



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

Para obtener el título de:

**MAESTRÍA EN CIENCIAS
MATEMÁTICAS APLICADAS E INDUSTRIALES**

Presenta:

Mariana Paola Ramos Gordillo

Asesor:

Dr. Gabriel Núñez Antonio

CDMX, México, a 12 de junio del 2020.

Índice general

Resumen	9
Introducción	11
1. Preliminares	17
1.1. Datos Composicionales	17
1.2. Datos Direccionales	36
1.3. Estadística Bayesiana	43
1.4. Métodos de Simulación	46
2. Un modelo Bayesiano para Datos Composicionales	57
2.1. El Modelo Gamma	57
2.2. El Modelo de la Gamma Propuesto	58
2.3. El Modelo Propuesto	61
3. Inferencia Bayesiana	67
3.1. Inferencia para el modelo Gamma	67
3.2. Inferencia para el modelo Gamma Propuesto	69
3.3. Inferencia para el Modelo Propuesto	74
4. Aplicaciones	77
4.1. Datos Simulados	77
4.2. Datos Reales	82
Conclusiones y Perspectivas	93
A. Códigos en R	99
A.1. Funciones para la Inferencia de los parámetros de una Gamma univariada	99
A.2. Función para la Inferencia del Modelo Propuesto	100
A.3. Funciones Auxiliares para la Inferencia del Modelo Propuesto	103

*A mi madre,
por ser un ejemplo de fortaleza y vida.*

*A mi padre,
por criarme y darme los valores que me caracterizan.*

*A Lu,
Por ser mi cómplice incondicional.*

*A Dani,
por ser mi compañero de vida.*

Agradecimientos

A CONACYT y al gobierno mexicano, por el apoyo brindado para desarrollar este proyecto.

A la Universidad Autónoma Metropolitana, por las facilidades otorgadas para mi desarrollo académico, profesional y personal.

Al Dr. Gabriel Nuñez Antonio, por dirigir este trabajo de tesis, por enseñarme el compromiso real a un proyecto y por la confianza en mi potencial.

Al Dr. Fabian Martínez Martínez, al Dr. Francisco Julián Ariza Hernández y al Act. Arturo González Morales por el tiempo dedicado a la revisión de este trabajo de tesis y sus puntuales observaciones y comentarios.

A mi familia, por apoyarme en cada meta que decido proponerme.

A mis compañeros de la MCMAI, pero sobretodo a los de mi generación por su gran amabilidad y compañerismo, así como por su disposición a compartir su conocimiento.

A mis amigos, que a pesar del tiempo, no dejan de mostrarme su cariño y apoyo en cualquier situación.

Resumen

En este trabajo se revisa la naturaleza de las variables composicionales, es decir, variables multivariadas donde la suma de sus componentes debe cumplir con la restricción de ser una constante, además se puntualizan los retos para especificar modelos que describan adecuadamente sus comportamientos. Se propone un análisis Bayesiano basado en el estudio de variables composicionales a través de modelos de mezclas de distribuciones Gamma multivariadas, y se describe la manera de cómo usar este modelo en la inferencia de variables composicionales.

La inferencia sobre los parámetros del modelo se basa en muestras de la densidad posterior, las cuales se obtienen mediante un *Slice Sampler* después de la introducción de variables latentes adecuadas. La metodología propuesta se ilustra utilizando conjuntos de datos simulados. Adicionalmente, se discute la aplicación de este modelo al análisis de datos como los obtenidos en la realización de los Conteos Rápidos organizados por el Instituto Nacional Electoral, en 2018.

