boldsymbol{\theta}^{(t+1)} = (\theta_1^{(t+1)}, \theta_2^{(t+1)}, \dots, \theta_d^{(t+1)})$  a partir de  $\boldsymbol{\theta}^{(t)} = (\theta_1^{(t)}, \theta_2^{(t)}, \dots, \theta_d^{(t)})$  de la siguiente manera:
  - a) Generar  $\theta_1^{(t+1)}$  de  $f_1(\theta_1 | \theta_2^{(t)}, \dots, \theta_d^{(t)})$
  - b) Generar  $\theta_2^{(t+1)}$  de  $f_2(\theta_2 | \theta_1^{(t+1)}, \theta_3^{(t)}, \dots, \theta_d^{(t)})$
  - c) Generar  $\theta_3^{(t+1)}$  de  $f_3(\theta_3 | \theta_1^{(t+1)}, \theta_2^{(t+1)}, \theta_4^{(t)}, \dots, \theta_d^{(t)})$
  - ⋮
  - d) Generar  $\theta_d^{(t+1)}$  de  $f_d(\theta_d | \theta_1^{(t+1)}, \theta_2^{(t+1)}, \dots, \theta_{d-1}^{(t+1)})$

3. Actualizar el contador a  $t = t + 1$  y regresar al paso 2, repetir este procedimiento hasta lograr la convergencia.

Si el procedimiento se repite para cada uno de los valores iniciales  $\theta_1^0, \theta_2^0, \dots, \theta_N^0$ , entonces al final de un número suficientemente grande de iteraciones “ $T$ ”, los valores  $\theta_1^T, \theta_2^T, \dots, \theta_N^T$  se pueden considerar como una muestra de tamaño “ $N$ ” de  $\pi(\boldsymbol{\theta} | \mathbf{x})$ . Lo anterior requeriría la generación de  $N$  cadenas por eso alternativamente se puede generar una sola cadena y a partir de “ $T$ ”, tomar los valores  $\theta^{T+k}, \theta^{T+2k}, \dots, \theta^{T+Nk}$ , donde “ $k$ ” se elije de manera que las correlaciones entre las observaciones sean independientes.

El Gibbs Sampling utiliza densidades de una dimensión para poder simular de la densidad multivariada, las cuales son relativamente fáciles de simular y justamente es por esa razón que este muestreo es una técnica ampliamente utilizada para obtener muestras de distribuciones finales multivariadas. La característica principal es que la simulación de un parámetro utiliza los valores obtenidos en los pasos anteriores al ciclo, para los parámetros que han sido simulados dentro del ciclo y mantiene los valores actuales para aquellos que aún no han sido generados.

### Slice Sampling

Perteneciente a la gran variedad de métodos MCMC el Slice Sampling se basa en la observación de que para muestrear una variable aleatoria se puede muestrear uniformemente de la región bajo el gráfico de su función de densidad.

Entonces el algoritmo Slice Sampling en su forma más simple obtiene una muestra de manera uniforme por debajo de la curva de la densidad conjunta  $\pi(\boldsymbol{\theta} | \mathbf{x})$ , de la siguiente manera:

1. Se elije un valor  $\theta_0$  para el cual  $\pi(\theta_0 | \mathbf{x}) > 0$  y se hace el contador  $t = 1$ .
2. Generar un valor  $\mathbf{u}$  de manera uniforme en el intervalo  $[0, \pi(\theta_{t-1} | \mathbf{x})]$ . Esto define una región de  $\boldsymbol{\theta}$  donde  $\pi(\boldsymbol{\theta} | \mathbf{x}) > \mathbf{u}$ , es decir, una región donde la densidad de probabilidad es al menos  $\mathbf{u}$ .
3. Generar un valor  $\theta_t$  de manera uniforme en la región delimitada en el paso anterior, actualizar el contador  $t = t+1$  y regresar al paso 2, repetir este procedimiento hasta lograr la convergencia.

Entonces el Slice Sampling permite obtener muestras utilizando lo que se conoce como *data augmentation* introduciendo estas variables latentes uniformes en el modelo propuesto como se verá más adelante, ver más en [Neal, 2003].

### Evaluación de Convergencia

Un elemento  $\boldsymbol{\theta}$  generado de una cadena de Markov es considerado una observación de la distribución estacionaria teóricamente, cuando el número de iteraciones tiende a infinito. En la práctica no es posible cumplir con esa condición. En su lugar se toma el valor de la cadena después de una cantidad suficientemente grande de iteraciones, de tal forma que dicho valor se aproxima a una observación de la distribución objetivo. A partir de ese valor, los siguientes estados de la cadena continuarán siendo considerados provenientes de la distribución estacionaria. Otra complicación práctica es que,

para obtener una muestra de la distribución estacionaria, las observaciones deben ser independientes entre sí y, los estados de una cadena de Markov son por su naturaleza dependientes entre sí. Estos dos aspectos forman parte fundamental del estudio de la convergencia de cualquier cadena simulada.

Existen dos formas para analizar la convergencia, uno es el enfoque teórico, con el cual se busca medir distancias y poder establecer cotas sobre las funciones de distribución generadas por una cadena. En particular, se puede estudiar la variación total de la distancia entre la distribución de la cadena en la iteración  $t$  y la distribución límite. También es posible analizar aspectos particulares derivados de la estructura probabilística de las cadenas. Si bien el enfoque teórico es importante para el estudio de la convergencia, en la práctica estos resultados pueden ser complicados de aplicar.

Un segundo enfoque es a través de métodos gráficos. Este enfoque analiza las propiedades de las observaciones obtenidas de una cadena para establecer la aproximación de ésta a la distribución límite. Uno de los principales inconvenientes de esta perspectiva es que no es posible garantizar la convergencia. A pesar de esta complicación el análisis de la convergencia desde la perspectiva gráfica es de gran utilidad en la práctica. Una referencia para el lector es el libro de [Gamerman and Lopes, 2006] en donde se detallan los distintos métodos y enfoques para el análisis de convergencia.

Un aspecto importante de los métodos MCMC es el hecho de establecer la cantidad de ciclos o iteraciones que se deben realizar para que la cadena generada se aproxime a la distribución estacionaria. A esta cantidad de iteraciones del algoritmo se le conoce como *período de calentamiento*. La obtención del periodo de calentamiento y por lo tanto la convergencia de la cadena, puede obtenerse a través de algunas verificaciones de convergencia basadas en técnicas gráficas propuestas en [Gelfand et al., 1990]. Uno de estos métodos consiste en realizar  $n$  simulaciones paralelas de la cadena que se desea estudiar y graficarlas durante un mismo número de iteraciones, posteriormente graficar los histogramas de los  $n$  valores obtenidos en la  $k$ -ésima iteración. Luego se grafican los histogramas de las  $m$  iteraciones posteriores a  $k$ . Generalmente, es necesario utilizar valores para  $m$  entre 10 y 50. Después de esto, se acepta la convergencia de la cadena si no existen diferencias notables entre los histogramas obtenidos. De esta forma el valor de  $k$  será considerado como el periodo de calentamiento.

Otra técnica gráfica requiere la simulación de una única cadena. En este caso si dicha cadena a partir de una iteración  $k$  presenta, cualitativamente, el mismo comportamiento; entonces, esto es un indicador de convergencia y por lo tanto  $k$  representa el periodo de calentamiento. Otro enfoque utiliza los promedios ergódicos para los cuales el periodo de calentamiento se establece como aquella iteración a partir de la cual los promedios ergódicos presentan un comportamiento aproximadamente asintótico.

Las conclusiones sobre la convergencia de la cadena, realizadas utilizando estos métodos gráficos, pueden ser reafirmadas si se realiza dicho análisis utilizando distintos valores iniciales para la cadena y seleccionando el mayor valor de iteraciones requeridas para poder afirmar que la cadena ha convergido. Sin embargo, debe resaltarse que las técnicas gráficas de análisis deben usarse con precaución puesto que la convergencia puede no ser notada o puede determinarse de forma equivocada dependiendo de la



---

escala que se utilice para realizar las gráficas. Además existen diversas cadenas que pueden exhibir todos los indicadores de convergencia sin haberla alcanzado realmente.



# Capítulo 2

## El Modelo Propuesto

En este capítulo se describe el modelo propuesto para la modelación de datos composicionales en el Símplex unitario  $(p + 1)$ -dimensional. La propuesta para analizar variables composicionales se basa en la definición de un modelo de mezclas infinitas de distribuciones normales univariadas y multivariadas, por lo tanto, se comienza describiendo el modelo normal univariado seguido del multivariado. Luego, se introducen los modelos de mezclas infinitas para ambos casos y los procesos Dirichlet en su representación Stick-Breaking. Por último, se termina representando el modelo propuesto a través del modelo de mezclas de procesos Dirichlet.

### 2.1. El modelo Normal Univariado

Una variable aleatoria continua  $x \in \mathbb{R}$ , tiene una distribución *normal univariada*, con parámetros  $\mu$  y  $\lambda$  (donde  $\mu \in \mathbb{R}$  y  $\lambda > 0$ ), si su densidad de probabilidad  $N(x | \mu, \lambda)$  es:

$$N(x | \mu, \lambda) = \lambda^{\frac{1}{2}} (2\pi)^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} \lambda (x - \mu)^2 \right\} \quad (2.1.1)$$

Con esperanza:

$$\mathbb{E}[x] = \mu \quad (2.1.2)$$

Y con varianza:

$$\mathbb{V}[x] = \lambda^{-1} \quad (2.1.3)$$

La distribución normal fue introducida por Carl Friedrich Gauss a principios del siglo XIX, al estudiar los errores de medida en los movimientos de los cuerpos celestes. Es una distribución que tiene forma de campana, es simétrica con respecto a  $\mu$  y puede tomar valores entre menos infinito y más infinito, ver más en [Reyes, 2012].

Algunas de las propiedades más importantes de esta distribución son las siguientes:

- Al ser simétrica la moda, la mediana y la media coinciden y son iguales a  $\mu$ .
- En el intervalo  $[\mu - 2\lambda, \mu + 2\lambda]$  se encuentra el 95.5 % de la distribución.
- Una transformación lineal de una variable aleatoria normal también se distribuye normalmente.
- Cualquier combinación lineal de variables aleatorias independientes e idénticamente distribuidas en forma normal, tiene una distribución normal.

Sin embargo la importancia de la distribución normal proviene principalmente del *Teorema central de limite*, que dice que bajo condiciones muy generales la suma (o media) de un conjunto de variables aleatorias se distribuye aproximadamente normal. En la Figura 2.1 se muestra el histograma de un conjunto de datos  $\mathbf{x}$  provenientes del modelo normal estándar univariado, es decir,  $\mathbf{x} \sim N(0, 1)$ .

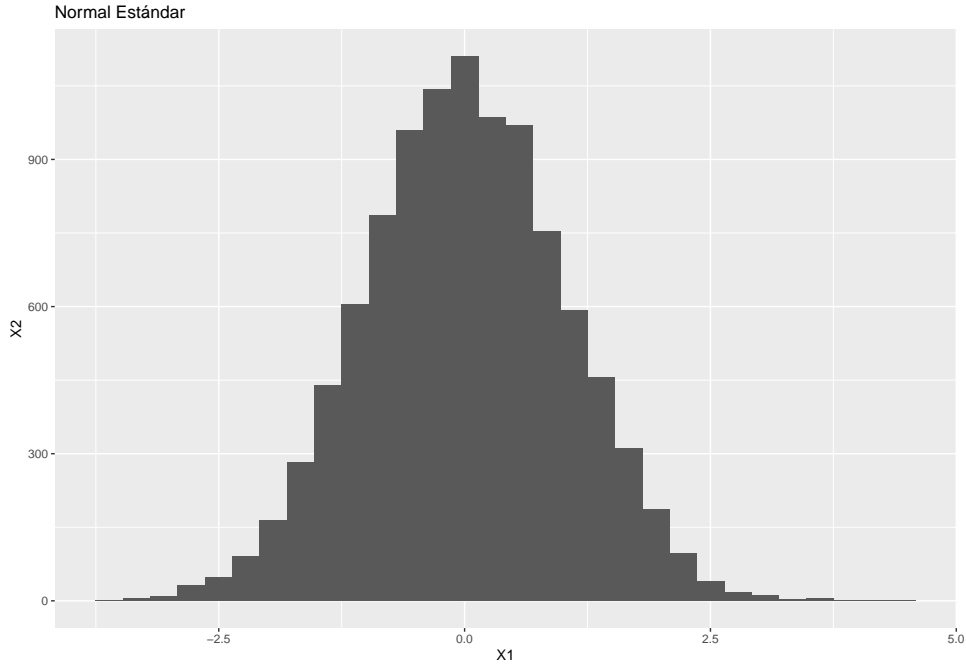


Figura 2.1: Histograma de la normal estándar.

Para una muestra aleatoria  $\mathbf{x} = \{x_1, x_2, \dots, x_n\}$  del modelo Normal Univariado, su verosimilitud esta dada por:

$$\begin{aligned}
 L(\mu, \lambda | X) &= \prod_{i=1}^n N(x_i | \mu, \lambda) \\
 &= \prod_{i=1}^n \lambda^{\frac{1}{2}} (2\pi)^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} \lambda (x_i - \mu)^2 \right\} \\
 &= \lambda^{\frac{n}{2}} (2\pi)^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2} \lambda \sum_{i=1}^n (x_i - \mu)^2 \right\} \\
 &= \lambda^{\frac{n}{2}} (2\pi)^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2} \lambda \left[ \sum_{i=1}^n x_i^2 - 2\mu \sum_{i=1}^n x_i + n\mu^2 \right] \right\}
 \end{aligned} \tag{2.1.4}$$

donde  $\sum_{i=1}^n x_i^2$  y  $\sum_{i=1}^n x_i$  son estadísticos suficientes para  $\mu$  y  $\lambda$ .

Uno de los principales objetivos es hacer inferencia de manera Bayesiana para los parámetros para  $\mu$  y  $\lambda$ , es decir, analizar la distribución posterior

$$\begin{aligned}
 \pi(\mu, \lambda | \mathbf{x}) &= \frac{L(\mu, \lambda | \mathbf{x})\pi(\mu, \lambda)}{\int L(\mu, \lambda | \mathbf{x})\pi(\mu, \lambda)d\mu d\lambda} \\
 &\propto L(\mu, \lambda | \mathbf{x})\pi(\mu, \lambda)
 \end{aligned} \tag{2.1.5}$$

En el siguiente capitulo se mostrará la forma de realizar dicha inferencia tanto para el modelo univariado como para el modelo multivariado.

## 2.2. El modelo Normal Multivariado

Un vector aleatorio continuo  $\mathbf{x} = (x_1, \dots, x_p) \in \mathbb{R}^p$ , tiene una distribución *normal multivariada* de dimensión  $p$ , con parámetros  $\boldsymbol{\mu}$  y  $\boldsymbol{\lambda}$  (donde  $\boldsymbol{\mu} = (\mu_1, \dots, \mu_p) \in \mathbb{R}^p$  y  $\boldsymbol{\lambda}$  es una matriz simétrica y positiva definida de dimensión  $p \times p$ ), si su densidad de probabilidad  $N_p(\mathbf{x} \mid \boldsymbol{\mu}, \boldsymbol{\lambda})$  es:

$$N_p(\mathbf{x} \mid \boldsymbol{\mu}, \boldsymbol{\lambda}) = |\boldsymbol{\lambda}|^{\frac{1}{2}} (2\pi)^{-\frac{p}{2}} \exp \left\{ -\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^t \boldsymbol{\lambda} (\mathbf{x} - \boldsymbol{\mu}) \right\} \quad (2.2.1)$$

Con vector de esperanza:

$$\mathbb{E}[\mathbf{x}] = \boldsymbol{\mu} \quad (2.2.2)$$

Y con matriz de covarianzas:

$$\text{Cov}[\mathbf{x}] = \boldsymbol{\lambda}^{-1} \quad (2.2.3)$$

Dentro de las distribuciones multivariantes la distribución normal multivariada es de especial importancia ya que es la generalización de la distribución normal univariada, por lo que, muchas de sus propiedades destacables serán extensiones naturales de las antes vistas para el caso univariado. Una propiedad destacable propia del caso multivariado sería que si  $\mathbf{x} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\lambda})$ , entonces cada componente  $x_i \sim N(\mu_i, \lambda_i)$  esto para cada  $i = 1, 2, \dots, p$  y donde cada  $\lambda_i$  es el elemento  $i$ -ésimo en la diagonal principal de la matriz de covarianzas  $\boldsymbol{\lambda}$ , ver más en [Montanero, 2018].

En la Figura 2.2 se muestra el gráfico de puntos de un conjunto de datos  $\mathbf{x}$  provenientes del modelo normal estándar bivariado, es decir,  $\mathbf{x} \sim N_2 \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right)$ .

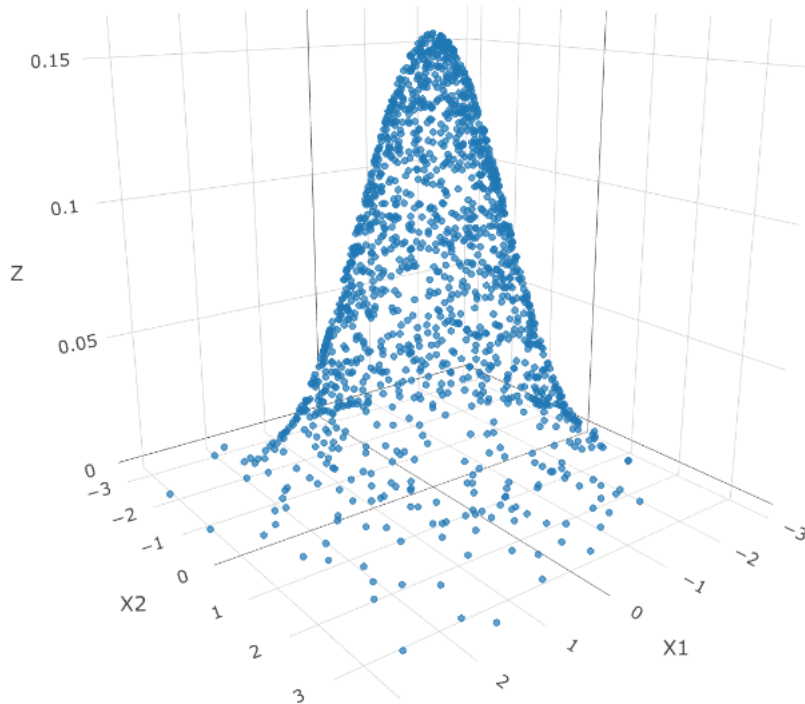


Figura 2.2: Gráfico de puntos de la normal estándar bivariada.

## 2.3. Un modelo no-paramétrico

El modelo de mezclas es un modelo probabilístico conocido por su flexibilidad e importantes ventajas, ver más en [Lindsay, 1995], es justo esta flexibilidad la que permite aproximar problemas no paramétricos, por lo que también suelen denominarse modelos semiparamétricos. Este modelo suele considerar la existencia de subpoblaciones dentro de una población general y consiste en una distribución de probabilidad que es una combinación convexa de varias distribuciones. Por lo tanto, hay distintas variantes, por ejemplo hay modelos de mezclas *finitas* o *infinitas* ya sea de la misma distribución o distribuciones distintas, es decir, todas las distribuciones pueden ser discretas, todas continuas o combinadas. En particular en el presente trabajo de tesis se está interesado en un modelo de mezclas infinito y cuyas distribuciones (continuas) sean iguales, es decir, estamos interesados en el modelo de mezclas infinitas de distribuciones normales univariadas y multivariadas. Cabe señalar que el motivo de presentar por separado el caso univariado del multivariado, cobrará sentido en el siguiente capítulo al mostrar la forma de realizar inferencias sobre los parámetros correspondientes a cada caso.

### El modelo de Mezcla Infinita de Normales Univariadas

Una variable aleatoria continua  $x \in \mathbb{R}$ , tiene una distribución de *mezcla infinita de normales univariadas*, con vector de pesos  $\boldsymbol{\omega} = (\omega_1, \omega_2, \dots)$  y parámetros  $\boldsymbol{\mu} = (\mu_1, \mu_2, \dots)$  y  $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \dots)$ , si su densidad de probabilidad  $f(x | \boldsymbol{\omega}, \boldsymbol{\mu}, \boldsymbol{\lambda})$  es:

$$f(x | \boldsymbol{\omega}, \boldsymbol{\mu}, \boldsymbol{\lambda}) = \sum_{i=1}^{\infty} \omega_i N(x | \mu_i, \lambda_i) \quad (2.3.1)$$

Donde cada  $N(x | \mu_i, \lambda_i)$  es como 2.1.1 y además  $\sum_{i=1}^{\infty} \omega_i = 1$ .

En la Figura 2.3 se puede observar el histograma de una mezcla de normales univariadas con tres componentes.

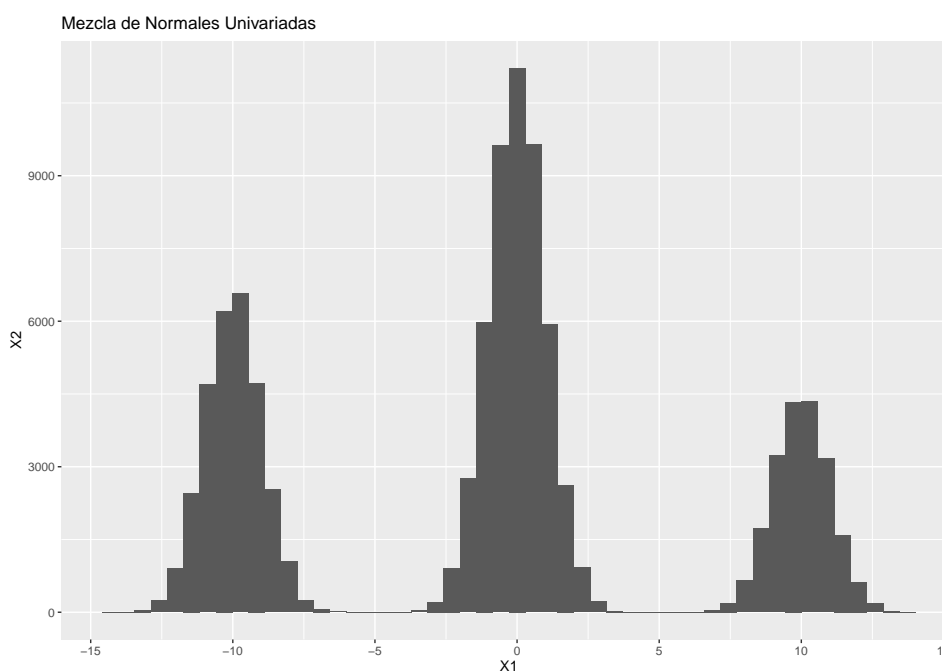


Figura 2.3: Histograma de una mezcla de normales univariadas con tres componentes.

### El modelo de Mezcla Infinita de Normales Multivariadas

Un vector aleatorio continuo  $\mathbf{x} = (x_1, \dots, x_p) \in \mathbb{R}^p$ , tiene una distribución de *mezcla infinita de normales multivariadas* de dimensión  $p$ , con vector de pesos  $\boldsymbol{\omega} = (\omega_1, \omega_2, \dots)$  y parámetros  $\boldsymbol{\mu} = (\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \dots)$  y  $\boldsymbol{\lambda} = (\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \dots)$ , si su densidad de probabilidad  $f(\mathbf{x} | \boldsymbol{\omega}, \boldsymbol{\mu}, \boldsymbol{\lambda})$  es:

$$f(\mathbf{x} | \boldsymbol{\omega}, \boldsymbol{\mu}, \boldsymbol{\lambda}) = \sum_{i=1}^{\infty} \omega_i N_p(\mathbf{x} | \boldsymbol{\mu}_i, \boldsymbol{\lambda}_i) \quad (2.3.2)$$

Donde cada  $N_p(\mathbf{x} | \boldsymbol{\mu}_i, \boldsymbol{\lambda}_i)$  es como 2.2.1 y además  $\sum_{i=1}^{\infty} \omega_i = 1$ .

En la Figura 2.4 se puede observar el gráfico de puntos de una mezcla de normales bivariadas con cuatro componentes.

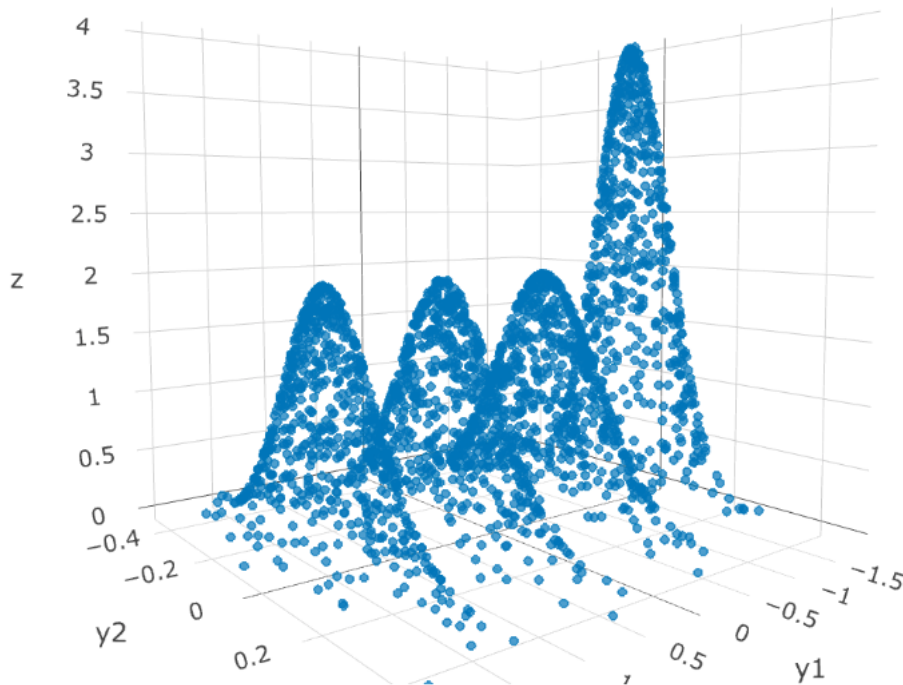


Figura 2.4: Gráfico de puntos de una mezcla de normales bivariadas con cuatro componentes.

Podemos observar que los modelos de mezclas son multimodales teniendo así la capacidad de modelar situaciones más complejas y reales. El principal problema y causa de la complejidad en este tipo de modelos es la colección infinita contable de parámetros  $\{\boldsymbol{\omega}, \boldsymbol{\mu}, \boldsymbol{\lambda}\}$  para los que hay que realizar inferencias.

Es importante mencionar que deben considerarse métodos de computación intensiva para inferir en modelos de mezclas, tales como, los *métodos MCMC*, *algoritmos EM*, entre otros. El enfoque bayesiano que utilizan los métodos MCMC nos permiten transformar la estructura compleja de un modelo de mezclas en un conjunto de estructuras simples utilizando *variables latentes*. No debemos olvidar que el principal objetivo es modelar, por lo que, aunque exista un problema de identificabilidad este no impactara de manera significativa este estudio.

## Procesos Dirichlet

En la teoría de probabilidad los *Procesos Dirichlet* son una familia de procesos estocásticos cuyas realizaciones son distribuciones de probabilidad, es decir, un proceso Dirichlet se puede entender como una variable aleatoria cuyo soporte son variables aleatorias, ver más en [Ferguson, 1973]. El proceso Dirichlet denotado como  $D(c, P_0)$  se especifica mediante una distribución base  $P_0$  y un número real positivo  $c$  llamado parámetro de concentración (también conocido como parámetro de escala), la distribución base es el valor esperado del proceso, es decir, el proceso de Dirichlet dibuja distribuciones “alrededor” de la distribución base. Sin embargo, incluso si la distribución base es continua, las distribuciones extraídas a partir de proceso de Dirichlet serán casi seguramente discretas, ver más en [Blackwell et al., 1973], y es justamente el parámetro de escala quien especifica qué tan fuerte es esta discretización.

Una aplicación particularmente importante de los procesos de Dirichlet es como una distribución de probabilidad previa en modelos de mezclas infinitas ya que son el conjugado previo para distribuciones discretas infinitas y no paramétricas. Entonces en el presente trabajo de tesis, se propone utilizar los procesos Dirichlet para llevar a cabo la inferencia sobre las colecciones infinitas  $\{\omega, \mu, \lambda\}$  de parámetros de los modelos propuestos. Cabe señalar que los resultados presentados a continuación para el modelo univariado, son completamente equivalentes para el modelo multivariado.

## Representación Stick-Breaking

Se sabe que una medida de probabilidad  $P$  se puede elegir de  $D(c, P_0)$  y que bajo un esquema de muestreo atribuible a [Sethuraman and Tiwari, 1982] e involucrando el llamado proceso *stick-breaking*, se puede representar como:

$$P = \sum_{j=1}^{\infty} \omega_j \delta_{\phi_j} \quad (2.3.3)$$

donde  $\delta_{\phi_j}$  denota la medida con una masa puntual de 1 en  $\phi_j = \{\mu_j, \lambda_j\}$ . Y donde además los pesos  $\omega_j$  serán obtenidos mediante el proceso stick-breaking el cual se dará más a detalle en el siguiente capítulo.

## El modelo de Mezclas de Procesos Dirichlet

Como ya se menciona las distribuciones aleatorias generadas por los procesos Dirichlet son discretas casi seguramente, por lo tanto, un modelo de *mezclas de procesos Dirichlet* (MDP) se basa en la idea de construir funciones de distribución absolutamente continuas. Así, el modelo propuesto se puede establecer como:

$$f(x | P) = \int N(x | \mu, \lambda) dP(\mu, \lambda) \quad (2.3.4)$$

Esta construcción de funciones de distribución absolutamente continuas se considero por primera vez en [Lo, 1984]. Luego, a partir de la ecuación 2.3.3 para  $P$ , el modelo de la ecuación 2.3.4 se puede representar como:

$$f(x | \omega, \mu, \lambda) = \sum_{j=1}^{\infty} \omega_j N(x | \mu_j, \lambda_j) \quad (2.3.5)$$



---

Entonces realizar inferencias sobre las colecciones infinitas de parámetros en los modelos de mezclas infinitos propuestos en las ecuaciones 2.3.1 y 2.3.2 se reduce a realizar inferencias en el modelo MDP, en otras palabras por el enfoque utilizado, se reduce a obtener muestras del modelo MDP. Esto se logrará mediante la propuesta de Walker en [Walker, 2007], que prácticamente se basa en muestrear el Modelo de Mezclas de Procesos Dirichlet mediante un Gibbs Sampling a través de la inclusión de variables latentes.



# Capítulo 3

## Inferencia Bayesiana del Modelo

Una de las principales ventajas de la metodología bayesiana es que evita el “sobreajuste” y con esto desvanece la difícil tarea de ajustar la complejidad del modelo. Por eso se comienza con la inferencia Bayesiana para el modelo normal univariado y multivariado, misma que será incluida en la inferencia del modelo propuesto. Por último, se presenta la inferencia al modelo de mezclas así como el algoritmo para llevar a cabo la implementación del mismo.

### 3.1. Inferencia para el Modelo Normal

El modelo normal es uno de los modelos de probabilidad más importante y sin duda alguna el más estudiado, por lo tanto, se cuenta con una amplia documentación con respecto a la inferencia de sus parámetros, tanto para el enfoque clásico como para el enfoque Bayesiano. Recordemos que bajo la metodología Bayesiana proponer distintas distribuciones iniciales conducirá a obtener distintas distribuciones posteriores, por ese motivo, en el presente trabajo de tesis se buscaron iniciales de tal forma que las posteriores fuesen lo más adecuadas posibles.

Por otro lado el objetivo principal de este capítulo es realizar inferencias sobre el modelo propuesto, es decir, sobre el modelo de mezclas infinitas. Por lo tanto, para el modelo normal únicamente se presentan las distribuciones posteriores obtenidas con las distribuciones iniciales propuestas en [Bernardo and Smith, 2009], si el lector desea obtener información sobre los cálculos de estos resultados puede ver más en [Geisser and Cornfield, 1963]. Las inferencias presentadas a continuación tienen la particularidad de ofrecernos distribuciones posteriores marginales y completamente especificadas, esto resulta muy conveniente al momento de la implementación computacional.

#### 3.1.1. Inferencia para el Modelo Normal Univariado

Dada una muestra aleatoria  $\mathbf{x} = \{x_1, x_2, \dots, x_n\}$  con  $x_i \in \mathbb{R}$  para toda  $i = 1, 2, \dots, n$  y los hiperparámetros “ $\alpha, \beta$ ” y “ $\mu_0, n_0$ ”, para las siguientes distribuciones iniciales respectivamente:

$$\pi(\lambda) = Ga(\lambda \mid \alpha, \beta) \quad (3.1.1)$$

$$\pi(\mu \mid \lambda) = N(\mu \mid \mu_0, n_0\lambda) \quad (3.1.2)$$

Se obtiene las siguientes distribuciones posteriores:

$$\pi(\mu | \mathbf{x}) = St\left(\mu | \mu_n, (n + n_0) \left(\alpha + \frac{1}{2}n\right) \beta_n^{-1}, 2\alpha + n\right) \quad (3.1.3)$$

$$\pi(\lambda | \mathbf{x}) = Ga\left(\lambda | \alpha + \frac{1}{2}n, \beta_n\right) \quad (3.1.4)$$

donde

$$\mu_n = (n_0 + n)^{-1} (n_0\mu_0 + n\bar{x})$$

$$\beta_n = \beta + \frac{1}{2}ns^2 + \frac{1}{2}(n_0 + n)^{-1} n_0n (\mu_0 - \bar{x})^2$$

y los estadísticos utilizados son  $\bar{x} = n^{-1} \sum_{i=1}^n x_i$  y  $s^2 = n^{-1} \sum_{i=1}^n (x_i - \bar{x})^2$ .

Entonces observamos que la inferencia implementada para el modelo normal univariado hace uso de las distribuciones Normal Univariada, Gamma y Student Univariada.

La distribución Normal Univariada ya fue formulada en 2.1.1, por lo tanto, a continuación se formulan las distribuciones Gamma y Student Univariada.

### Distribución Gamma

Una variable aleatoria continua  $x \in \mathbb{R}_+$ , tiene una distribución *Gamma*, con parámetros  $\alpha$  y  $\beta$  (donde  $\alpha > 0$  y  $\beta > 0$ ), si su densidad de probabilidad  $Ga(x | \alpha, \beta)$  es:

$$Ga(x | \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp\{-\beta x\} \quad (3.1.5)$$

Con esperanza:

$$\mathbb{E}[x] = \frac{\alpha}{\beta} \quad (3.1.6)$$

Y con varianza:

$$\mathbb{V}[x] = \frac{\alpha}{\beta^2} \quad (3.1.7)$$

### Distribución Student Univariada

Una variable aleatoria continua  $x \in \mathbb{R}$ , tiene una distribución *Student univariada*, con parámetros  $\mu$ ,  $\lambda$  y  $\alpha$  (donde  $\mu \in \mathbb{R}$ ,  $\lambda > 0$  y  $\alpha > 0$ ), si su densidad de probabilidad  $St(x | \mu, \lambda, \alpha)$  es:

$$St(x | \mu, \lambda, \alpha) = \frac{\Gamma\left(\frac{(\alpha+1)}{2}\right)}{\Gamma\left(\frac{\alpha}{2}\right)} \left(\frac{\lambda}{\alpha\pi}\right)^{\frac{1}{2}} \left[1 + \frac{1}{\alpha}\lambda(x - \mu)^2\right]^{-\frac{(\alpha+1)}{2}} \quad (3.1.8)$$

Con esperanza:

$$\mathbb{E}[x] = \mu \quad (3.1.9)$$

Y con varianza:

$$\mathbb{V}[x] = \lambda^{-1} \left(\frac{\alpha}{(\alpha - 2)}\right) \quad (3.1.10)$$

### 3.1.2. Inferencia para el Modelo Normal Multivariado

Dada una muestra aleatoria  $\mathbf{x} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$  con  $\mathbf{x}_i \in \mathbb{R}^p$  donde  $p \geq 2$  para toda  $i = 1, 2, \dots, n$  y los hiperparámetros “ $\alpha, \boldsymbol{\beta}$ ” y “ $\boldsymbol{\mu}_0, n_0$ ”, para las siguientes distribuciones iniciales respectivamente:

$$\pi(\boldsymbol{\lambda}) = Wi_p(\boldsymbol{\lambda} \mid \alpha, \boldsymbol{\beta}) \quad (3.1.11)$$

$$\pi(\boldsymbol{\mu} \mid \boldsymbol{\lambda}) = N_p(\boldsymbol{\mu} \mid \boldsymbol{\mu}_0, n_0 \boldsymbol{\lambda}) \quad (3.1.12)$$

Se obtiene las siguientes distribuciones posteriores:

$$\pi(\boldsymbol{\mu} \mid \mathbf{x}) = St_p(\boldsymbol{\mu} \mid \boldsymbol{\mu}_n, (n + n_0) \alpha_n \boldsymbol{\beta}_n^{-1}, 2\alpha_n) \quad (3.1.13)$$

$$\pi(\boldsymbol{\lambda} \mid \mathbf{x}) = Wi_p\left(\boldsymbol{\lambda} \mid \alpha + \frac{1}{2}n, \boldsymbol{\beta}_n\right) \quad (3.1.14)$$

donde

$$\boldsymbol{\mu}_n = (n_0 + n)^{-1} (n_0 \boldsymbol{\mu}_0 + n \bar{\mathbf{x}})$$

$$\boldsymbol{\beta}_n = \boldsymbol{\beta} + \frac{1}{2} \mathbf{S} + \frac{1}{2} (n_0 + n)^{-1} n_0 n (\boldsymbol{\mu}_0 - \bar{\mathbf{x}}) (\boldsymbol{\mu}_0 - \bar{\mathbf{x}})^t$$

$$\alpha_n = \alpha + \frac{1}{2}n - \frac{1}{2}(p - 1)$$

y los estadísticos utilizados son  $\bar{\mathbf{x}} = n^{-1} \sum_{i=1}^n \mathbf{x}_i$  y  $\mathbf{S} = \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^t$ .

Entonces observamos que la inferencia implementada para el modelo normal multivariado hace uso de las distribuciones Normal Multivariada, Whishart y Student Multivariada.

La distribución Normal Multivariada ya fue formulada en 2.2.1, por lo tanto, a continuación se formulan las distribuciones Whishart y Student Multivariada.

#### Distribución Wishart

Una matriz  $\mathbf{x}$  simétrica y positiva definida de cantidades aleatorias  $x_{ij} = x_{ji}$ , para  $i, j = 1, \dots, p$  con  $p \geq 2$ , tiene una distribución *Wishart* de dimensión  $p$ , con parámetros  $\alpha$  y  $\boldsymbol{\beta}$  (donde  $2\alpha > p - 1$  y  $\boldsymbol{\beta}$  es una matriz simétrica y no singular de dimensión  $p \times p$ ), si la densidad  $Wi_p(\mathbf{x} \mid \alpha, \boldsymbol{\beta})$  del vector aleatorio  $\frac{p(p+1)}{2}$  dimensional de las entradas de  $\mathbf{x}$  es:

$$Wi_p(\mathbf{x} \mid \alpha, \boldsymbol{\beta}) = \frac{\pi^{\frac{-p(p-1)}{4}} |\boldsymbol{\beta}|^\alpha}{\prod_{l=1}^p \Gamma\left(\frac{1}{2}(2\alpha + 1 - l)\right)} |\mathbf{x}|^{\alpha - \frac{(p+1)}{2}} \exp\{-tr(\boldsymbol{\beta}\mathbf{x})\} \quad (3.1.15)$$

Con esperanza:

$$\mathbb{E}[\mathbf{x}] = \alpha \boldsymbol{\beta}^{-1} \quad (3.1.16)$$

Y con el elemento  $ij^{th}$  de su matriz de covarianzas:

$$\text{Cov}[x_{ij}] = \frac{\alpha}{2} (\sigma_{ij}^2 + \sigma_{ii}\sigma_{jj}) \quad (3.1.17)$$

Donde  $\sigma_{ij}$  es el elemento  $ij^{th}$  de la matriz de covarianzas  $\boldsymbol{\Sigma} = \boldsymbol{\beta}^{-1}$ .

### Distribución Student Multivariada

Un vector aleatorio continuo  $\mathbf{x} = (x_1, x_2, \dots, x_p) \in \mathbb{R}^p$  con  $p \geq 2$ , tiene una distribución *Student multivariada* de dimensión  $p$ , con parámetros  $\boldsymbol{\mu}$ ,  $\boldsymbol{\lambda}$  y  $\alpha$  (donde  $\boldsymbol{\mu} = (\mu_1, \mu_2, \dots, \mu_p) \in \mathbb{R}^p$ ,  $\boldsymbol{\lambda}$  es una matriz simétrica y positiva definida de dimensión  $p \times p$  y  $\alpha > 0$ ), si su densidad de probabilidad  $St_p(\mathbf{x} \mid \boldsymbol{\mu}, \boldsymbol{\lambda}, \alpha)$  es:

$$St_p(\mathbf{x} \mid \boldsymbol{\mu}, \boldsymbol{\lambda}, \alpha) = \frac{\Gamma\left(\frac{\alpha+p}{2}\right)}{\Gamma\left(\frac{\alpha}{2}\right) (\alpha\pi)^{\frac{p}{2}}} |\boldsymbol{\lambda}|^{\frac{1}{2}} \left[1 + \frac{1}{\alpha} (\mathbf{x} - \boldsymbol{\mu})^t \boldsymbol{\lambda} (\mathbf{x} - \boldsymbol{\mu})\right]^{-\frac{(\alpha+p)}{2}} \quad (3.1.18)$$

Con esperanza:

$$\mathbb{E}[\mathbf{x}] = \boldsymbol{\mu} \quad (3.1.19)$$

Y con matriz de covarianzas:

$$\text{Cov}[\mathbf{x}] = \boldsymbol{\lambda}^{-1} \left( \frac{\alpha}{(\alpha - 2)} \right) \quad (3.1.20)$$

## 3.2. Inferencia para el Modelo no-paramétrico

A continuación retomaremos el modelo propuesto en las ecuaciones 2.3.1 y 2.3.2, con el propósito realizar inferencias sobre sus parámetros. Nótese que, aunque los parámetros asociados al modelo en cualquiera de los casos (univariado o multivariado) es la colección infinita  $\{\boldsymbol{\omega}, \boldsymbol{\mu}, \boldsymbol{\lambda}\}$ , en el fondo, el parámetro de interés es el modelo subyacente a los datos, es decir, el modelo de mezclas.

Antes de comenzar con la inferencia, recordemos que los modelos de mezclas ofrecen una gran versatilidad en la descripción de datos, lo que los convierte en herramientas ampliamente utilizadas, sin embargo, no fue hasta que [Escobar and West, 1995] propusieron una implementación computacional, que se generó un aumento en el interés por este tipo de modelos. Una vez más, aunque los resultados presentados a continuación se hacen para el modelo de mezclas univariado, son completamente equivalentes para el modelo multivariado.

Comenzamos obteniendo los pesos involucrados en la ecuación 2.3.3 mediante el proceso stick-breaking:

$$\begin{aligned} \omega_1 &= \nu_1 \\ \omega_j &= \nu_j \prod_{l < j} (1 - \nu_l) \quad \forall j > 1 \end{aligned} \quad (3.2.1)$$

donde las variables  $\nu_1, \nu_2, \dots$  son variables aleatorias independientes e indéticamente distribuidas  $Beta(1, c)$ , con  $c$  el parámetro de escala del proceso Dirichlet.

Luego para llevar a cabo inferencias sobre el modelo propuesto, hemos visto que necesitaremos obtener muestras del modelo de Mezclas de Procesos Dirichlet (MDP), para esto se utiliza el método dado por Walker en [Walker, 2007], el cual propone muestrear el modelo MDP a través de un Gibb Sampling en donde cada iteración estará basada en el esquema Slice Sampling, que se ocupara de la infinitud mediante la inclusión de variables latentes. Por lo tanto, se necesitan especificar todas las condicionales finales

completas de los parámetros involucrados. En principio se deberían muestrear, en cada paso del Gibbs Sampling, toda la colección infinita de parámetros  $\{\boldsymbol{\omega}, \boldsymbol{\mu}, \boldsymbol{\lambda}\}$ , sin embargo, como se verá a continuación, solo será necesario muestrear un conjunto finito de parámetros.