Introducción

En la modelación estadística de fenómenos reales existen diversas ramas del conocimiento en las cuales un investigador puede encontrarse con *variables composicionales*; es decir, datos que representan composiciones. Por ejemplo, en Petrología el estudio de la composición de los diferentes óxidos principales o los minerales básicos que conforman geoquímicamente una roca. En Agricultura, el análisis de la composición de la tierra en términos de sus diferentes minerales es relevante. En Economía, la composición de un portafolio de inversión o la composición del gasto de los hogares es de importancia. En los procesos electorales, el estudio de la proporción de votos a favor de cada fuerza política, en alguna elección federal o estatal, es de relevancia en el estudio de las preferencias electorales. Cada uno de los ejemplos anteriores está asociado a variables de naturaleza composicional cuya adecuada modelación requiere de cierto cuidado en términos estadísticos. Algunos problemas de interés están dados por las siguientes preguntas: ¿Existe una variabilidad que pueda ser descrita por los datos?, ¿existe alguna composición central alrededor de la cual pueda explicarse esa variabilidad?

En la etapa inicial del planteamiento y la modelación de cualquier fenómeno real, como un problema estadístico, es crucial el reconocimiento y la definición de un espacio muestral adecuado para el análisis de los datos con los que se está trabajando. En este caso, los datos composicionales son vectores con la restricción de que sus componentes están sujetos a sumar una constante c y cuyos elementos son no negativos. Así, los datos composicionales tienen al *simplex D -dimensional* como el espacio muestral comúnmente asociado. Históricamente, la estructura algebraica de \mathbb{R}^k resultó familiar y muy intuitiva para la modelación estadística de datos, propiciando el desarrollo de una gama extensa de métodos e interpretaciones naturales. Por otro lado, hasta que se reconoció la topología del simplex (ver, por ejemplo Aitchison (1982) [2])

se tuvo un gran impulso en el análisis estadístico de los *datos composicionales*.

La propuesta de Aitchison(1982) [2] para el tratamiento de datos composicionales es relativamente reciente y ha ofrecido un panorama general para el análisis de variables aleatorias definidas en el simplex. Sin embargo, aún quedan aspectos por resolver desde el punto de vista metodológico y hoy en día son tema de discusión (ver por ejemplo, Pawlosky-Glahn y Egozcue (2015) [24] y las referencias ahí incluídas).

Las restricciones de no negatividad y de suma constante introducen un reto importante en la modelación de variables aleatorias en el simplex. El enfoque de Aitchison basado en log-cocientes para el análisis de datos composicionales ha sido una fuente de discusión en las últimas décadas. Esta aproximación de Aitchison ha permitido realizar análisis estadísticos frecuentistas una vez que se ha aplicado una transformación a los datos originales. Con la idea de llevar las variables composicionales a espacios más manejables como \mathbb{R}^k y posteriormente regresar al Simplex a través de algunas *isometrías*. Esto da una ventaja en el análisis, debido a la gran cantidad de procedimientos estadísticos disponibles en \mathbb{R}^k .

Aún con las transformaciones propuestas por Aitchison, el estudio de este tipo de datos es complejo, y debido a ello, el desarrollo de una metodología para su análisis tuvo un crecimiento relativamente lento. Sin embargo, en los últimos años el desarrollo computacional y la creación de diversas herramientas, han permitido la generación y el desarrollo de nuevos métodos que permiten analizar este tipo de datos de forma más adecuada. Las aportaciones realizadas van desde el desarrollo de métodos de visualización y graficación, hasta propuestas de modelos que permiten realizar inferencias sobre observaciones composicionales.

Así, el análisis de datos en el simplex involucra variables que por su naturaleza solo son definidos en ciertos subconjuntos de \mathbb{R}^k . Es por ello que, desde un punto de vista metodológico, resulta importante tener distribuciones de probabilidad definidas en esos subconjuntos de \mathbb{R}^k . Específicamente, es necesario contar con modelos de probabilidad que describan variables restringidas al primer ortante de \mathbb{R}^k .