Entonces para estimar el modelo vía un Gibbs Sampling, se introduce una variable latente  $u$  tal que la densidad conjunta de  $(x, u)$  dado  $(\boldsymbol{\omega}, \boldsymbol{\mu}, \boldsymbol{\lambda})$  está dada por

$$f(x, u | \boldsymbol{\omega}, \boldsymbol{\mu}, \boldsymbol{\lambda}) = \sum_{j=1}^{\infty} \mathbf{1}(u < \omega_j) N(x | \mu_j, \lambda_j) \quad (3.2.2)$$

donde  $\mathbf{1}(\cdot)$  es la función indicadora sobre la condición  $(\cdot)$ .

La introducción de la variable latente  $u$  es crucial en el proceso de inferencia puesto que permite lidiar con el problema de la “infinitud” en cada paso del algoritmo.

Claramente, la integración sobre  $u$  nos devuelve la densidad deseada  $f(x | \boldsymbol{\omega}, \boldsymbol{\mu}, \boldsymbol{\lambda})$ , por lo tanto, la densidad conjunta existe y también existirá una densidad marginal para  $u$ . Alternativamente podemos escribir la ecuación 3.2.2 como:

$$f(x, u | \boldsymbol{\omega}, \boldsymbol{\mu}, \boldsymbol{\lambda}) = \sum_{j=1}^{\infty} \omega_j U(u | 0, \omega_j) N(x | \mu_j, \lambda_j) \quad (3.2.3)$$

Así, con probabilidad  $\omega_j$ ,  $x$  y  $u$  son independientes, por lo tanto, la densidad marginal para  $u$  está dada por:

$$\begin{aligned} f(u | \boldsymbol{\omega}) &= \sum_{j=1}^{\infty} \omega_j U(u | 0, \omega_j) \\ &= \sum_{j=1}^{\infty} \mathbf{1}(u < \omega_j) \end{aligned} \quad (3.2.4)$$

Por otro lado, si se define el siguiente conjunto:

$$A(u | \boldsymbol{\omega}) = \{j : \omega_j > u\}$$

entonces la ecuación 3.2.2 se puede escribir como:

$$f(x, u | \boldsymbol{\omega}, \boldsymbol{\mu}, \boldsymbol{\lambda}) = \sum_{j \in A(u | \boldsymbol{\omega})} N(x | \mu_j, \lambda_j) \quad (3.2.5)$$

Por lo tanto:

$$\begin{aligned} f(x | u, \boldsymbol{\omega}, \boldsymbol{\mu}, \boldsymbol{\lambda}) &= \frac{f(x, u | \boldsymbol{\omega}, \boldsymbol{\mu}, \boldsymbol{\lambda})}{f(u | \boldsymbol{\omega})} \\ &= \frac{\sum_{j \in A(u | \boldsymbol{\omega})} N(x | \mu_j, \lambda_j)}{\sum_{j=1}^{\infty} \mathbf{1}(u < \omega_j)} \\ &= \frac{1}{f(u | \boldsymbol{\omega})} \sum_{j \in A(u | \boldsymbol{\omega})} N(x | \mu_j, \lambda_j) \end{aligned} \quad (3.2.6)$$

Así, dada la variable  $u$  se tiene un modelo de mezcla finita con pesos todos iguales a  $1/f(u | \boldsymbol{\omega})$ . Por lo tanto pasar de una suma infinita a una suma finita, dada  $u$ , va a

hacer mucha diferencia cuando se trate de un muestreo.

Luego, es común en el contexto de mezclas, introducir variables latentes indicadoras  $z_i$  que indiquen el componente de la mezcla al cual pertenece la observación  $x_i$  para toda  $i = 1, \dots, n$ .

Por lo tanto, consideremos la siguiente densidad conjunta:

$$f(x, z = k, u | \boldsymbol{\omega}, \boldsymbol{\mu}, \boldsymbol{\lambda}) = N(x | \mu_k, \lambda_k) \mathbb{1}(k \in A(u | \boldsymbol{\omega})) \quad (3.2.7)$$

Entonces, la verosimilitud completa basada en una muestra de tamaño  $n$  se ve fácilmente como:

$$f(\{x_i, u_i, z_i = k_i\}_{i=1}^n | \boldsymbol{\omega}, \boldsymbol{\mu}, \boldsymbol{\lambda}) = \prod_{i=1}^n N(x_i | \mu_{k_i}, \lambda_{k_i}) \mathbb{1}(u_i < \omega_{k_i}) \quad (3.2.8)$$

A continuación se presenta el conjunto de funciones de densidad condicionales finales completas que se necesitan para poder implementar el Gibbs Sampling.

### Densidades condicionales finales completas

1. Para  $i = 1, 2, \dots, n$  se tiene:

$$f(u_i | \dots) \sim U(0, \omega_{k_i})$$

2. Para  $j = 1, 2, \dots$  se tiene:

$$\pi(\mu_j, \lambda_j | \dots) = f(\mu_j, \lambda_j | \dots) \propto P_0(\mu_j, \lambda_j) \prod_{k_i=j} N(x_i | \mu_j, \lambda_j)$$

Si no existiera un  $k_i$  igual a  $j$  entonces  $(\mu_j, \lambda_j)$  se debe simular de su respectiva distribución inicial  $P_0(\mu_j, \lambda_j) = \pi(\mu_j, \lambda_j) = \pi(\mu_j | \lambda_j) \pi(\lambda_j)$ , es decir, el producto de las densidades dadas por las ecuaciones 3.1.2 y 3.1.1 para el caso univariado o por las ecuaciones 3.1.12 y 3.1.11 para el caso multivariado.

3. De la verosimilitud 3.2.8 se tiene:

$$\begin{aligned} f(\boldsymbol{\nu} | \dots) &\propto \pi(\boldsymbol{\nu}) \prod_{i=1}^n \mathbb{1}(u_i < \omega_{k_i}) \\ &\propto \pi(\boldsymbol{\nu}) \prod_{i=1}^n \mathbb{1}(u_i < \nu_1 \prod_{l < k_i} (1 - \nu_l)) \end{aligned}$$

Recordando que:

$$\begin{aligned} \omega_1 &= \nu_1 \\ \omega_j &= \nu_j \prod_{l < j} (1 - \nu_l) \quad \forall j > 1 \end{aligned}$$

Por lo que, de la colección infinita de  $\{\boldsymbol{\nu}\}$  solo las  $\nu_j \leq k^*$ , donde



$k^* = \max\{k_1, k_2, \dots, k_n\}$ , serán afectadas y solo para esos índices se muestrean de la distribución condicional final completa, las cuales están dadas por:

$$\pi(\nu_j) \propto \text{Beta}(a_j, b_j)$$

donde

$$a_j = 1 + \sum_{i=1}^n \mathbb{1}(z_i = j) \quad \text{y} \quad b_j = \alpha + \sum_{i=1}^n \mathbb{1}(z_i > j)$$

Sin embargo, para los  $\nu_j$  con  $j > k^*$ ,  $\nu_j \sim \text{Beta}(1, c)$ .

4. De la densidad 3.2.8 se tiene:

$$P(z_i = k \mid \dots) \propto N(x_i \mid \mu_k, \lambda_k) \mathbb{1}(k \in A(u_i \mid \boldsymbol{\omega}))$$

Los valores que toma cada  $z_i$  es sobre un conjunto finito, el cual es de la forma  $\{k : \omega_k > u_i\}$ . Por lo que solo se necesitan muestrear tantos  $\omega_k$  hasta asegurar que se tienen todos los  $\omega_k > u_i$  para toda  $i = 1, \dots, n$ . Para encontrar todos los  $\omega_k$  en cada paso del algoritmo, se propone muestrear  $k^*$  pesos  $\omega_j$  hasta que se cumpla que

$$\sum_{j=1}^{k^*} \omega_j > 1 - u^*$$

donde  $u^* = \min\{u_1, u_2, \dots, u_n\}$ . Ya que con lo anterior se garantiza que todos los  $\omega_k$  con  $k = 1, \dots, k^*$  cumplan con  $\omega_k > u_i$  para toda  $i = 1, \dots, n$ .

Ahora bien, aunque se esta utilizando en su totalidad el algoritmo dado por Walker en [Walker, 2007], vamos a complementarlo con una propuesta más eficiente de dicho muestreo realizada por Kalli en [Kalli et al., 2011], la cual señala que si se considera la densidad conjunta dada por la ecuación 3.2.8 se presentan varios problemas con el muestreo debidos en parte a la correlación entre las variables  $u$  y  $\boldsymbol{\omega}$ . Adicionalmente señala que la actualización de  $u$  puede llevar a simular más pesos  $\omega$  de los necesarios. Por lo anterior proponen considerar un Slice Sampling más general definiendo la siguiente densidad conjunta:

$$f(x, z = k, u \mid \boldsymbol{\omega}, \boldsymbol{\mu}, \boldsymbol{\lambda}) = \xi_k^{-1} \mathbb{1}(u < \xi_k) \omega_k N(x \mid \mu_k, \lambda_k) \quad (3.2.9)$$

donde  $\xi_1, \xi_2, \xi_3, \dots$  es una sucesión determinista de valores positivos decreciente. Algunas opciones que producen muestreos eficientes son  $\xi_j = (1 - \kappa)\kappa^{j-1}$  con  $\kappa = 1/2$  y  $\xi_j = e^{-j}$ .

Así, la verosimilitud para los datos completos basada sobre una muestra de tamaño  $n$  está dada por:

$$f(\{x_i, z_i = k_i, u_i\}_{i=1}^n \mid \boldsymbol{\xi}, \boldsymbol{\omega}, \boldsymbol{\mu}, \boldsymbol{\lambda}) = \prod_{i=1}^n \mathbb{1}(u_i < \xi_{k_i}) \omega_{k_i} / \xi_{k_i} N(x_i \mid \mu_{k_i}, \lambda_{k_i}) \quad (3.2.10)$$

Si  $\boldsymbol{\xi}$  y  $\boldsymbol{\nu}$  son condicionalmente independientes entonces las condicionales completas para la implementación del Gibbs Sampling, puedan ser especificadas de la siguiente manera:

1. Para  $j = 1, 2, \dots$  se tiene:

$$f(\mu_j, \lambda_j | \dots) \propto N(\mu_j | \mu_0, n_0 \lambda_j) Ga(\lambda_j | \alpha, \beta) \prod_{k_i=j} N(x_i | \mu_j, \lambda_j)$$

2. Para  $j = 1, 2, \dots$  se tiene:

$$f(\nu_j | \dots) \propto Beta(\nu_j | a_j, b_j), \text{ donde:}$$

$$a_j = 1 + \sum_{i=1}^n \mathbf{1}(z_i = j) \quad \text{y} \quad b_j = c + \sum_{i=1}^n \mathbf{1}(z_i > j)$$

3. Para  $i = 1, 2, \dots, n$  se tiene:

$$f(u_i | \dots) \propto \mathbf{1}(0 < u_i < \xi_{k_i})$$

4.  $Pr(z_i = k | \dots) \propto \mathbf{1}(k : \xi_k > u_i) \omega_k / \xi_k N(x_i | \mu_k, \lambda_k)$

Entonces para proceder con la cadena solo se necesita muestrear hasta un entero  $N$  para el que se han encontrado todos los  $k$  apropiados para realizar el paso 4 exactamente. De hecho es fácil encontrar el conjunto de  $k$  requeridos, ya que será del tipo  $\{1, \dots, N\}$  donde  $N = \max\{N_i\}$  y  $N_i$  es el entero más grande  $l$  para el cual  $\xi_l > u_i$ . Para mayores detalles del algoritmo el lector puede consultar [Kalli et al., 2011].

# Capítulo 4

## Aplicaciones

En este capítulo ilustraremos el potencial que puede tener el *Modelo de Mezclas de Normales Multivariadas* en la descripción de datos composicionales. Comenzaremos con una serie de ejemplos con datos simulados, con el objetivo de evaluar la eficiencia del modelo. Lo anterior, a través del análisis de la distribución predictiva. Seguido de esto se pondrá a prueba el modelo con un conjunto de datos reales derivado de ejercicios de *Conteo Rápidos* realizados por el Instituto Nacional Electoral (INE) en México. En este caso, además de ofrecer la distribución predictiva como una descripción al comportamiento de la población, se ofrece un estimador puntual, una región de probabilidad e intervalos de probabilidad marginales para el parámetro de interés que es la proporción de votos de las fuerzas electorales involucradas en dicho proceso.

### 4.1. Datos Simulados

A continuación presentamos dos ejemplos con datos simulados, con el fin de evaluar y analizar la metodología propuesta. Ambos ejemplos, representan patrones de datos poco convencionales, el primero representa un patrón de datos bastante complejo de modelar y el segundo representa un patrón de datos bastante extremo al momento de recuperar ciertas características de la población de la que proviene. En cualquier caso, como se podrá ver, la metodología propuesta es capaz de describir adecuadamente las características de los datos.

#### Ejemplo1: Una mezcla de Normales Bivariadas

El primer conjunto de datos simulados lo construimos a partir del siguiente modelo de mezclas de normales bivariadas con 3 componentes:

$$\mathbf{x} \sim f(\mathbf{x}|\boldsymbol{\omega}, \boldsymbol{\mu}, \boldsymbol{\lambda}) \quad \text{donde,}$$

$$f(\mathbf{x}|\boldsymbol{\omega}, \boldsymbol{\mu}, \boldsymbol{\lambda}) = \sum_{i=1}^3 \omega_i N_2(\mathbf{x}|\boldsymbol{\mu}_i, \boldsymbol{\lambda}_i),$$

y cuyos parámetros son los siguientes:

$$\text{Pesos: } \omega_1 = 0.34, \omega_2 = 0.33 \text{ y } \omega_3 = 0.33$$

$$\text{Medias: } \boldsymbol{\mu}_1 = \begin{pmatrix} -0.8 \\ -0.8 \end{pmatrix}, \boldsymbol{\mu}_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \text{ y } \boldsymbol{\mu}_3 = \begin{pmatrix} 0.8 \\ -0.8 \end{pmatrix}$$

$$\text{Precisiones: } \boldsymbol{\lambda}_1 = \begin{pmatrix} 50 & 0 \\ 0 & 50 \end{pmatrix}, \boldsymbol{\lambda}_2 = \begin{pmatrix} 50 & 0 \\ 0 & 50 \end{pmatrix} \text{ y } \boldsymbol{\lambda}_3 = \begin{pmatrix} 50 & 0 \\ 0 & 50 \end{pmatrix}$$

### Muestra

A partir de este modelo se genera una muestra de tamaño 500 a la que llamaremos  $\mathbf{X}_{\text{Sim1}}$ , notemos que la muestra está en  $\mathbb{R}^2$  y por lo tanto podemos visualizarla en el diagrama de dispersión de la Figura 4.1.

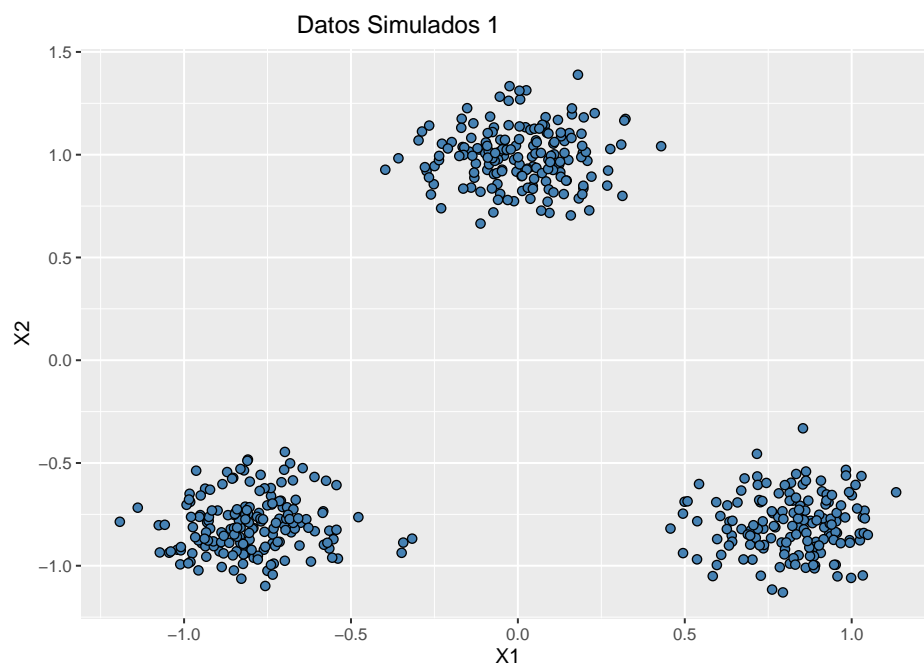


Figura 4.1: Muestra  $\mathbf{X}_{\text{Sim1}}$  (Azul).

Posteriormente, para obtener una muestra  $\mathbf{Y}_{\text{Sim1}}$  de datos composicionales le aplicamos la transformación  $ilr^{-1}$  a  $\mathbf{X}_{\text{Sim1}}$ , es decir,  $\mathbf{Y}_{\text{Sim1}} = ilr^{-1}(\mathbf{X}_{\text{Sim1}})$  la cual ya se encontrará en  $\mathbb{S}^3$  y por lo tanto podemos visualizarla en el diagrama ternario de la Figura 4.2.

Podemos observar que se trata de una muestra de datos composicionales con un patrón bastante peculiar que trataremos de modelar con la metodología propuesta. Para lo anterior, especificaremos todos los insumos necesarios necesarios para llevar a cabo la implementación de la metodología propuesta y así obtener una muestra de la distribución predictiva final. Observemos que un paso previo sería pasar la muestra  $\mathbf{Y}_{\text{Sim1}}$  de datos composicionales a los números reales bajo la transformación  $ilr$  pero esto ya lo tenemos en  $\mathbf{X}_{\text{Sim1}}$  por la manera en como se construyó la muestra.

### Especificación de los parámetros del algoritmo

A continuación se especifican los valores de arranque que tomarán los parámetros y demás valores numéricos requeridos para la implementación del análisis correspondiente.

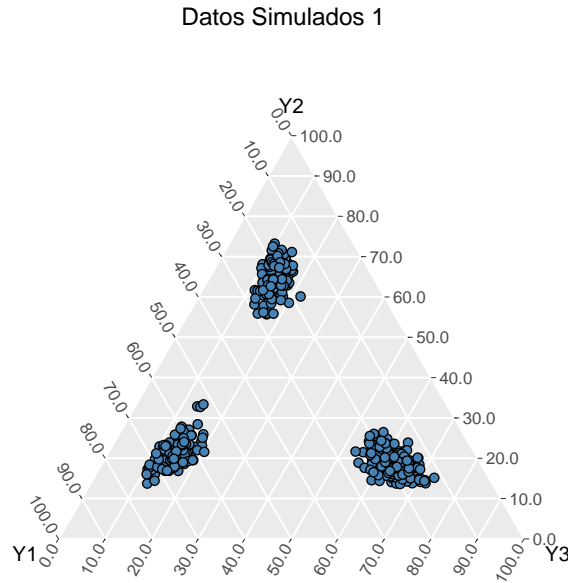


Figura 4.2: Muestra  $\mathbf{Y}_{\text{Sim1}}$  (Azul).

Comenzamos especificando en la Cuadro 4.1 los hiperparámetros del modelo normal, es decir, especificando distribuciones iniciales para  $\boldsymbol{\mu}$  y  $\boldsymbol{\lambda}$ .

Distribución Inicial	Hiperparámetros
$\boldsymbol{\mu} \mid \boldsymbol{\lambda} \sim N_2(\boldsymbol{\mu}_0, n_0\boldsymbol{\lambda})$	$\boldsymbol{\mu}_0 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$
	$n_0 = 1$
$\boldsymbol{\lambda} \sim Wi_2(\alpha, \boldsymbol{\beta})$	$\alpha = 1.00033$
	$\boldsymbol{\beta} = \begin{pmatrix} 1.00033 & 0 \\ 0 & 1.00033 \end{pmatrix}$

Cuadro 4.1: Hiperparámetros utilizados en el algoritmo.

Después de especificar la distribución inicial, cuya justificación se dará más adelante, en el Cuadro 4.2 se especifican los valores de arranque (numérico) necesarios para correr el algoritmo asociado.

$c$	$k.ini$	Burn	Conv	Lag	sizeMCMC
3	3	1,000	179,000	10	2,000

Cuadro 4.2: Valores de arranque utilizados en el algoritmo.

Para este ejemplo, se considera un periodo de calentamiento de 1,000 iteraciones seguidas de 179,000 iteraciones para alcanzar la convergencia y solo a partir de ese momento se toma una muestra de la distribución predictiva final de tamaño 2,000 cada 10 iteraciones; realizando así un total de 200,000 iteraciones del algoritmo. Los valores de arranque para  $c$  y  $k.ini$  son el resultado de realizar múltiples combinaciones y corridas del algoritmo tanto con datos simulados como con datos reales, obteniendo

con estos un equilibrio entre tiempo de cómputo y el ajuste de la distribución predictiva final.

### Distribución Predictiva

Así, tras un tiempo de cómputo de 7.6 horas el algoritmo obtiene una muestra de tamaño 2,000 tanto de la distribución predictiva final en  $\mathbb{R}^2$  que llamaremos  $\mathbf{X}_{\text{PredSim1}}$ , como de valores  $k$  y  $k_0$  que representan el número de componentes muestreados y no vacíos, respectivamente, en cada paso del algoritmo. La muestra de la distribución predictiva final puede ser visualizada junto a la muestra de los datos originales  $\mathbf{X}_{\text{Sim1}}$  en el diagrama de dispersión de la Figura 4.3.

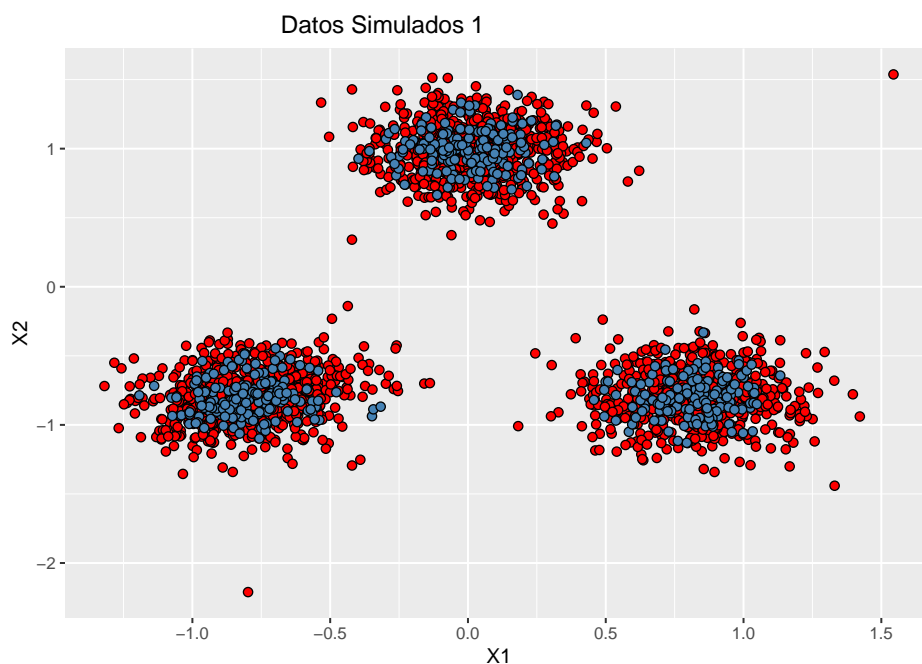


Figura 4.3: Muestra de la distribución predictiva final  $\mathbf{X}_{\text{PredSim1}}$  (Rojo) junto a la muestra original de datos  $\mathbf{X}_{\text{Sim1}}$  (Azul).

Bajo la transformación  $ilr^{-1}$  podemos obtener a  $\mathbf{Y}_{\text{PredSim1}} = ilr^{-1}(\mathbf{X}_{\text{PredSim1}})$ , misma que podemos visualizar junto a la muestra  $\mathbf{Y}_{\text{Sim1}}$  en la Figura 4.4.

Podemos observar que la muestra de la distribución predictiva final es capaz de modelar adecuadamente los tres cúmulos que inicialmente son una de las características que más resalta en este primer conjunto de datos simulados, sin embargo, también podemos notar que existen algunos puntos que se alejan demasiado de dichos cúmulos estando incluso en los lados y vértices del diagrama ternario.

Esto puede deberse a la falta de convergencia por parte de la cadena generada por el algoritmo, recordemos que para poder afirmar que la muestra  $\mathbf{X}_{\text{PredSim1}}$ , obtenida a partir de dicha cadena, proviene de la distribución predictiva final del modelo propuesto, se requiere que los valores de dicha muestra sean tomados a partir de la convergencia de esta y además deben ser tomados de tal manera que estén lo menos autocorrelacionados entre sí.

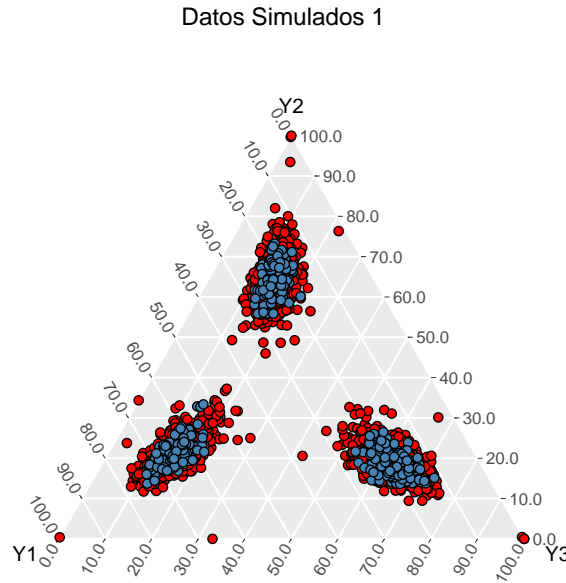


Figura 4.4: Muestra de la distribución predictiva final  $\mathbf{Y}_{\text{PredSim1}}$  (Rojo) junto a la muestra  $\mathbf{Y}_{\text{Sim1}}$  (Azul).

### Análisis de Convergencia y Autocorrelación

En la Figura 4.5 se analizan gráficamente los promedios ergódicos y autocorrelaciones de  $k$  y  $k_0$  como herramienta para monitorear la convergencia de la cadena y la autocorrelación entre los valores de la muestra obtenida por el algoritmo. Se puede observar que la autocorrelación entre los valores de las muestras para ambos parámetros es pequeño y que a pesar de que los promedios ergódicos muestran pequeñas fluctuaciones, estos son ya a una escala muy pequeña.

### Región de Probabilidad

Una vez que se tiene una muestra de la distribución predictiva final estamos en posibilidad de llevar a cabo inferencias sobre características poblacionales. Comenzamos obteniendo una región de probabilidad a través de una aproximación por *kernel normal* (Ver, por ejemplo, ...) a la densidad de la muestra de la distribución predictiva final. Las curvas de nivel de dicha aproximación pueden visualizarse en la Figura 4.6, conjuntamente con las muestras  $\mathbf{X}_{\text{Sim1}}$  y  $\mathbf{X}_{\text{PredSim1}}$ . Entonces, para obtener la región de probabilidad tomamos los puntos de la curva de nivel que acumula una probabilidad del 95 % y los llevamos al simplex bajo la transformación  $ilr^{-1}$ . Así, la correspondiente región de probabilidad junto con la muestra  $\mathbf{Y}_{\text{Sim1}}$  se pueden visualizar en el diagrama ternario de la Figura 4.7. En la Figura 4.7 se puede observar cómo esta región de probabilidad, derivada de la distribución predictiva obtenida bajo la metodología propuesta, describe adecuadamente el comportamiento general de los datos. Particularmente, la región esta conformada por tres componentes, uno para cada uno de los cúmulos característicos del conjunto de datos simulados.

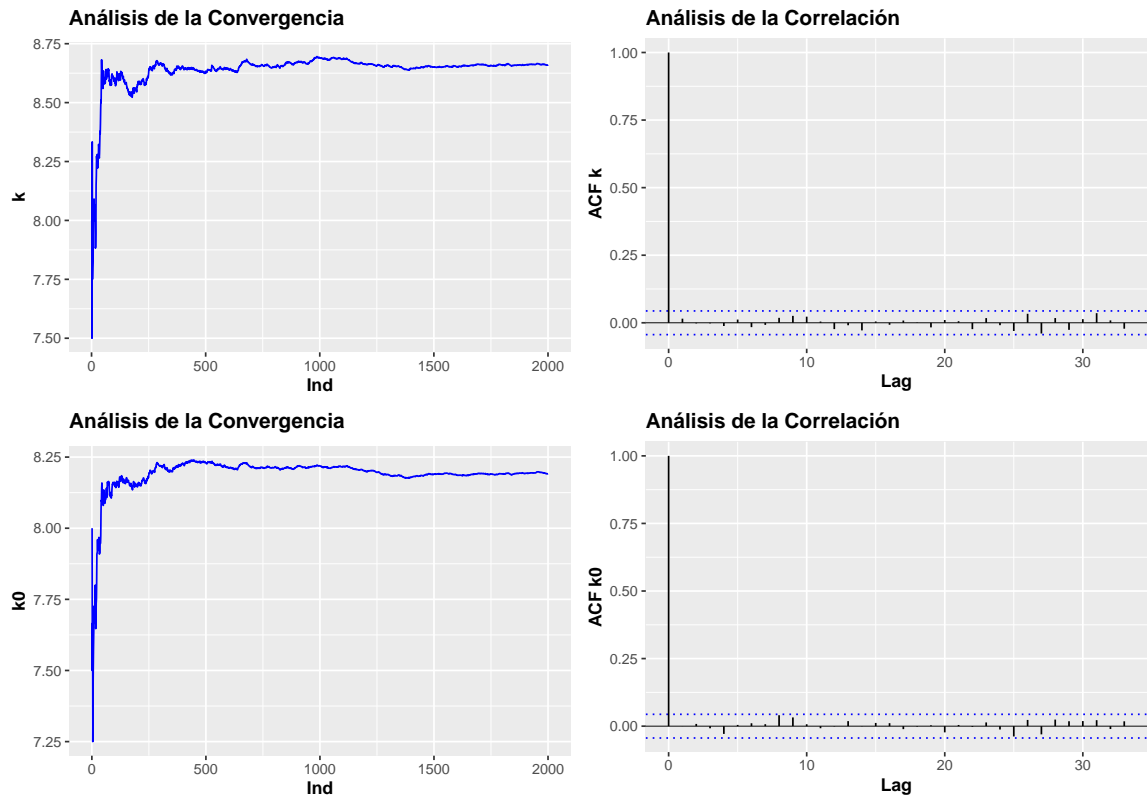


Figura 4.5: Análisis de convergencia y autocorrelación para los datos simulados 1.

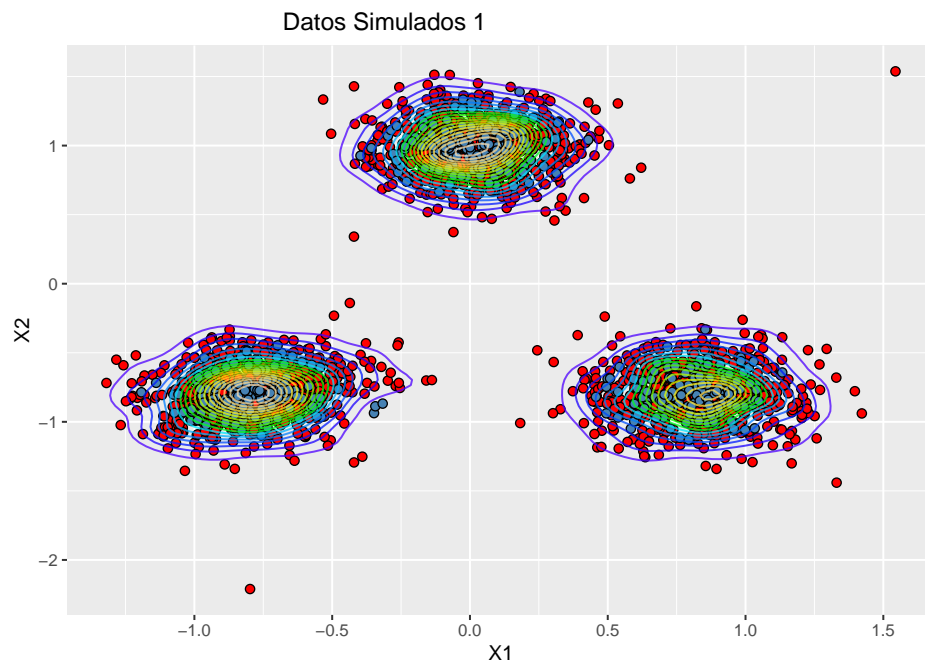


Figura 4.6: Muestra  $\mathbf{X}_{\text{Sim}1}$  (Azul) y muestra de la distribución predictiva final  $\mathbf{X}_{\text{PredSim}1}$  (Rojo) junto a las curvas de nivel de su densidad estimada.



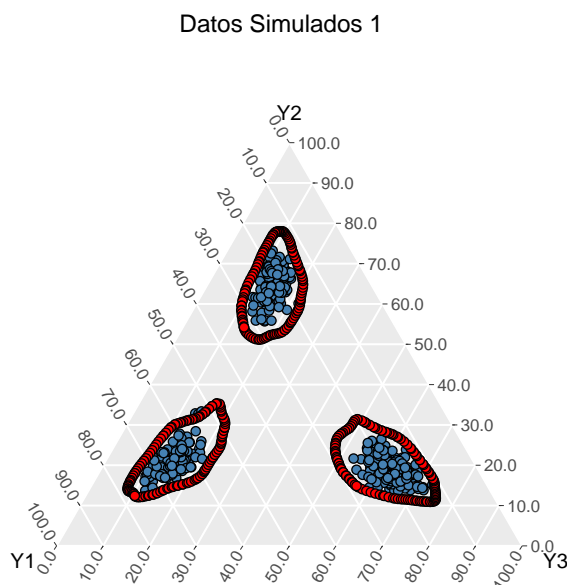


Figura 4.7: Región de probabilidad(Rojo). Muestra (simulada) de los datos originales,  $\mathbf{Y}_{\text{Sim1}}$  (Azul).

### Intervalos de Probabilidad Marginales

La región de probabilidad nos brinda información conjunta de las componentes pero en ocasiones resulta conveniente contar con información marginal, por lo que a continuación se obtendrán intervalos de probabilidad marginales para cada una de las componentes involucradas. Así, tomando marginalmente para cada componente de la muestra  $\mathbf{Y}_{\text{PredSim1}}$  los cuantiles que acumulan 2.5 % de probabilidad en sus extremos, obtenemos los límites inferior y superior de los intervalos de probabilidad marginales, estos se muestran en el Cuadro 4.3.

Intervalos de Probabilidad Marginales		
Componente	Inferior	Superior
Y1	13.42 %	72.70 %
Y2	13.71 %	72.45 %
Y3	09.34 %	69.74 %

Cuadro 4.3: Intervalos de probabilidad marginales para cada componente.

Una forma de evaluar estos intervalos es procediendo de la misma manera como se obtuvo en un principio la muestra  $\mathbf{Y}_{\text{Sim1}}$ . Así, se generó una muestra de tamaño 100,000 y la consideraremos como nuestra población, a continuación obtuvimos los correspondientes intervalos marginales, los cuales se muestran en el Cuadro 4.4.

Intervalos Poblacionales		
Componente	Inferior	Superior
Y1	14.68 %	70.38 %
Y2	14.73 %	70.39 %
Y3	10.51 %	68.41 %

Cuadro 4.4: Intervalos marginales para la población.

Se puede ver que los intervalos de probabilidad marginales obtenidos a través de la muestra de la distribución predictiva final contienen a los intervalos poblacionales.

Todos los resultados anteriores muestran que con la metodología propuesta se pueden describir adecuadamente datos composicionales con comportamientos no estándar.

### Breve justificación de la especificación inicial

Por último daremos una breve justificación a la elección de los hiperparámetros del Cuadro 4.1 que especifican las distribuciones iniciales del modelo base normal involucrado en la metodología propuesta. Desde un inicio comenzamos eligiendo hiperparámetros que determinaran distribuciones iniciales no informativas en  $\mathbb{R}^2$ , para  $\boldsymbol{\mu}$  y  $\boldsymbol{\lambda}$ , obteniendo resultados como los mostrados en la Figura 4.8, donde es evidente la alta variabilidad que contiene la muestra de la distribución predictiva final en el simplex unitario. El comportamiento anterior, se atribuye principalmente a la falta de convergencia por parte de la cadena que al proceso estocástico.

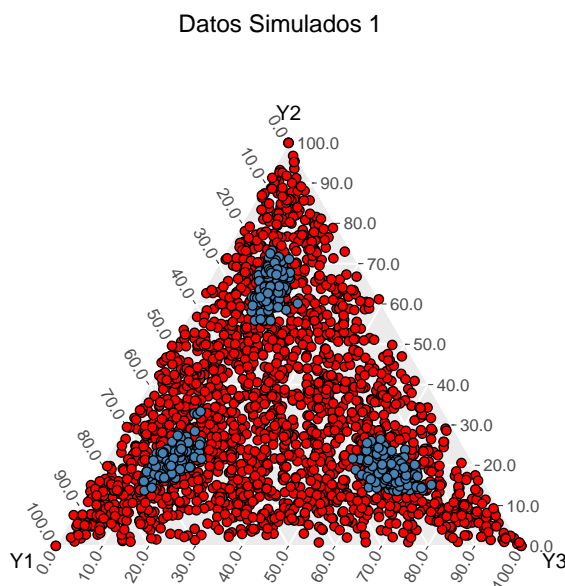


Figura 4.8: Muestra de la distribución predictiva final  $\mathbf{Y}_{\text{PredSim1}}$  (Rojo) junto a la muestra de datos originales  $\mathbf{Y}_{\text{Sim1}}$  (Azul). Análisis derivado con distribuciones iniciales no informativas en  $\mathbb{R}^2$ .

Así, como una forma de analizar el impacto de la especificación inicial en la convergencia del proceso, en la Figura 4.9 se muestran los resultados obtenidos con distri-

bucciones iniciales informativas (en  $\mathbb{R}^2$ ). En este caso se puede observar que la muestra de la distribución predictiva final tiene muy poca variabilidad.

Es importante señalar que ambos resultados se obtuvieron utilizando los mismos valores de arranque del Cuadro 4.2, entonces además de comparar la variabilidad también podemos comparar entre otras cosas el tiempo de cómputo, que resulta ser casi del doble para las iniciales no informativas que para las iniciales informativas. En este punto se debe recordar que en el problema de inferencia para modelos no-paramétricos (mezclas infinitas contables de densidades), la distribución inicial juega un papel crucial en el análisis (como queda de manifiesto en este ejemplo). Derivado del análisis anterior, se propuso establecer una especificación inicial no informativa en el simplex y no en  $\mathbb{R}^2$ . Lo anterior, derivó en las iniciales especificadas por los hiperparámetros del Cuadro 4.1. Adicionalmente, con esta especificación, se obtiene un equilibrio entre la variabilidad de los resultados mostrados en las Figuras 4.8 y 4.9, llegando incluso a mejorar entre otras cosas el tiempo de cómputo.

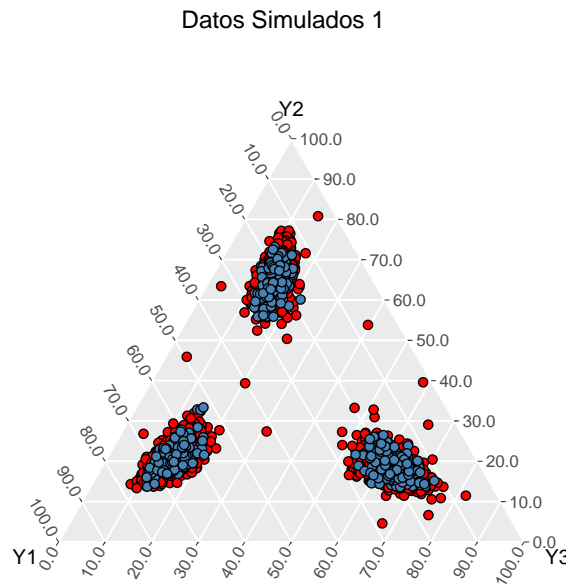


Figura 4.9: Muestra de la distribución predictiva final  $\mathbf{Y}_{\text{PredSim1}}$  (Rojo) junto a la muestra  $\mathbf{Y}_{\text{Sim1}}$  (Azul) con distribuciones iniciales informativas.

## Ejemplo 2: Un mezcla de Gammas trivariadas

Para este segundo ejemplo, se simuló un conjunto de datos a partir del siguiente modelo de mezclas de 2 componentes de distribuciones Gammas Proyectadas trivariadas (Ver, Núñez-Antonio y Geneyro, 2019).

$$\boldsymbol{\theta} \sim f(\boldsymbol{\theta}|\boldsymbol{\omega}, \boldsymbol{\alpha}, \boldsymbol{\beta}) \quad \text{donde,}$$

$$f(\boldsymbol{\theta}|\boldsymbol{\omega}, \boldsymbol{\alpha}, \boldsymbol{\beta}) = \sum_{i=1}^3 \omega_i GP_3(\boldsymbol{\theta}|\boldsymbol{\alpha}_i, \boldsymbol{\beta}_i),$$

con parámetros dados por:

$$\begin{aligned} \text{Pesos: } \omega_1 &= 0.45, \omega_2 = 0.55, \\ \text{Alfa's: } \boldsymbol{\alpha}_1 &= \begin{pmatrix} 5 \\ 2 \\ 0.2 \end{pmatrix} \text{ y } \boldsymbol{\alpha}_2 = \begin{pmatrix} 5 \\ 3 \\ 0.1 \end{pmatrix}, \\ \text{Beta's: } \boldsymbol{\beta}_1 &= \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \text{ y } \boldsymbol{\beta}_2 = \begin{pmatrix} 1 \\ 0.05 \\ 0.01 \end{pmatrix}. \end{aligned}$$

### Muestra

A partir de la especificación anterior, se generó una muestra de 250 datos composicionales a la que se le denominará  $\mathbf{Y}_{\text{Sim2}}$ , la cual se puede visualizar en el diagrama ternario de la Figura 4.10. Cabe señalar que dichos datos composicionales son ternas positivas de  $\boldsymbol{\theta}$ 's bajo el operador clausura.

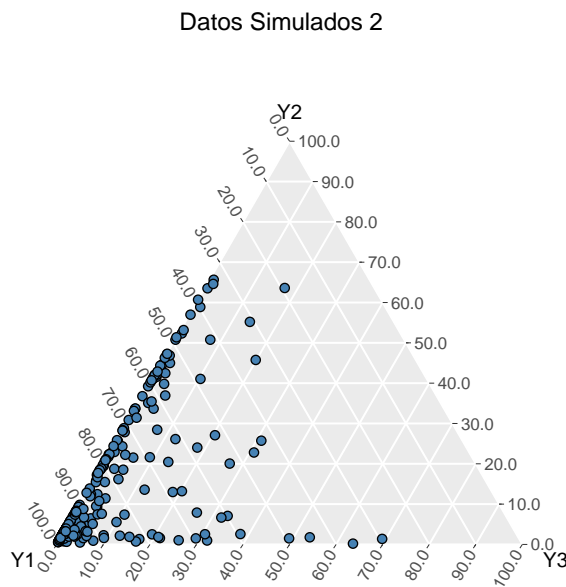


Figura 4.10: Muestra  $\mathbf{Y}_{\text{Sim2}}$  (Azul).

Se debe señalar que este conjunto de datos describen el comportamiento de datos composicionales con un patrón bastante extremo, el cual trataremos de modelar con la metodología propuesta en este trabajo.

Como primer paso, mapeamos la muestra de  $\mathbb{S}^3$  a  $\mathbb{R}^2$  mediante la transformación  $ilr$ , es decir, hacemos  $\mathbf{X}_{\text{Sim2}} = ilr(\mathbf{Y}_{\text{Sim2}})$ . Una vez en  $\mathbb{R}^2$  podemos visualizarla en el diagrama de dispersión de la Figura 4.11 y podemos especificar todos los parámetros necesarios para llevar a cabo el algoritmo con el propósito de obtener nuevamente una muestra de la distribución predictiva final.

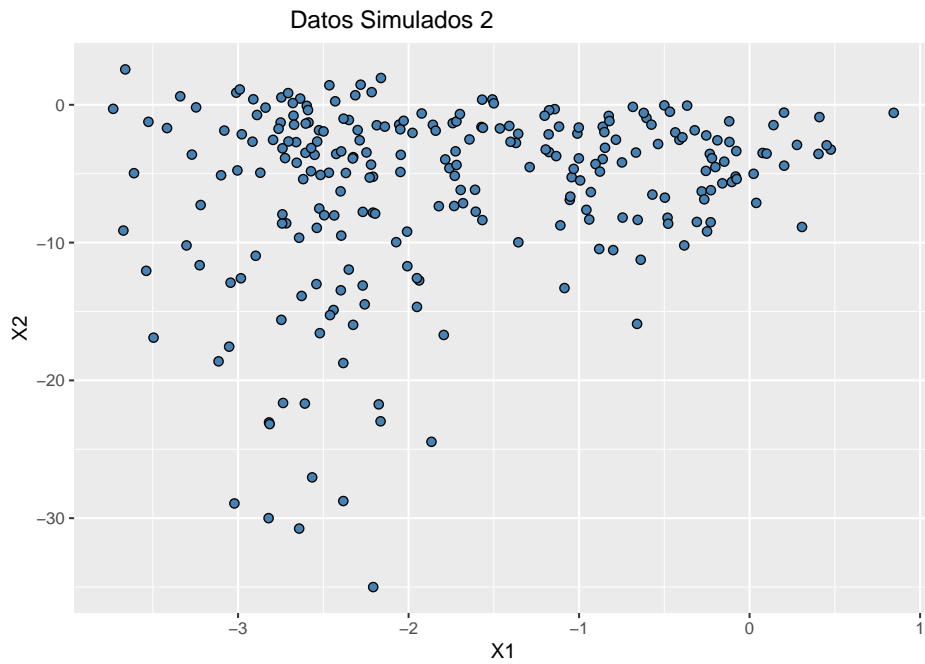


Figura 4.11: Muestra  $\mathbf{X}_{\text{Sim2}}$ (Azul).

## Distribución Predictiva

Decidimos utilizar los mismo valores de arranque para todos los parámetros del algoritmo, es decir, utilizamos los valores que están contenidos en los Cuadros 4.1 y 4.2 con la intención de ser consistentes. Entonces, tras un tiempo de cómputo de 21.6 horas el algoritmo recorre 200,000 iteraciones, obteniendo una muestra de tamaño 2,000 tanto de la distribución predictiva final en  $\mathbb{R}^2$  que llamaremos  $\mathbf{X}_{\text{PredSim2}}$ , como de valores  $k$  y  $k_0$  que representan el número de componentes muestreados y no vacíos, respectivamente, en cada paso del algoritmo. La Figura 4.12 presenta la distribución predictiva final junto la muestra original  $\mathbf{X}_{\text{Sim2}}$ .

Posteriormente, bajo la transformación  $ilr^{-1}$  podemos obtener a  $\mathbf{Y}_{\text{PredSim2}} = ilr^{-1}(\mathbf{X}_{\text{PredSim2}})$ , misma que podemos visualizar junto a la muestra  $\mathbf{Y}_{\text{Sim2}}$  en la Figura 4.13. De la Figura 4.13 se puede observar que, la muestra de la distribución predictiva final presenta gran variabilidad.

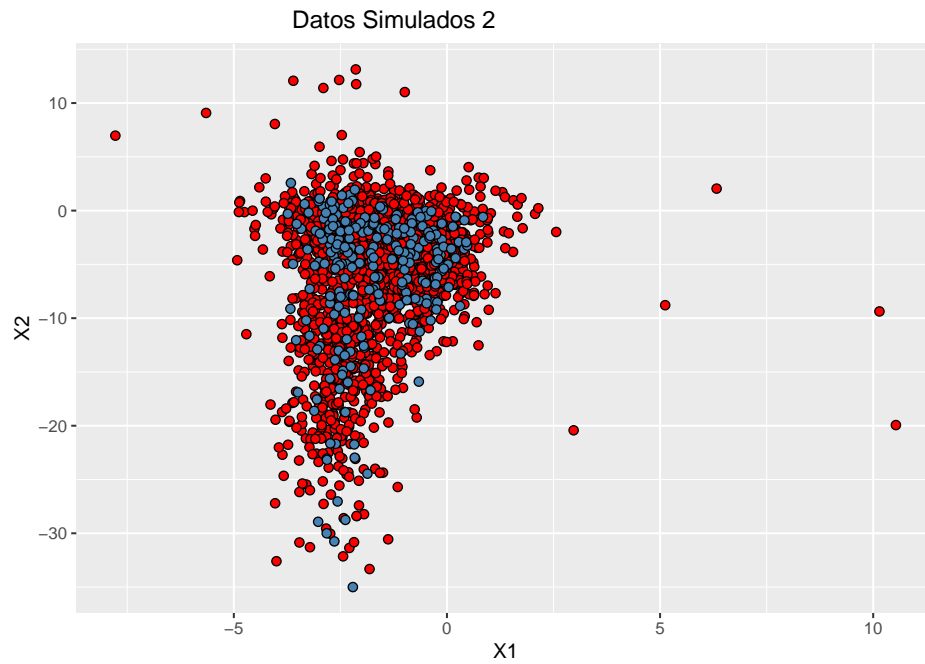


Figura 4.12: Muestra de la distribución predictiva final  $\mathbf{X}_{\text{PredSim2}}$  (Rojo) junto con la muestra original  $\mathbf{X}_{\text{Sim2}}$  (Azul).

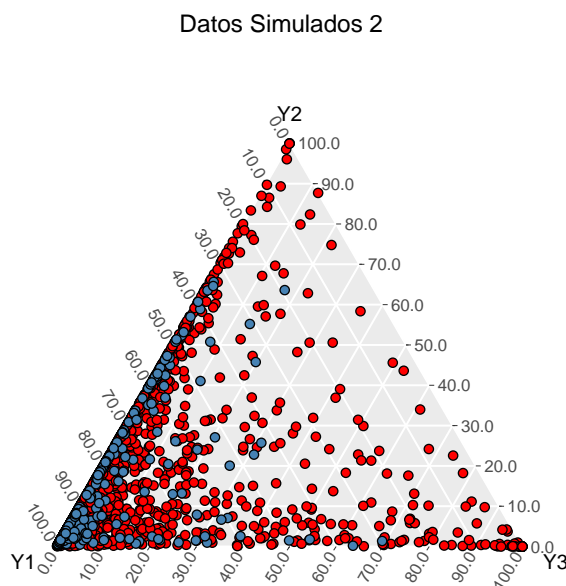


Figura 4.13: Muestra de la distribución predictiva final  $\mathbf{Y}_{\text{PredSim2}}$  (Rojo) junto a la muestra  $\mathbf{Y}_{\text{Sim2}}$  (Azul).

### Análisis de Convergencia y Correlación

En la Figura 4.14 se analizan gráficamente los promedios ergódicos y autocorrelaciones de  $k$  y  $k_0$ , entonces podemos observar que la autocorrelación entre los valores de las muestras para ambos parámetros en general es pequeña, la inestabilidad de los

promedios ergódicos indica la falta de convergencia por parte de la cadena. Esto impacta en la variabilidad de la muestra de la distribución predictiva final, quizás en un futuro para muestras con patrones tan extremos sea conveniente aumentar el número de iteraciones tanto de calentamiento como de convergencia para mejorar dichos promedios, así como, aumentar el *Lag* con el que se toman los valores provenientes de la cadena, para disminuir aún más su autocorrelación.

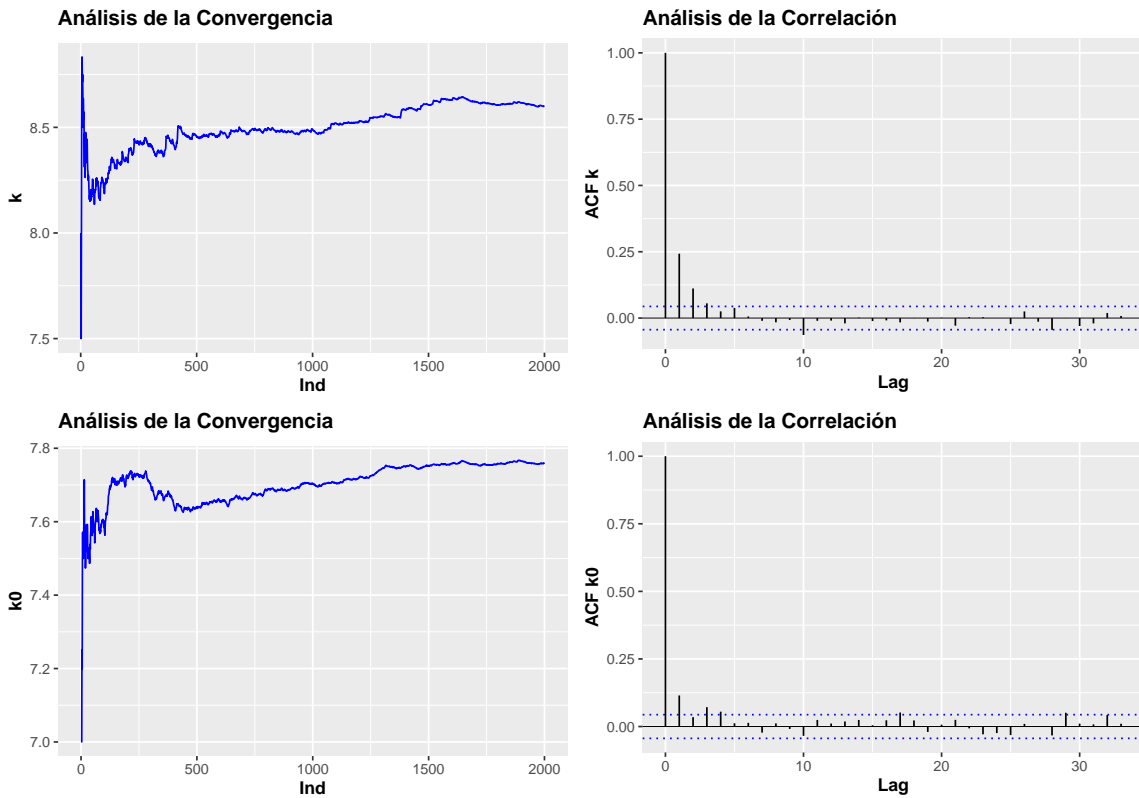


Figura 4.14: Análisis de convergencia y autocorrelación para los datos simulados 2.

### Estimadores Puntuales

Al principio de esta sección mencionamos el interés particular para este conjunto de datos simulados, que es poder recuperar ciertas características de la población de la que proviene. Entonces, así como se obtuvo la muestra  $\mathbf{Y}_{\text{Sim2}}$ , se generó una muestra de tamaño 100,000, la cual se consideró como nuestra población y obtuvimos la media poblacional, la cual es reportada en el Cuadro 4.5.

Media Poblacional		
Componente		
Y1	Y2	Y3
92.14 %	07.84 %	00.02 %

Cuadro 4.5: Media poblacional para cada componente.

Posteriormente, a partir de la muestra  $\mathbf{X}_{\text{PredSim2}}$  se obtienen los correspondientes estimadores puntuales, Cuadro 4.6. A continuación se da una breve descripción de este proceso.

- Media: Se obtiene sacando la media marginal, es decir, por componente de  $\mathbf{X}_{\text{PredSim2}}$ , para después pasar dicho valor al simplex bajo la transformación  $ilr^{-1}$ . Nótese que resulta ser el centro muestral para datos composicionales.
- Mediana: Se obtiene sacando la mediana marginal, es decir, por componente de  $\mathbf{X}_{\text{PredSim2}}$ , para después pasar dicho valor al simplex bajo la transformación  $ilr^{-1}$ .

Estimadores Puntuales			
Estimador	Componente		
	Y1	Y2	Y3
Media	91.58 %	08.40 %	00.02 %
Mediana	92.76 %	07.02 %	00.22 %

Cuadro 4.6: Estimadores puntuales propuestos para cada componente.

En los diagramas ternarios de la Figura 4.15 se representan visualmente estos estimadores puntuales junto con la media poblacional y la muestra  $\mathbf{Y}_{\text{Sim2}}$ . Podemos observar que la media resulta estar más cerca del valor poblacional, entonces la metodología propuesta al menos para tres componentes está siendo capaz de recuperar características importantes en poblaciones con un comportamiento un tanto extremo.

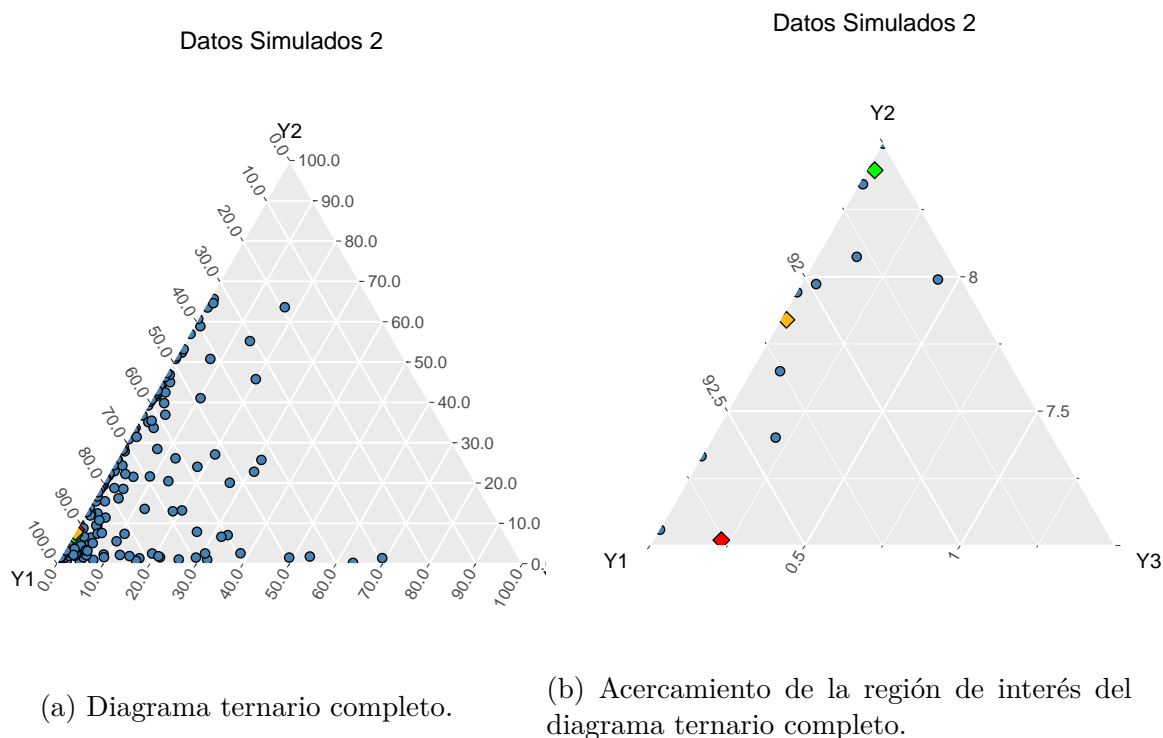


Figura 4.15: Media poblacional (Dorado) y muestra  $\mathbf{Y}_{\text{Sim2}}$  (Azul). Estimadores puntuales: media composicional (en Verde) y mediana marginal (en Rojo).



## 4.2. Datos Reales

### Conteo Rápido

El Conteo Rápido es un ejercicio científico estadístico, cuyo principal objetivo es contribuir a dar certidumbre, confianza y transparencia a un proceso electoral, ofreciendo una estimación de las tendencias de la votación de alta calidad estadística la misma noche de la elección [Estok et al., 2002]. Este ejercicio nos permite conocer con mucha rapidez a tan solo pocas horas de finalizada la votación, cómo serán los resultados al término del conteo de todos los votos.

Este tipo de ejercicios estadísticos es realizado con regularidad en muchos países. En México el primer ejercicio de este tipo se llevo a cabo hace más de dos décadas y tuvo por objetivo estimar los resultados de la elección presidencial de 1994. Inicialmente se decidió implementar conteos rápidos para estimar las tendencias de la votación en cada elección presidencial, sin embargo, en el año 2016 el Consejo General del Instituto Nacional Electoral, decide organizar conteos rápidos tanto para elecciones federales como locales.

En esta sección, aplicaremos la metodología propuesta para estimar las tendencias de la votación para la gubernatura del estado de Morelos llevada a cabo en 2018. Los resultados serán brevemente comparados con los obtenidos por el Comité Técnico de Conteos Rápidos coordinado por el INE [General, 2018], con el objetivo de motivar e impulsar los retos que más adelante se puntualizaran en las conclusiones de este trabajo de tesis.

### Elecciones para la Gubernatura del Estado de Morelos, 2018

El 01 de julio de 2018 en el Estado de Morelos en México, se llevaron a cabo las votaciones para la gubernatura del estado. Con la finalidad de conocer con rapidez las tendencias de los resultados al término del conteo de los votos, el INE nombró científicos expertos en matemáticas y estadística, para crear el Comité Técnico Asesor de Conteo Rápido (COTECORA), quien fue el encargado de llevar a cabo el ejercicio del conteo rápido. De las 200 casillas que integran la muestra construida por el COTECORA, se recibió información de 160 casillas, las cuales representan un 80 % de la muestra total y con las cuales se llevaron a cabo las estimaciones correspondientes. En este trabajo, nosotros nos basaremos en dicha información, la cual se encuentra disponible en el sitio web del INE [Muestra de Morelos, 2018].

### Muestra

Realizamos un tratamiento a la base original para unificar las variables de interés y así poder formar una base de datos composicionales de 3 componentes. Lo anterior, aglomerando los partidos políticos en tres coaliciones (componentes) definidas de la siguiente manera:

- Coalición MORENA: Incluye los votos para Morena, Partido Encuentro Social, Partido del Trabajo, y sus combinaciones.

- Coalición PAN: Incluye los votos para el Partido Acción Nacional, Movimiento Ciudadano y su combinación.
- Coalición Otros: Incluye los votos del Partido Revolucionario Institucional, Partido de la Revolución Democrática, Partido Social Demócrata de Morelos, Partido Verde Ecologista de México, Partido Nueva Alianza, Partido Humanista, Candidatos Independientes, Votos Nulos, Candidatos No Registrados y sus combinaciones en los casos en las que los haya.

La nueva base conformada por estas 3 componentes fue sometida al operador clausura para llevarla al simplex de dimensión tres  $S^3$  en donde recibió el nombre de  $\mathbf{Y}_{\text{Mor}}$ . Esta muestra, se puede visualizar en el diagrama ternario de la Figura 4.16. Posteriormente bajo la transformación  $ilr$  se llevo al espacio real  $\mathbb{R}^2$  donde recibió el nombre de  $\mathbf{X}_{\text{Mor}}$ , es decir,  $\mathbf{X}_{\text{Mor}} = ilr(\mathbf{Y}_{\text{Mor}})$  y podemos visualizarla en el diagrama de dispersión de la Figura 4.17.

### Distribución Predictiva

Un vez más, con la intención de ser consistentes, los valores de arranque a utilizar en el algoritmo para este conjunto de datos son los mismos que se utilizaron en los ejemplos con datos simulados, es decir, los contenidos en los Cuadros 4.1 y 4.2. Así, tras un tiempo de cómputo de 1.7 horas, el algoritmo recorre 200,000 iteraciones, obteniendo una muestra de tamaño 2,000 tanto de la distribución predictiva final en  $\mathbb{R}^2$  que llamaremos  $\mathbf{X}_{\text{PredMor}}$ , como de valores  $k$  y  $k_0$  que representan el número de componentes muestreados y no vacíos respectivamente en cada paso del algoritmo. La muestra de la distribución predictiva final puede ser visualizada, junto a la muestra  $\mathbf{X}_{\text{Mor}}$ , en el diagrama de dispersión de la Figura 4.18. Posteriormente, bajo la transformación  $ilr^{-1}$  podemos obtener a  $\mathbf{Y}_{\text{PredMor}} = ilr^{-1}(\mathbf{X}_{\text{PredMor}})$ , misma que podemos visualizar junto a la muestra  $\mathbf{Y}_{\text{Mor}}$  en el diagrama ternario de la Figura 4.19.

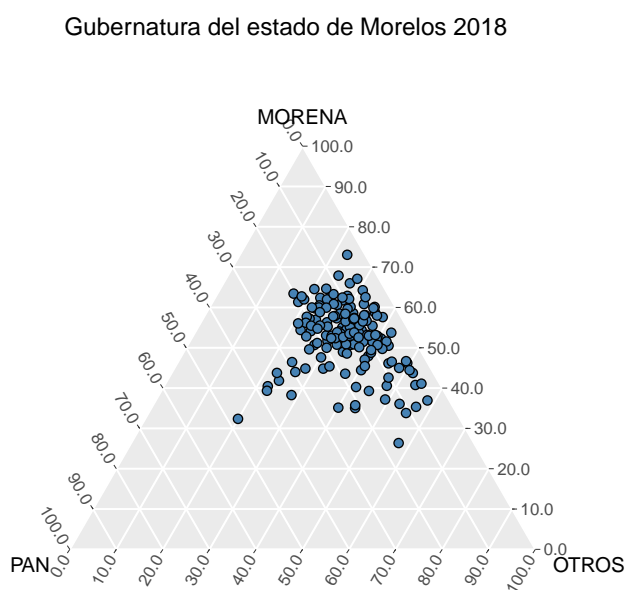


Figura 4.16: Muestra de votos,  $\mathbf{Y}_{\text{Mor}}$  (Azul).

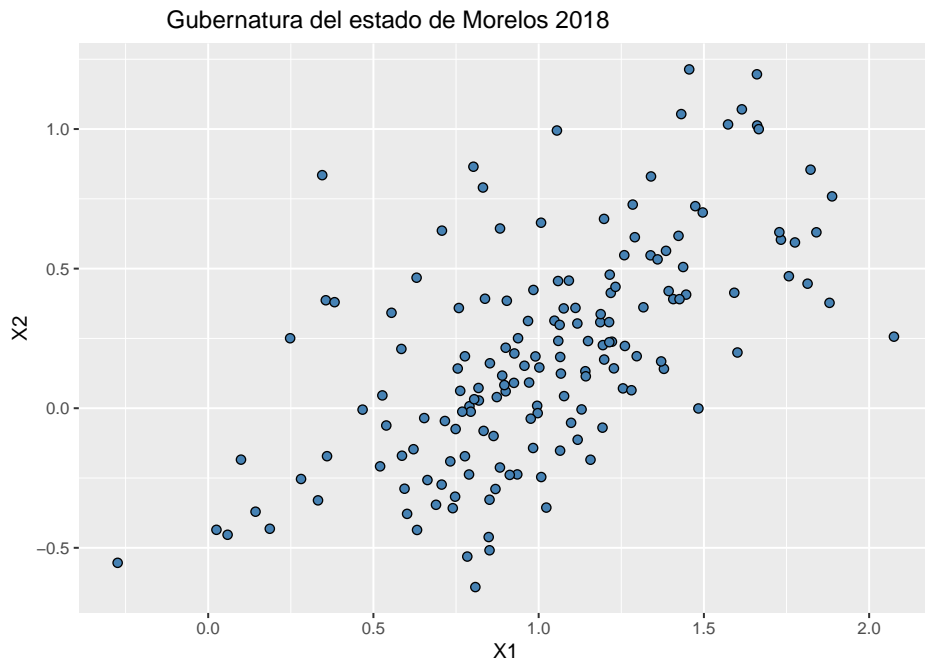


Figura 4.17: Muestra de votos  $\mathbf{X}_{\text{Mor}}(\text{Azul})$ , en  $\mathbb{R}^2$ .

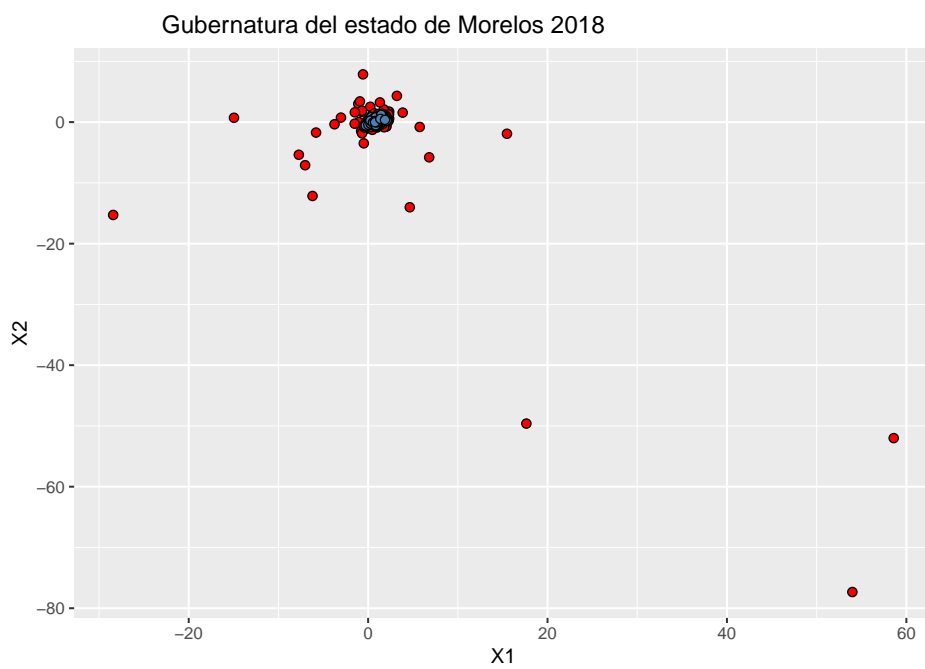


Figura 4.18: Muestra de la distribución predictiva final  $\mathbf{X}_{\text{PredMor}}(\text{Rojo})$ . Muestra de votos de casillas  $\mathbf{X}_{\text{Mor}}(\text{Azul})$ .

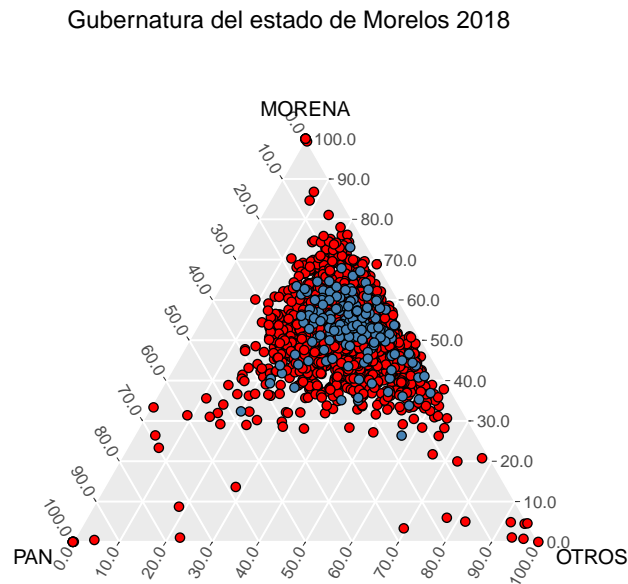


Figura 4.19: Muestra de la distribución predictiva final  $Y_{PredMor}$  (Rojo) junto a la muestra  $Y_{Mor}$  (Azul).

Podemos observar que la muestra de la distribución predictiva final presenta gran variabilidad, pues existen algunos puntos que se alejan demasiado estando incluso en los lados y vértices del diagrama ternario. A continuación se analiza la convergencia y autocorrelación del proceso, para ver si están pueden estar involucrados en dicho comportamiento.

### Análisis de Convergencia y Correlación

En la Figura 4.20 se analizan gráficamente los promedios ergódicos y autocorrelaciones de  $k$  y  $k_0$  como herramienta para monitorear la convergencia de la cadena y, la autocorrelación entre los valores de la muestra obtenida por el algoritmo. Se puede observar que la autocorrelación entre los valores de las muestras para ambos parámetros es pequeño y que a pesar de que los promedios ergódicos muestran pequeñas fluctuaciones, estas son a una escala muy pequeña. Derivado de lo anterior, concluimos que la cadena ha alcanzado la convergencia.

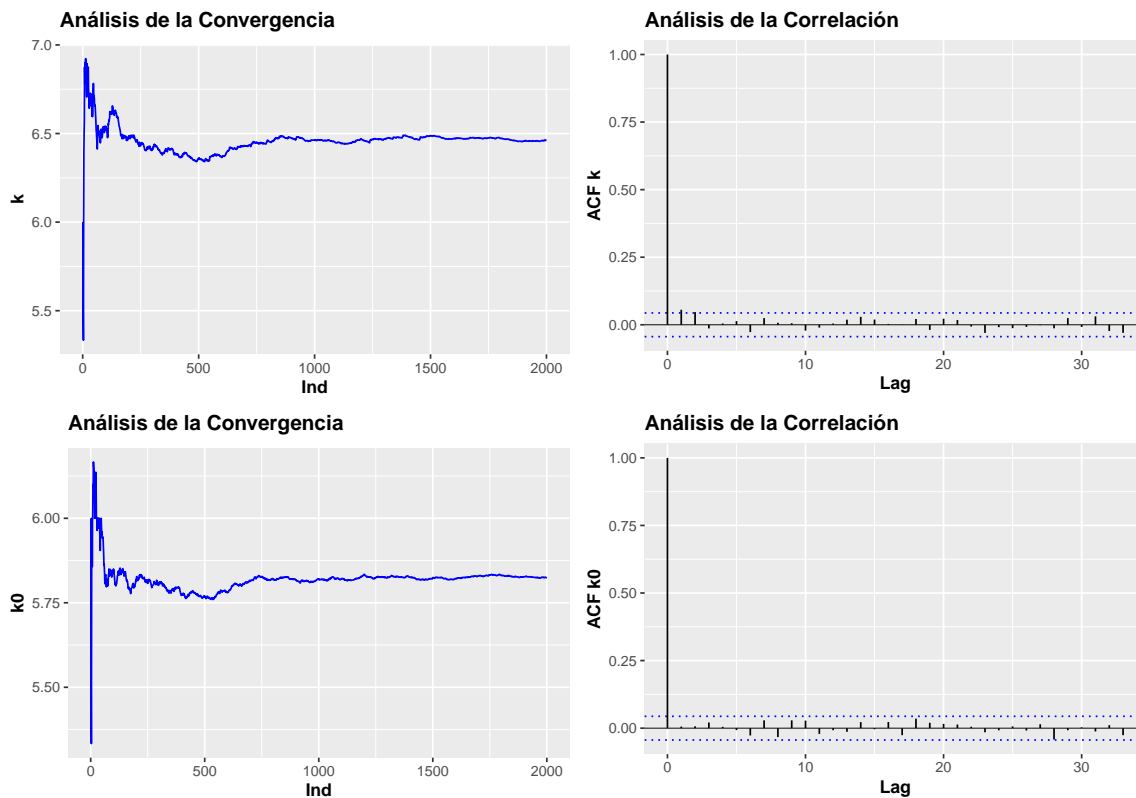


Figura 4.20: Análisis de convergencia y autocorrelación para los datos de la gubernatura del estado de Morelos.

### Estimadores Puntuales

En el Cuadro 4.7 se presentan los resultados del cómputo distrital para este ejercicio electoral, es decir, los resultados reales después de sumar los resultados contenidos en las actas de escrutinio y cómputo de cada una de las casillas [Cómputos en Morelos, 2018].

Cómputo Distrital		
Coalición		
MORENA	PAN	OTROS
52.59 %	14.05 %	33.36 %

Cuadro 4.7: Cómputo distrital para las coaliciones.

Entonces a través de la muestra  $\mathbf{X}_{\text{PredMor}}$  obtenemos los mismos estimadores puntuales propuestos en el segundo ejemplo con datos simulados, estos se pueden observar en el Cuadro 4.8.

Estimadores Puntuales			
Estimador	Coalición		
	MORENA	PAN	OTROS
Media	56.72 %	13.10 %	30.18 %
Mediana	53.61 %	12.84 %	33.55 %

Cuadro 4.8: Estimadores puntuales propuestos para las coaliciones.

En los diagramas ternarios de la Figura 4.21 se representan visualmente estos estimadores puntuales junto con el cómputo distrital y la muestra  $\mathbf{Y}_{\text{Mor}}$ . En este caso la mediana es la que resulta estar más cerca al cómputo distrital. De lo anterior, se puede concluir que con la metodología propuesta estamos en posición de poder brindar un buen estimador puntual que respete la restricción de la suma constante, la cual es la restricción natural de este tipo de datos.

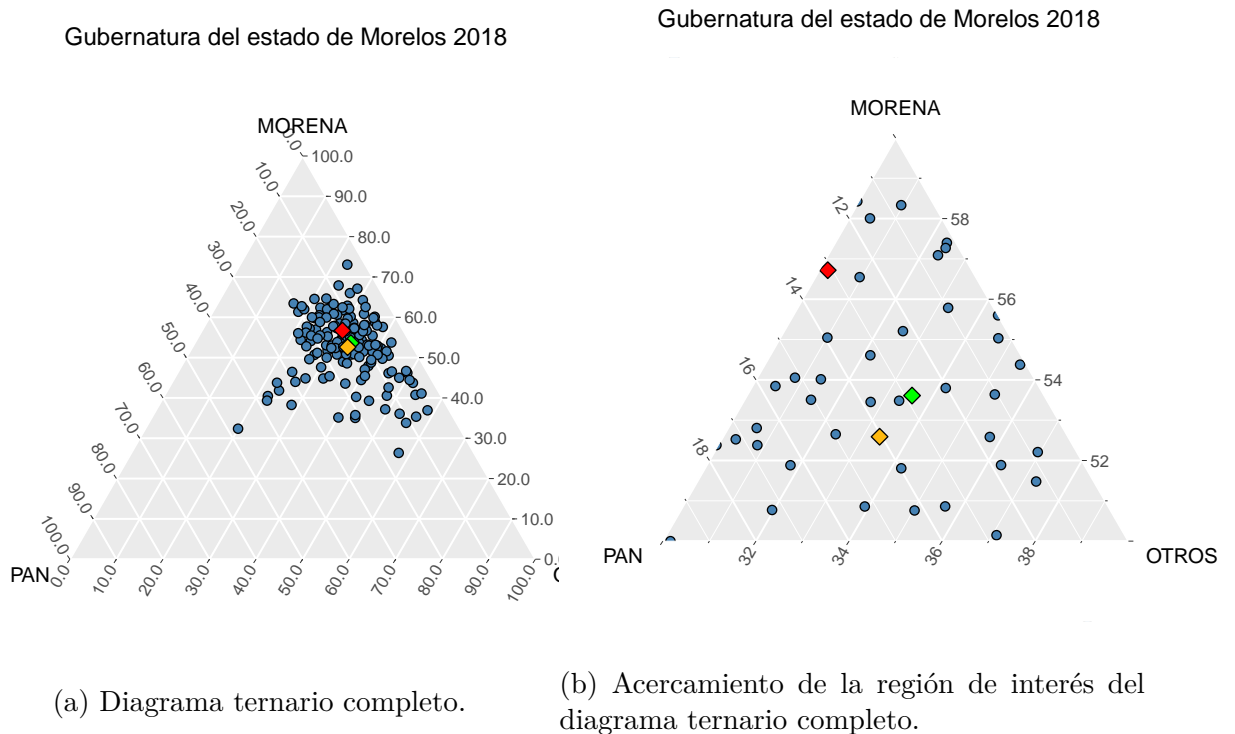


Figura 4.21: Cómputo distrital (Dorado) y muestra de casillas  $\mathbf{Y}_{\text{Mor}}$  (Azul). Estimadores puntuales: Media composicional (Rojo) y mediana marginal (Verde).

### Intervalos de Probabilidad Marginales

Una vez que hemos sido capaces de obtener un estimador puntual para la variable de interés que es la proporción de votos para cada coalición, nuestro próximo objetivo será brindar una región de probabilidad conjunta e intervalos de probabilidad marginales. Para ello se modificó el algoritmo de la siguiente manera: En cada iteración (en equilibrio) del algoritmo, se generó una muestra aleatoria de 1,000 realizaciones del modelo (vectores reales de dimensión dos) y para esta muestra se calculó su mediana. Entonces al finalizar el algoritmo se obtuvo una muestra de 2,000 medianas en  $\mathbb{R}^2$  que llamaremos  $\Delta_{\text{PredMor}}$  y cuya visualización junto a la muestra  $\mathbf{X}_{\text{Mor}}$  se encuentra en el diagrama de dispersión de la Figura 4.22.

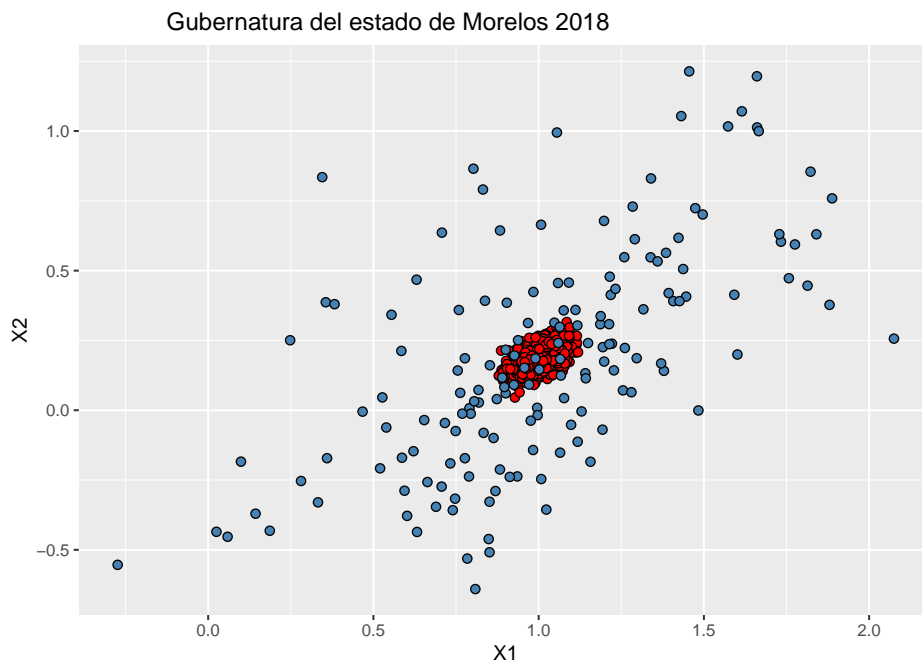


Figura 4.22: Muestra de medianas  $\Delta_{\text{PredMor}}$  (Rojo). Muestra de casillas  $\mathbf{X}_{\text{Mor}}$  (Azul).

En los diagramas ternarios de la Figura 4.23 se puede visualizar la muestra  $\mathbf{Y}_{\text{Mor}}$ , el cómputo distrital y la nube de medianas llevadas al simplex bajo la transformación  $ilr^{-1}$ . Podemos ver que el cómputo distrital está contenido dentro de la nube de medianas, entonces tomando marginalmente para cada componente de esta muestra de medianas en el simplex los cuantiles que acumulan 2.5% de probabilidad en sus extremos obtenemos los límites inferior y superior de los intervalos de probabilidad marginales mostrados en el Cuadro 4.9.

Intervalos de Probabilidad Marginales		
Coalición	Inferior	Superior
MORENA	52.20 %	55.31 %
PAN	11.80 %	14.22 %
OTROS	31.70 %	34.96 %

Cuadro 4.9: Intervalos de probabilidad marginales para las coaliciones.

Satisfactoriamente dichos intervalos contiene al cómputo distrital para cualquiera de las tres coaliciones.

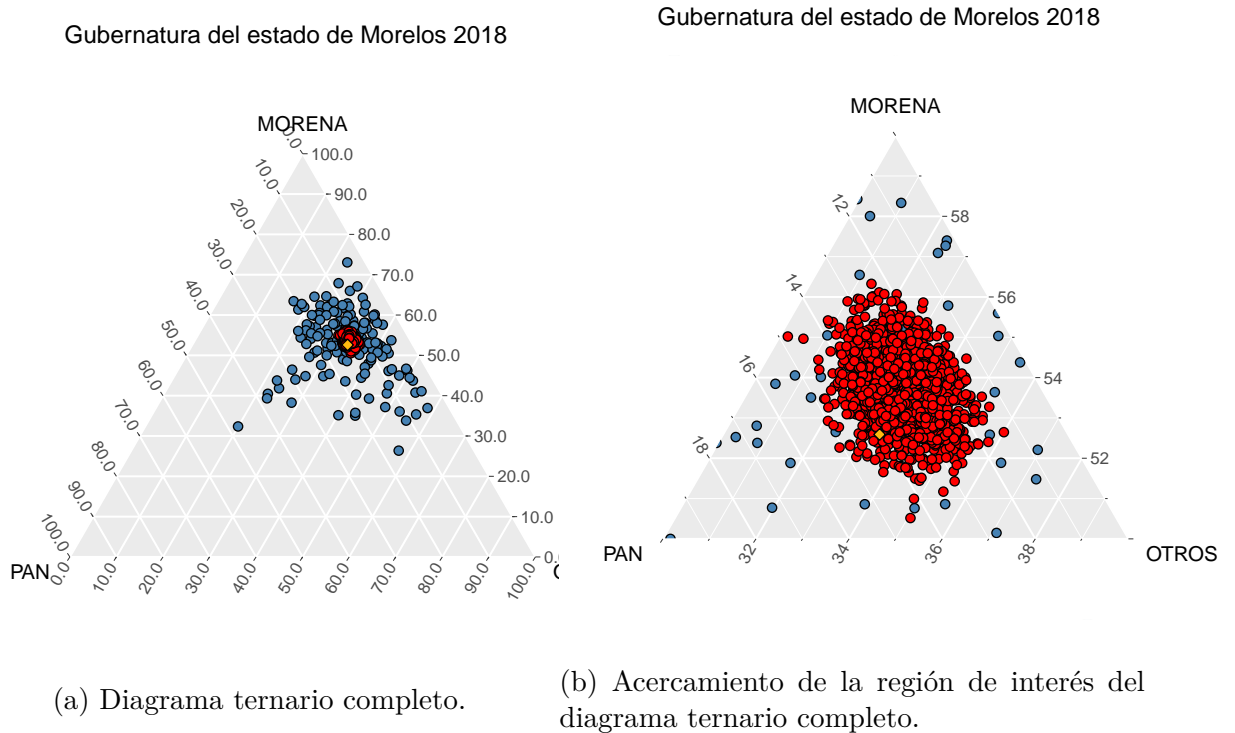


Figura 4.23: Muestra de medianas  $\Delta_{\text{PredMor}}$  (Rojo) bajo la transformación *ilr*, muestra de casillas  $Y_{\text{Mor}}$  (Azul) y cómputo distrital (Dorado).

### Región de Probabilidad

Por último se obtiene la región de probabilidad de la misma forma en que se obtuvo para el primer conjunto de datos simulados, en los diagramas ternarios de la Figura 4.24 se puede visualizar dicha región. La región de probabilidad contiene al cómputo distrital tal como se ve en la Figura 4.24a. De lo anterior, se desprende que con la metodología propuesta en este trabajo, se pueden obtener estimaciones razonables tanto de manera puntual como en forma de región e intervalos marginales, para el verdadero valor de los parámetros involucrados, es decir, para la proporción real de votos de las fuerzas políticas involucradas en este tipo de ejercicios Electorales.

### Breve comparación con los resultados del INE

En el Cuadro 4.10 mostramos dos de los intervalos del conteo rápido brindados por el INE la noche de la elección, ya que estos son comparables con los intervalos de probabilidad marginales mostrados en el Cuadro 4.9.

Intervalos del INE		
Coalición	Inferior	Superior
MORENA	51.00 %	53.80 %
PAN	13.40 %	16.10 %

Cuadro 4.10: Resultados del conteo rápido realizado por el INE.

Las coaliciones comparables en este caso son MORENA y PAN. Se puede observar que el respectivo cómputo distrital correspondiente a cada una de estas dos coaliciones



está contenido en el correspondiente intervalo de probabilidad derivado de la aplicación de nuestra metodología. Además, la diferencia porcentual más grande para nuestros intervalos es de 3.11 % puntos porcentuales, mientras que la de los intervalos del INE es de 2.8 % puntos porcentuales. Esto quiere decir, que los intervalos obtenidos mediante la metodología propuesta exceden a lo más en 0.31 % puntos porcentuales a los intervalos dados por el COTECORA en el conteo rápido, lo cual consideramos que es una precisión razonable.

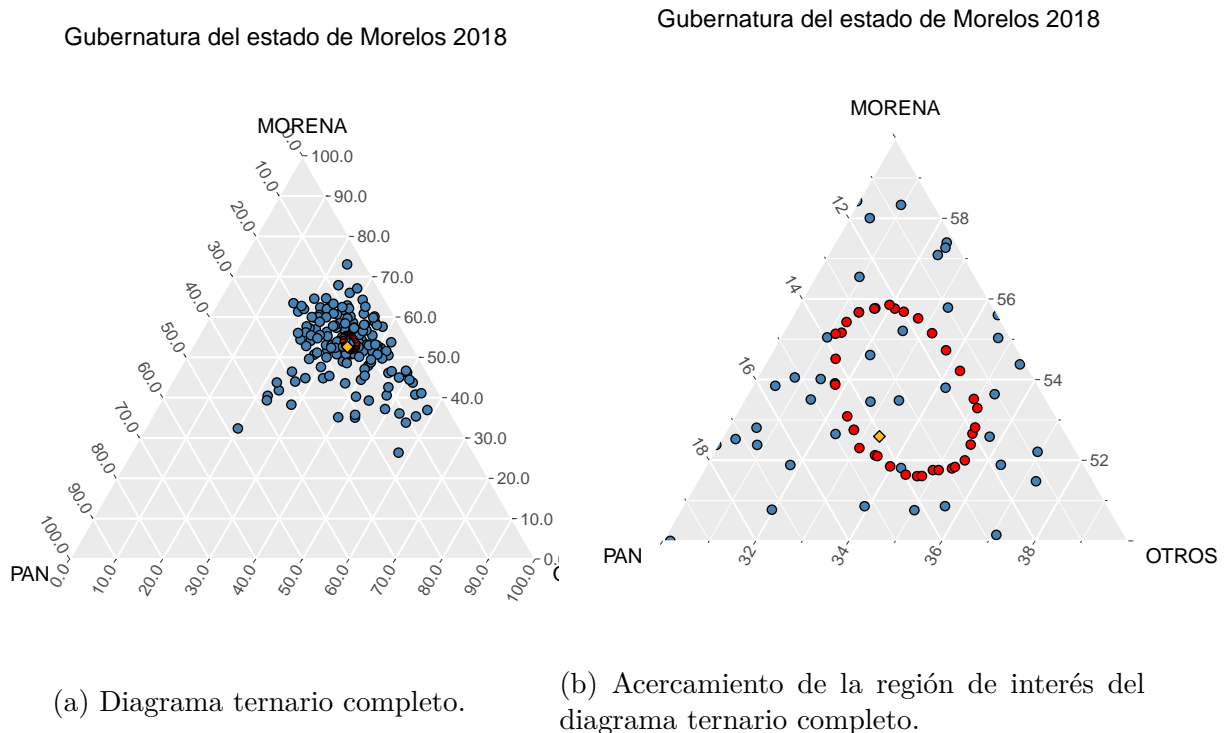


Figura 4.24: Región de probabilidad(Rojo) y cómputo distrital(Dorado) junto a la muestra  $Y_{Mor}$ (Azul).

## Conclusión

Los resultados obtenidos permiten concluir que la metodología propuesta nos permite obtener inferencias adecuadas en poblaciones con patrones poco convencionales y extremos. Del mismo modo permite inferir adecuadamente sobre características de interés de un problema específico, como lo es la proporción de votos en ejercicios como los Conteo Rápidos.



# Conclusiones y Perspectivas

Como se ha dejado ver en este trabajo, desde el punto de vista metodológico, el análisis de datos composicionales representa un reto mayúsculo. La restricción de la suma constante y la correlación negativa entre las variables son de las características más importantes que hay que tomar en cuenta para su estudio. Esto último ya que cualquier perturbación en alguna de las componentes genera cambios directamente en las demás componentes. La metodología propuesta en este trabajo de tesis toma en cuenta la naturaleza topológica de los datos, es decir, nunca pierde de vista que los datos son composicionales, respetando en todo momento estos aspectos fundamentales además de que lo hace desde el enfoque de la estadística Bayesiana.

El análisis e inferencia de los modelos de mezclas infinitas de distribuciones desde el enfoque bayesiano representa un tema complejo ya que, aunque existen algoritmos para implementarlo, se siguen haciendo propuestas metodológicas para mejorar los tiempos de convergencia. El modelo no-parámtrico (basado en mezclas infinitas de distribuciones normales univariadas y multivariadas) fue viable gracias al uso de la transformación log-cociente *ilr* que mapea el simplex en el espacio real. Sin embargo, el simplex es un subconjunto del mismo espacio real ya que sus elementos son vectores con entradas reales no negativas. Resultaría interesante poder contrastar los resultados obtenidos en este trabajo de tesis, con una metodología que considere modelos de mezclas con distribuciones definidas en ortantes positivos de tal manera que podamos omitir cualquier tipo de transformación log-cociente.

Con la metodología propuesta se esta en condiciones de ofrecer, no solo una muestra de la distribución predictiva final, sino también regiones de probabilidad, intervalos de probabilidad marginales y estimadores puntuales, para cada componente involucrada en el vector composicional. Estos últimos cumpliendo con la restricción de que sus partes suman uno, lo cual no ocurre si se aplican otros métodos estadísticos convencionales, como los que se utilizan actualmente en los ejercicios de conteo rápido realizados por el Instituto Nacional Electoral (INE).

Los resultados obtenidos derivados del análisis de los datos reales sobre procesos electorales, son motivantes. Lo anterior, en el sentido de lo cerca que están de los obtenidos por el INE, con metodologías donde no se respeta su naturaleza composicional, impulsando así la propuesta y desarrollo de metodologías que involucren un adecuado análisis de datos composicionales.

Aunque quedan cosas por hacer, consideramos que la propuesta desarrollada en este trabajo de tesis es fundamental y pionera en el análisis Bayesiano de datos composicionales.



# Bibliografía

- [Aitchison, 1982] Aitchison, J. (1982). The statistical analysis of compositional data. *Journal of the Royal Statistical Society: Series B (Methodological)*, 44(2):139–160. V, VII, 4, 8
- [Aitchison, 2003] Aitchison, J. (2003). A concise guide to compositional data analysis. In *CDA Workshop, Girona*. 11
- [Aldrich et al., 1995] Aldrich, J. et al. (1995). Correlations genuine and spurious in pearson and yule. *Statistical science*, 10(4):364–376. 6, 7
- [Bernardo and Smith, 2009] Bernardo, J. M. and Smith, A. F. (2009). *Bayesian theory*, volume 405. John Wiley & Sons. 17, 35
- [Billheimer, 2001] Billheimer, D. (2001). Modelado de receptor composicional. *Environmetrics: el diario oficial de la International Environmetrics Society*, 12(5):451–467. 5
- [Blackwell et al., 1973] Blackwell, D., MacQueen, J. B., et al. (1973). Ferguson distributions via pólya urn schemes. *The annals of statistics*, 1(2):353–355. 32
- [Chayes, 1960] Chayes, F. (1960). On correlation between variables of constant sum. *Journal of Geophysical research*, 65(12):4185–4193. 6
- [Correa Morales and Barrera Causil, 2018] Correa Morales, J. C. and Barrera Causil, C. J. (2018). *Introducción a la estadística bayesiana: notas de clase*. Instituto Tecnológico Metropolitano. 17
- [Cómputos en Morelos, 2018] Cómputos en Morelos (2018). <http://computos.impepac.mx>. 61
- [Diharce, 2008] Diharce, E. V. (2008). Técnicas de simulación para el análisis estadístico de datos de medición. *Centro de investigación en Matemáticas*. 19
- [Escobar and West, 1995] Escobar, M. D. and West, M. (1995). Bayesian density estimation and inference using mixtures. *Journal of the american statistical association*, 90(430):577–588. 38
- [Estok et al., 2002] Estok, M., Nevitte, N., and Cowan, G. (2002). El conteo rápido y la observación de elecciones. *Instituto Nacional Demócrata para Asuntos Internacionales (NDI)*. 57
- [Ferguson, 1973] Ferguson, T. S. (1973). A bayesian analysis of some nonparametric problems. *The annals of statistics*, pages 209–230. 32

- [Gamerman and Lopes, 2006] Gamerman, D. and Lopes, H. F. (2006). *Markov chain Monte Carlo: stochastic simulation for Bayesian inference*. CRC Press. 24
- [Geisser and Cornfield, 1963] Geisser, S. and Cornfield, J. (1963). Posterior distributions for multivariate normal parameters. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 368–376. 35
- [Gelfand et al., 1990] Gelfand, A. E., Hills, S. E., Racine-Poon, A., and Smith, A. F. (1990). Illustration of bayesian inference in normal data models using gibbs sampling. *Journal of the American Statistical Association*, 85(412):972–985. 22, 24
- [Geman and Geman, 1984] Geman, S. and Geman, D. (1984). Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on pattern analysis and machine intelligence*, (6):721–741. 22
- [General, 2018] General, C. (2018). Informe de actividades del comité técnico asesor de los conteos rápidos para los procesos electorales federal y locales 2017-2018, junio de 2018. 57
- [Hoff, 2009] Hoff, P. D. (2009). *A first course in Bayesian statistical methods*, volume 580. Springer. 17
- [Kalli et al., 2011] Kalli, M., Griffin, J. E., and Walker, S. G. (2011). Slice sampling mixture models. *Statistics and computing*, 21(1):93–105. 41, 42
- [Lindsay, 1995] Lindsay, B. G. (1995). Mixture models: theory, geometry and applications. In *NSF-CBMS regional conference series in probability and statistics*, pages i–163. JSTOR. 30
- [Lo, 1984] Lo, A. Y. (1984). On a class of bayesian nonparametric estimates: I. density estimates. *The annals of statistics*, pages 351–357. 32
- [Lynch, 2007] Lynch, S. M. (2007). *Introduction to applied Bayesian statistics and estimation for social scientists*. Springer Science & Business Media. 17
- [Martín-Fernández et al., 2004] Martín-Fernández, J., Daunis-i Estadella, J., and Mateu-Figueras, G. (2004). Clasificación paramétrica de datos composicionales: aproximación metodológica. In *XXVIII Congreso Nacional de Estad. e Invest. Oper., Cádiz (E), CD-ROM*. 15
- [McAlister, 1879] McAlister, D. (1879). Xiii. the law of the geometric mean. *Proceedings of the Royal Society of London*, 29(196-199):367–376. 16
- [Mendoza and Regueiro, 2011] Mendoza, M. and Regueiro, P. (2011). Estadística bayesiana. *Instituto Tecnológico de México*. 17
- [Montanero, 2018] Montanero, J. (2018). Manual abreviado de análisis estadístico multivariante. *España: Universidad de Extremadura*. 29
- [Muestra de Morelos, 2018] Muestra de Morelos (2018). <https://www.ine.mx/conteos-rapidos-procesos-electorales-federal-locales-2017-2018/>. 57
- [Muestra de Puebla, 2019] Muestra de Puebla (2019). <https://www.ine.mx/conteos-rapidos-procesos-electorales-locales-2018-2019/>. 11

- [Neal, 2003] Neal, R. M. (2003). Slice sampling. *Annals of statistics*, pages 705–741. 23
- [Pawłowsky-Glahn and Egozcue, 2001] Pawłowsky-Glahn, V. and Egozcue, J. J. (2001). Geometric approach to statistical analysis on the simplex. *Stochastic Environmental Research and Risk Assessment*, 15(5):384–398. 14
- [Pawłowsky-Glahn et al., 2015] Pawłowsky-Glahn, V., Egozcue, J. J., and Tolosana-Delgado, R. (2015). *Modeling and analysis of compositional data*. John Wiley & Sons. 10
- [Pearson, 1897] Pearson, K. (1897). Mathematical contributions to the theory of evolution.—on a form of spurious correlation which may arise when indices are used in the measurement of organs. *Proceedings of the royal society of london*, 60(359-367):489–498. V, 6
- [Reyes, 2012] Reyes, G. M. (2012). La función de probabilidad normal: Características y aplicaciones. *eXtoikos*, (6):107–110. 27
- [Rossi et al., 2012] Rossi, P. E., Allenby, G. M., and McCulloch, R. (2012). *Bayesian statistics and marketing*. John Wiley & Sons. 17
- [Sethuraman and Tiwari, 1982] Sethuraman, J. and Tiwari, R. C. (1982). Convergence of dirichlet measures and the interpretation of their parameter. In *Statistical decision theory and related topics III*, pages 305–315. Elsevier. 32
- [Van den Boogaart and Tolosana-Delgado, 2013] Van den Boogaart, K. G. and Tolosana-Delgado, R. (2013). *Analyzing compositional data with R*, volume 122. Springer. 10, 13
- [Walker, 2007] Walker, S. G. (2007). Sampling the dirichlet mixture model with slices. *Communications in Statistics—Simulation and Computation*<sup>®</sup>, 36(1):45–54. 33, 38, 41





# Apéndice A

## Códigos en R

### A.1. Códigos utilizados en R

#### Modelo de Mezclas de Normales "p"variadas con "k"componentes

##### Función para evaluar su densidad

```
#####  
###Funcion para evaluar de la Densidad del###  
#####  
###Modelo de Mezclas de Normales "p"-variadas con "k" componentes###  
#####  
#Autor: Arnoldo Daniel Miranda Fournier  
#Fecha de creacion: 17/07/2019  
#####  
###Evaluacion de la Densidad###  
#Para evaluar la densidad de una muestra de tamaño "n" de  
#vectores(escalares) reales  $y=\{y_1, y_2, \dots, y_n\}$  del Modelo de Mezclas  
#de Normales "p"-variado con "k" componentes, se necesita tener los  
#siguientes parametros en una lista:  
#w: Vector de dimension "k", que contiene los pesos de cada una de las "k" componenet  
#mu: Arreglo(Vector) de dimension "p x k"("k"), que contiene las medias para cada una  
#lambda: Vector de arreglos(Vector) de dimension "p x p x k"("k"), que contiene las p  
#Al mismo tiempo la muestra de vectores(escalares) reales  $y=\{y_1, y_2, \dots, y_n\}$   
#debe estar contenida en un arreglo(vector) de dimension "n x p"("n").  
#####  
###Argumentos###  
#y: Arreglo(Vector) de dimension "n x p"("n") que contiene la muestra que se evaluara  
#par: Lista que contiene los parametros "w", "mu" y "lambda".  
#univ: Argumento logico que indica si la dimension del modelo de mezclas es univariad  
#####  
require(LaplacesDemon)  
#Se utiliza el paquete "LaplacesDemon" que contiene las funciones "dmvnp" y "dnormp"  
require(compiler)  
#Se utiliza el paquete "compiler" que contiene la funcion "cmpfun" la cual permite co  
#####  
###Funcion para evaluar la densidad###
```

```
d.MMNp <- function(y, par, univ = TRUE){
  d <- numeric(1)
  if(univ){
    for(j in 1:length(par$w)){
      d <- d + par$w[j]*dnormp(y, par$mu[j], par$lambda[j])
    }
  }else{
    for(j in 1:length(par$w)){
      d <- d + par$w[j]*dmvnp(y, par$mu[ , j], par$lambda[ , , j])
    }
  }
  return(d)
}
d.MMNp <- cmpfun(d.MMNp)
```

### Función para realizar su simulación

```
#####
###Funcion para la simulacion del###
###Modelo de Mezclas de Normales "p"-variadas con "k" componentes###
#####
#Autor: Arnoldo Daniel Miranda Fournier
#Fecha de creacion: 17/07/2019
#####
###Simulacion###
#Para simular una muestra de tamaño "n" de
#vectores(escalares) reales  $y=\{y_1, y_2, \dots, y_n\}$  del Modelo de Mezclas
#de Normales "p"-variado con "k" componentes, se necesita tener los
#siguientes parametros en una lista:
#w: Vector de dimension "k", que contiene los pesos de cada una de las "k" compo
#mu: Arreglo(Vector) de dimension "p x k"("k"), que contiene las medias para cad
#lambda: Vector de arreglos(Vector) de dimension "p x p x k"("k"), que contiene
#####
###Argumentos###
#n: Entero positivo que indica el tamaño de la muestra a simular.
#par: Lista que contiene los parametros "w", "mu" y "lambda".
#univ: Argumento logico que indica si la dimension del modelo de mezclas es univ
#####
require(LaplacesDemon)
#Se utiliza el paquete "LaplacesDemon" que contiene las funciones "rmvnp" y "rno
require(compiler)
#Se utiliza el paquete "compiler" que contiene la funcion "cmpfun" la cual perm
#####
###Funcion para simular###
r.MMNp <- function(n, par, univ = TRUE){
  nk <- hist(runif(n), breaks = c(0, cumsum(par$w)), plot = FALSE)$counts
  muestra <- NULL
  if(univ){
```

```

    for (k in 1:length(nk)){
      muestra <- c(muestra, rnormp(nk[k], par$mu[k], par$lambda[k]))
    }
  }else{
    for (k in 1:length(nk)){
      muestra <- rbind(muestra, rmvnp(nk[k], par$mu[ , k], par$lambda[ , , k]))
    }
  }
  return(muestra)
}
r.MMNp <- cmpfun(r.MMNp)
#####

```

## Modelo de Mezclas de Normales Bivariadas con "k" componentes

Función para su graficación en los reales

```

#####
###Funcion para graficar en R~2 una muestra o poblacion del###
###Modelo de Mezclas de Normales "2"-variadas con "k" componentes###
###y en caso de contar con los parametros su densidad real###
#####
#Autor: Arnoldo Daniel Miranda Fournier
#Fecha de creacion: 15/08/2019
#####
require(ggplot2)
#Se utiliza el paquete "ggplot2" para realizar las graficas.
require(compiler)
#Se utiliza el paquete "compiler" que contiene la funcion "cmpfun" la cual permite co
#####
###Argumentos###
#y: Arreglo de dimension "n x 2" que contiene la muestra o poblacion del modelo de me
#par.real: Lista que contiene los parametros "w", "mu" y "lambda" del modelo de mezcl
#den.real: Argumento logico que indica si se desea utilizar la densidad del modelo de
#den.est: Argumento logico que indica si se desea utilizar una densidad estimada por
#curvas: Argumento logico que indica si se desea añadir al grafico las curvas de nive
#Pob: Argumento logico que indica si se trata de la poblacion, por defecto es FALSE.
#####
###Funcion###
g.MMN2 <- function(y, par.real = NULL, den.real = FALSE, curvas = FALSE, Pob = FALSE,
  if(is.null(par.real) & den.real){
    g <- ggplot(data = data.frame(y), aes(x = X1, y = X2)) +
      xlab(expression(y[1])) +
      ylab(expression(y[2])) +
      ggtitle(ifelse(Pob, expression(italic("Diagrama de Dispersión de la población")),
        #Se agregan los puntos:
        geom_point(mapping = aes(fill = factor(1)), color = "black", shape = 21, size = 2
        scale_fill_manual(values = "steelblue")
      if(curvas){

```

```

    g <- g +
      geom_density_2d(mapping = aes(color = ..level..), h = h, bins = bins) +
      scale_colour_gradientn(colors = topo.colors(100, .75), name = expression(1))
  }
  return(g)
}else{
d.MMN2 <- function(y){
  d <- numeric(1)
  for(j in 1:length(par.real$w)){
    d <- d + par.real$w[j]*dmvnp(y, par.real$mu[ , j], par.real$lambda[ , , j])
  }
  return(d)
}
d.MMN2 <- cmpfun(d.MMN2)
z <- d.MMN2(y)
g <- ggplot(data = data.frame(y, z), aes(x = X1, y = X2)) +
  xlab(expression(y[1])) +
  ylab(expression(y[2])) +
  ggtitle(ifelse(Pob, expression(bold("Diagrama de Dispersión de la población")),
  #Se agregan los puntos:
  geom_point(mapping = aes(fill = z), color = "black", shape = 21, size = 2, show.legend = FALSE) +
  scale_fill_gradientn(colors = topo.colors(100, .75), name = expression(bold("Densidad de la población")))
  if(curvas){
    g <- g +
      geom_density_2d(mapping = aes(color = ..level..), h = h, bins = bins, show.legend = FALSE) +
      scale_colour_gradientn(colors = topo.colors(100, .75), name = expression(1))
  }
  return(g)
}
}
g.MMN2 <- cmpfun(g.MMN2)
#####

```

### Función para su graficación en el Símpex

```

#####
###Funcion para graficar en  $S^3$  una muestra o poblacion del###
###Modelo de Mezclas de Normales "2"-variadas con "k" componentes###
#####
#Autor: Arnoldo Daniel Miranda Fournier
#Fecha de creacion: 15/08/2019
#####
require(ggplot2)
#Se utiliza el paquete "ggplot2" para realizar las graficas.
require(ggtern)
#Se utiliza el paquete "ggtern" para realizar el diagrama ternario.
require(compiler)
#Se utiliza el paquete "compiler" que contiene la funcion "cmpfun" la cual permite
#####

```

```

###Argumentos###
#y: Arreglo de dimension "n x 2" que contiene la muestra o poblacion del modelo de me
#den.est: Argumento logico que indica si se desea utilizar una densidad estimada por
#Pob: Argumento logico que indica si se trata de la poblacion, por defecto es FALSE.
#####
###Diagrama Ternario###
#Con la libreria "ggtern"
g.S3 <- function(x, den.est = FALSE, Pob = FALSE){
  g <- ggtern(data = as.data.frame(x), mapping = aes(X1, X2, X3)) +
    geom_mask() +
    #Dibujamos los puntos:
    geom_point(color = "black", fill = "steelblue", shape = 21, size = 2) +
    limit_tern(breaks = c(0, .1, .2, .3, .4, .5, .6, .7, .8, .9, 1), labels = sprintf("%",
    theme_showarrows() +
    ggtitle(ifelse(Pob, expression(bold("Diagrama Ternario de la población")), expression(
    xlab(expression(x[1])) +
    ylab(expression(x[2])) +
    zlab(expression(x[3]))
  if(den.est){
    #Dibujamos la densidad estimada:
    g <- g +
      geom_density_tern(aes(color=..level..)) +
      scale_color_gradientn(colors = topo.colors(100, .65), name = expression(bold("Der
  }
  return(g)
}
g.S3 <- cmpfun(g.S3)
#####

```

## Inferencia del MMNp

### Funciones a utilizar en su Inferencia

```

#####
###Funciones a utilizar en el Script###
###Inferencia_MMNp(Funcion)###
#####
#Autor: Arnoldo Daniel Miranda Fournier
#Fecha de creacion: 17/07/2019
#####
require(LaplacesDemon)
#Se utiliza el paquete "LaplacesDemon" que contiene las funciones "rmvnp" y "rnormp"
require(compiler)
#Se utiliza el paquete "compiler" que contiene la funcion "cmpfun" la cual permite co
#####
###Funcion que calcula los parametros para las distribuciones posteriores marginales
###Argumentos###
#y: Arreglo(Vector) de dimension "n x p"("n") que contiene la muestra con la que se l
#mu0: Primer hiperparametro de la distribucion inicial para "mu", tiene que ser un ve

```

```

#n0: Segundo hiperparametro de la distribucion inicial para "mu", tiene que ser
#alfa0: Primer hiperparametro de la distribucion inicial para "lambda", tiene que
#beta0: Segundo hiperparametro de la distribucion inicial para "lambda", tiene que
#univ: Argumento logico que indica si la dimension del modelo de mezclas es univ
#####
i.MNp <- function(y, mu0, n0, alfa0, beta0, univ = TRUE){
  if(univ){
    n <- length(y)
    ybar <- mean(y)
    ifelse(n == 1, S <- 0, S <- var(y)*(n-1))
    mu <- (n0*mu0 + n*ybar)/(n0 + n)
    betan <- beta0 + (0.5)*(S + ((n*n0*(mu0 - ybar)^2)/(n + n0)))
    beta <- 1/(2*betan)
    alfa1 <- alfa2 <- 2*alfa0 + n
    Sigma <- betan/((n0 + n)*(alfa1/2))
  }else{
    ifelse(is.array(y), c(n <- nrow(y), p <- ncol(y)), c(n <- NCOL(y), p <- NROW(y)),
    ifelse(n == 1, c(ybar <- y, S <- 0*diag(p)), c(ybar <- apply(y, 2, mean), S <- 0*diag(p))),
    mu <- (n0*mu0 + n*ybar)/(n0 + n)
    betan <- beta0 + (0.5)*(S + ((n*n0*((mu0 - ybar)%*%t(mu0 - ybar)))/(n + n0)))
    beta <- as.symmetric.matrix(solve(2*betan))
    alfa1 <- 2*alfa0 + n - p + 1
    Sigma <- betan/((n0 + n)*(alfa1/2))
    alfa2 <- 2*alfa0 + n
  }
  return(list(mu = mu, Sigma = Sigma, alfa1 = alfa1, alfa2 = alfa2, beta = beta))
}
i.MNp <- cmpfun(i.MNp)
#####
###Funcion que simula un vector(escalar) real con densidad St_p(St) con p>=2(p=1)
###Argumentos###
#par: Lista que contiene los parametros "mu", "lambda" y "alfa" del Modelo St p-
#univ: Argumento logico que indica si la dimension del modelo de mezclas es univ
#####
r.Stp <- function(par, univ = TRUE){
  ifelse(univ, return(rst(1, par$mu, par$Sigma, par$alfa1)), return(rmvt(1, par$mu, par$Sigma)))
}
r.Stp <- cmpfun(r.Stp)
#####
###Funcion que simula una matriz(escalar) definida positiva(positivo) con densidad Wishart
###Argumentos###
#par: Lista que contiene los parametros "alfa" y "beta" del Modelo Wishart p-var
#univ: Argumento logico que indica si la dimension del modelo de mezclas es univ
#####
r.Wip <- function(par, univ = TRUE){
  ifelse(univ, return(rgamma(1, shape = par$alfa2, scale = par$beta)), return(Laplace(1, par$mu, par$Sigma)))
}
r.Wip <- cmpfun(r.Wip)

```

```
#####
###Funcion que simula un vector(escalar) real con densidad  $N_p(N)$  con  $p \geq 2$  ( $p=1$ )###
###Argumentos###
#par: Lista que contiene los parametros "mu" y "lambda" del Modelo Normal  $p$ -variado,
#univ: Argumento logico que indica si la dimension del modelo de mezclas es univariad
#####
r.Np <- function(par, univ = TRUE){
  ifelse(univ, return(rnormp(1, par$mu, par$lambda)), return(rmvnp(1, par$mu, par$lambda)
}
r.Np <- cmpfun(r.Np)
#####
###Funcion que evalua la densidad del Modelo de Mezclas de Normales  $p$ -variado(univariad)
###Argumentos###
#y: Arreglo(Vector) de dimension " $n \times p$ " (" $n$ ") que contiene la muestra que se evaluara
#par: Lista que contiene los parametros "w", "mu" y "lambda" del Modelo de Mezclas de
#univ: Argumento logico que indica si la dimension del modelo de mezclas es univariad
#####
d.MMNp <- function(y, par, univ = TRUE){
  d <- numeric(1)
  if(univ){
    for(j in 1:length(par$w)){
      d <- d + par$w[j]*dnormp(y, par$mu[j], par$lambda[j])
    }
  }else{
    for(j in 1:length(par$w)){
      d <- d + par$w[j]*dmvnp(y, par$mu[ , j], par$lambda[ , , j])
    }
  }
  return(d)
}
d.MMNp <- cmpfun(d.MMNp)
#####
```

### Función que realiza su Inferencia

```
#####
###Funcion para realizar la inferencia al###
###Modelo de Mezclas de Normales " $p$ "-variadas con " $k$ " componentes###
#####
#Autor: Arnoldo Daniel Miranda Fournier
#Fecha de creacion: 17/07/2019
#####
###Argumentos###
#y: Muestra de tamaño " $n$ " proveniente del MMNp de dimension " $p$ ".
#mu0: Primer hiperparametro de la distribucion inicial para "mu", tiene que ser un ve
#n0: Segundo hiperparametro de la distribucion inicial para "mu", tiene que ser un es
#alfa0: Primer hiperparametro de la distribucion inicial para "lambda", tiene que ser
#beta0: Segundo hiperparametro de la distribucion inicial para "lambda", tiene que se
#c: Parametro de escala del proceso de Dirichlet, tiene que ser un escalar positivo.
```

```

#k.ini: Numero de componentes iniciales, tiene que ser un entero positivo.
#y.malla: Malla de tamaño "500" para graficar la distribucion predictiva para el
#sizeMCMC: Tamaño de la cadena final(muestra) con la que nos quedaremos, obtenid
#Burn: Periodo de calentamiento de la cadena.
#Conv: Periodo de convergencia de la cadena.
#Lag: Numero de saltos para disminuir la correlacion entre los elementos de la c
#####
###Funcion###
i.MMNp <- function(y, mu0, n0, alfa0, beta0, y.malla = NULL, c, k.ini, sizeMCMC, I
  ###Tamaño y dimension de la muestra ingresada###
  ifelse(is.vector(y), c(n <- length(y), p <- 1), c(n <- nrow(y), p <- ncol(y)))

  ###Numero de iteraciones a realizar en el Gibbs###
  ite.Gibbs <- Burn + Conv + Lag*sizeMCMC

  ###Funciones y parametros iniciales a utilizar###
  #####
  ###Especificamente para el Modelo de Mezclas de Normales p-variado###
  source('Inferencia_MMNp(Funciones).R')
  ifelse(p == 1, par.ini <- list(alfa2 = 2*alfa0, beta = 1/(2*beta0)), par.ini <
  #####

  ###Objetos a guardar###
  k <- k0 <- integer(sizeMCMC)
  if(!is.null(y.malla)){
    f.pred <- array(0, c(ifelse(p == 1, length(y.malla), nrow(y.malla)), sizeMCMC
  }
  ifelse(p == 1, y.pred <- numeric(sizeMCMC), y.pred <- array(0, dim = c(sizeMCMC

  ###Numero de iteraciones maximo para comenzar a guardar los objetos###
  ite.max <- (Burn + Conv)

  ###Contador auxiliar para guardar los objetos###
  cont <- 1

  ###Variables latentes###
  z <- floor(runif(n)*k.ini) + 1
  u <- exp(-z)*runif(n)
  ki <- as.integer(floor(-log(u)))

  ###Variables Temporales###
  k.temp <- max(ki)
  k0.temp <- sum(tabulate(ki, k.temp) > 0)
  nj <- tabulate(z, k.temp)
  v.temp <- w.temp <- numeric(n)
  #####
  ###Especificamente para el Modelo de Mezclas de Normales p-variado###
  ifelse(p == 1, par.temp <- list(mu = numeric(k.temp), lambda = numeric(k.temp))

```



```
#####

###Iteraciones###
for (ite in 1:ite.Gibbs) {
  temp <- 0
  for (j in 1:k.temp){
    if(nj[j] > 0){
      ###Muestreo de la distribucion posterior###
      #####
      ###Especificamente para el Modelo de Mezclas de Normales p-variado###
      if(p == 1){
        par.pos <- i.MNp(y[which(z == j)], mu0, n0, alfa0, beta0)
        par.temp$mu[j] <- r.Stp(par.pos)
        par.temp$lambda[j] <- r.Wip(par.pos)
      }else{
        par.pos <- i.MNp(y[which(z == j), ], mu0, n0, alfa0, beta0, univ = FALSE)
        par.temp$mu[ , j] <- r.Stp(par.pos, univ = FALSE)
        par.temp$lambda[ , , j] <- r.Wip(par.pos, univ = FALSE)
      }
      #####

      temp <- temp + nj[j]
      v.temp[j] <- rbeta(1, 1 + nj[j], c + (n - temp))
    }else{
      ###Muestreo de la distribucion inicial###
      #####
      ###Especificamente para el Modelo de Mezclas de Normales p-variado###
      if(p == 1){
        par.temp$lambda[j] <- r.Wip(par.ini)
        par.temp$mu[j] <- r.Np(list(mu = mu0, lambda = n0*par.temp$lambda[j]))
      }else{
        par.temp$lambda[ , , j] <- r.Wip(par.ini, univ = FALSE)
        par.temp$mu[ , j] <- r.Np(list(mu = mu0, lambda = n0*par.temp$lambda[ , , j]))
      }
      #####

      v.temp[j] <- rbeta(1, 1, c)
    }
    ###Stick-Breaking para los pesos w###
    ifelse(j == 1, w.temp[j] <- v.temp[1], w.temp[j] <- w.temp[j-1]*(v.temp[j]/v.temp[1]))
  }

  ###Bloque de almacenamiento###
  if((ite > ite.max) & ((ite %% Lag) == 0)){
    ###Densidad predictiva evaluada en la malla###
    if(!is.null(y.malla)){
      #####
      ###Especificamente para el Modelo de Mezclas de Normales p-variado###
    }
  }
}

```

```

par.temp$w <- w.temp[1:k.temp]/sum(w.temp[1:k.temp])
ifelse(p == 1, f.pred[ , cont] <- d.MMnp(y.malla, par.temp), f.pred[ , con
#####
}

###Muestrear de la predictiva###
u.pred <- runif(1)
ind.w.ord <- order(w.temp[1:k.temp])
ind.pred <- ind.w.ord[which(u.pred < w.temp[ind.w.ord])[1]]
if(is.na(ind.pred)){
  ###Muestreo de la predictiva mediante la distribucion inicial###
  #####
  ###Especificamente para el Modelo de Mezclas de Normales p-variado###
  if(p == 1){
    lambda <- r.Wip(par.ini)
    mu <- r.Np(list(mu = mu0, lambda = n0*lambda))
    y.pred[cont] <- r.Np(list(mu = mu, lambda = lambda))
  }else{
    lambda <- r.Wip(par.ini, univ = FALSE)
    mu <- r.Np(list(mu = mu0, lambda = n0*lambda), univ = FALSE)
    y.pred[cont, ] <- r.Np(list(mu = mu, lambda = lambda), univ = FALSE)
  }
  #####
} else{
  ###Muestreo de la predictiva mediante la distribucion final###
  #####
  ###Especificamente para el Modelo de Mezclas de Normales p-variado###
  if(p == 1){
    y.pred[cont] <- r.Np(list(mu = par.temp$mu[ind.pred], lambda = par.temp$
  } else{
    y.pred[cont, ] <- r.Np(list(mu = par.temp$mu[ , ind.pred], lambda = par.
  }
  #####
}

k[cont] <- k.temp
k0[cont] <- k0.temp
cont <- cont + 1
}

###Actualizacion de las variables latentes###
for(i in 1:n){
  #####
  ###Especificamente para el Modelo de Mezclas de Normales p-variado###
  if(p == 1){
    prob <- exp(1:ki[i])*w.temp[1:ki[i]]*dnormp(y[i], par.temp$mu[1:ki[i]], pa
  } else{
    prob <- numeric(ki[i])
  }
}

```

```

    for (j in 1:ki[i]) {
      prob[j] <- exp(j)*w.temp[j]*dmvnp(y[i, ], par.temp$mu[ , j], par.temp$lambda[ , j])
    }
  }
  #####

z[i] <- sample.int(ki[i], 1, prob = prob/sum(prob))
u[i] <- exp(-z[i])*runif(1)
ki[i] <- as.integer(floor(-log(u[i])))
}

###Actualizacion de Variables Temporales###
k.temp <- max(ki)
k0.temp <- sum(tabulate(ki, k.temp) > 0)
nj <- tabulate(z, k.temp)
#####
###Específicamente para el Modelo de Mezclas de Normales p-variado###
ifelse(p == 1, par.temp <- list(mu = numeric(k.temp), lambda = numeric(k.temp)), par.temp)
#####
}

###Resultados###
if(!is.null(y.malla)){
  return(list(k = k, k0 = k0, y.pred = y.pred, f.pred = f.pred))
}
return(list(k = k, k0 = k0, y.pred = y.pred))
}
i.MMNp <- cmpfun(i.MMNp)
#####

```

## Función para graficar el Análisis de Convergencia y Correlación

```

#####
###Funcion para graficar el Análisis de convergencia y correlacion de la inferencia de
###Modelo de Mezclas de Normales "p"-variadas con "k" componentes###
#####
#Autor: Arnoldo Daniel Miranda Fournier
#Fecha de creacion: 15/08/2019
#####
require(ggplot2)
#Se utiliza el paquete "ggplot2" para realizar las graficas.
require(gridExtra)
#Se utiliza el paquete "gridExtra" para adjuntar las graficas.
require(compiler)
#Se utiliza el paquete "compiler" que contiene la funcion "cmpfun" la cual permite co
#####
###Argumentos###
#par.est: Lista que contiene la muestra final de "k" y "k0" provenientes de la inferencia
#k1: Argumento logico que indica si se grafica el Análisis de convergencia para "k",

```

```

#k2: Argumento logico que indica si se grafica el Análisis de correlacion para "
#k01: Argumento logico que indica si se grafica el Análisis de convergencia para
#k02: Argumento logico que indica si se grafica el Análisis de correlacion para
#####
###Análisis de la Convergencia###
g.i.MMNp <- function(par.est, k1 = TRUE, k2 = TRUE, k01 = TRUE, k02 = TRUE){
  if((k2) | (k02)){
    qacf <- function(x, k = TRUE, conf.level = 0.95) {
      ciline <- qnorm((1 - conf.level)/2)/sqrt(length(x))
      bacf <- acf(x, plot = FALSE)
      bacfdf <- with(bacf, data.frame(lag, acf))
      q <- ggplot(aes(x = lag, y = acf), data = bacfdf) +
        geom_hline(yintercept = -ciline, color = "blue", size = .5, lty = 3) +
        geom_hline(yintercept = ciline, color = "blue", size = .5, lty = 3) +
        geom_hline(yintercept = 0, color = "black", size = 0.3) +
        geom_segment(mapping = aes(xend = lag, yend = 0)) +
        ggtitle(expression(bold("Análisis de la Correlacion")) +
          xlab(expression(bold("Lag"))))
      ifelse(k, q <- q + ylab(expression(bold("ACF k"))), q <- q + ylab(expression(bold("ACF k"))))
      return(q)
    }
  }
}

k <- data.frame(ind = (1:length(par.est$k)), k = cumsum(par.est$k)/(1:length(par.est$k)))
k0 <- data.frame(ind = (1:length(par.est$k0)), k0 = cumsum(par.est$k0)/(1:length(par.est$k0)))
if((k1) & (k2) & (k01) & (k02)){
  g1 <- ggplot(data = k) +
    geom_line(mapping = aes(x = ind, y = k), colour = "blue") +
    ggtitle(expression(bold("Análisis de la Convergencia"))) +
    xlab(expression(bold("Ind"))) +
    ylab(expression(bold("k")))

  g2 <- qacf(par.est$k)

  g01 <- ggplot(data = k0) +
    geom_line(mapping = aes(x = ind, y = k0), colour = "blue") +
    ggtitle(expression(bold("Análisis de la Convergencia"))) +
    xlab(expression(bold("Ind"))) +
    ylab(expression(bold("k0")))

  g02 <- qacf(par.est$k0, k = FALSE)

  grid.arrange(g1, g2, g01, g02, ncol = 2, nrow = 2)
}else{
  if((k1) & (k2)){
    g1 <- ggplot(data = k) +
      geom_line(mapping = aes(x = ind, y = k), colour = "blue") +
      ggtitle(expression(bold("Análisis de la Convergencia"))) +

```

```

      xlab(expression(bold("Ind"))) +
      ylab(expression(bold("k")))

    g2 <- qacf(par.est$k)

    grid.arrange(g1, g2, ncol = 2, nrow = 1)
  }else if((k01) & (k02)){
    g01 <- ggplot(data = k0) +
      geom_line(mapping = aes(x = ind, y = k0), colour = "blue") +
      ggtitle(expression(bold("Análisis de la Convergencia"))) +
      xlab(expression(bold("Ind"))) +
      ylab(expression(bold("k0")))

    g02 <- qacf(par.est$k0, k = FALSE)

    grid.arrange(g01, g02, ncol = 2, nrow = 1)
  }else if((k1) & (k01)){
    g1 <- ggplot(data = k) +
      geom_line(mapping = aes(x = ind, y = k), colour = "blue") +
      ggtitle(expression(bold("Análisis de la Convergencia"))) +
      xlab(expression(bold("Ind"))) +
      ylab(expression(bold("k")))

    g01 <- ggplot(data = k0) +
      geom_line(mapping = aes(x = ind, y = k0), colour = "blue") +
      ggtitle(expression(bold("Análisis de la Convergencia"))) +
      xlab(expression(bold("Ind"))) +
      ylab(expression(bold("k0")))

    grid.arrange(g1, g01, ncol = 2, nrow = 1)
  }else if((k2) & (k02)){
    g2 <- qacf(par.est$k)

    g02 <- qacf(par.est$k0, k = FALSE)

    grid.arrange(g2, g02, ncol = 2, nrow = 1)
  }else if(k1){
    g1 <- ggplot(data = k) +
      geom_line(mapping = aes(x = ind, y = k), colour = "blue") +
      ggtitle(expression(bold("Análisis de la Convergencia"))) +
      xlab(expression(bold("Ind"))) +
      ylab(expression(bold("k")))

    return(g1)
  }else if(k2){
    g2 <- qacf(par.est$k)

    return(g2)
  }

```

```

}else if(k01){
  g01 <- ggplot(data = k0) +
    geom_line(mapping = aes(x = ind, y = k0), colour = "blue") +
    ggtitle(expression(bold("Análisis de la Convergencia"))) +
    xlab(expression(bold("Ind"))) +
    ylab(expression(bold("k0")))

  return(g01)
}else if(k02){
  g02 <- qacf(par.est$k0, k = FALSE)

  return(g02)
}
}
}
g.i.MMNp <- cmpfun(g.i.MMNp)
#####

```

### Regiones de confianza caso $p = 2$

```
###
```

```
y.pred <- par.est$y.pred
```

```
#Diagrama de Dispersion de la muestra de la "predictiva" en  $R^2$ :
```

```
g.Mue.Pred.MMN2(y.pred = y.pred, den.est = TRUE, h = NULL, bins = NULL)
```

```
#Obtenemos la densidad estimada:
```

```
den.ker <- MASS::kde2d(y.pred[ , 1], y.pred[ , 2], n = nrow(y.pred)*1)
```

```
#Obtenemos la curva de nivel:
```

```
line.cont <- contourLines(den.ker$x, den.ker$y, den.ker$z, nlevels = 20)
contour(den.ker$x, den.ker$y, den.ker$z, nlevels = 20)
```

```
#Obtenemos los puntos correspondientes a la curva de nivel:
```

```
#Reg.Conf.R2 <- data.frame(x = line.cont[[1]][["x"]], y = line.cont[[1]][["y"]])
```

```
Reg.Conf.R2 <- NULL
```

```
for (i in 1:22) {
```

```
  df2 <- data.frame(x = line.cont[[i]][["x"]], y = line.cont[[i]][["y"]])
```

```
  Reg.Conf.R2 <- rbind(Reg.Conf.R2, df2)
```

```
}
```

```
#Transformación  $ilr^{-1}$  de la curva de nivel de la muestra de la distribución pre
```

```
Reg.Conf.S2 <- compositions::ilrInv(Reg.Conf.R2)
```

```
colnames(Reg.Conf.S2) <- c("X1", "X2", "X3")
```

```
#Diagrama Ternario de la muestra y de la curva de nivel en  $S^3$ :
```

```
g.Mue.Pred.S3(x = x, x.pred = Reg.Conf.S2, den.est = FALSE)
```

```
#Intervalo de Confianza Marginal para la componente 1:
```

```
#Poblacional:
```

```
(IC.X1.Pob <- quantile(X[ , 1], probs = c(0.025, 0.975), na.rm = TRUE))
```

```
#Predictivo:
```

```
(IC.X1.Pred <- quantile(Reg.Conf.S2[ , 1], probs = c(0.025, 0.975), na.rm = TRUE))
```

```
#Intervalo de Confianza Marginal para la componente 2:
```

```
#Poblacional:
```

```
(IC.X2.Pob <- quantile(X[ , 2], probs = c(0.025, 0.975), na.rm = TRUE))
```

```
#Predictivo:
```

```
(IC.X2.Pred <- quantile(Reg.Conf.S2[ , 2], probs = c(0.025, 0.975), na.rm = TRUE))
```

```
#Intervalo de Confianza Marginal para la componente 3:
```

```
#Poblacional:
```

```
(IC.X3.Pob <- quantile(X[ , 3], probs = c(0.025, 0.975), na.rm = TRUE))
```

```
#Predictivo:
```

```
(IC.X3.Pred <- quantile(Reg.Conf.S2[ , 3], probs = c(0.025, 0.975), na.rm = TRUE))
```