Con el fin de contribuir a extender las propuestas metodológicas para el análisis de datos composicionales, en este trabajo de tesis se propone un modelo de mezclas para describir este tipo de datos, el cual como se verá ofrece una gran flexibilidad en la modelación. Los modelos de mezclas se encuentran entre las herramientas más utilizadas en muchas aplicaciones en las que se requieren modelar datos con características como multimodalidad, heterogeneidad o conglomerados, asimetría, etc. Dichas mezclas se puede visualizar de forma práctica como la suma ponderada de componentes que son, en sí mismas, distribuciones de probabilidad (generalmente, todas las componentes pertenecen a una misma familia de distribución, aunque esto no es necesario) y pueden tener componentes finitos o infinitos. La densidad de una mezcla puede verse de la forma

$$f(y|\omega, \phi) = \sum_{j=1}^{\infty} \omega_j K(y|\phi_j)$$

donde la colección infinita de ω 's son no negativos y K es el kernel, en este caso el mismo para todos. El gran desafío aquí es realizar inferencias estadísticas en este tipo de modelos. Estos retos incluyen, por ejemplo, estimar los correspondientes vectores de pesos, estimar el número de componentes en la mezcla significativamente diferentes de cero, la estimación de la correspondiente distribución predictiva, etc.

En contraste, el estudio de *datos direccionales* (variables que representan direcciones) ha tenido un mayor desarrollo en los últimos años. Ver por ejemplo, Wang y Gerfand (2007) [34], Scely y Welsh (2011)[29], así como Núñez-Antonio, et al. (2015)[21] y las referencias allí incluidas. Los datos direccionales aparecen comúnmente en ciencias como la Biología, Geofísica, Meteorología y Ecología. Algunas de sus principales aplicaciones son las direcciones de viento, los datos de orientación y migración de aves o algunas otras especies, las direcciones de propagación de fisuras en el concreto u otros materiales, etc. Para una revisión y análisis, desde un punto de vista frecuentista, el lector puede referirse a Mardia (1976)[18] y Mardia y Jupp (2000)[19]. Por su parte, Núñez-Antonio, et. al (2015)[21], presentan un modelo Bayesiano no-paramétrico para datos direccionales definidos en el círculo unitario (*datos circulares*) el cual se tomará como modelo base en el desarrollo de esta

tesis.

La metodología propuesta en este trabajo se basa en el siguiente modelo de mezclas con un kernel definido en la esfera unitaria y analizado en Núñez-Antonio y Geneyro (2019) [22].

$$f(\boldsymbol{\theta}) = \int GP(\boldsymbol{\theta}|\alpha, \beta) P(d\theta), \quad (1)$$

donde $GP(\boldsymbol{\theta}|\cdot)$ es una densidad Gamma proyectada y P define un Proceso Dirichlet (Ferguson (1973) [10]).

Así, el modelo propuesto para analizar datos composicionales es un modelo para datos direccionales generado por (1) y posteriormente mapeado al simplex D -dimensional. Aunque la idea de modelar datos composicionales a través de modelos para datos direccionales no es nueva (ver por ejemplo, Aitchison (1982) [2]), hasta ahora no se contaba con una metodología más o menos general para hacer inferencia en modelos para variables direccionales.

Para llevar a cabo inferencias en el modelo propuesto, dada su complejidad, se propone realizar inferencias desde un enfoque Bayesiano de la estadística. Lo anterior, nos lleva al uso de procedimientos y algoritmos de simulación directa como el muestreo de rechazo adaptativo (ARS, por sus siglas en inglés) y a algoritmos pertenecientes a la familia de los métodos de Monte Carlo vía Cadenas de Markov (MCMC, por sus siglas en inglés) como el *Gibbs sampler* y el *Slice sampler*.

El modelo se probará con datos simulados para evaluar su eficacia y se aplicará a datos del conteo rápido de las elecciones estatales, en este caso, para la elección de Morelos en 2018.

La estructura de esta tesis es la siguiente. En el Capítulo 1 se ofrece una introducción al análisis de datos composicionales, haciendo énfasis en los principales problemas que se presentan en su modelación. Se describe la naturaleza de variables direccionales, así como algunas herramientas para su análisis. Adicionalmente se da una breve introducción a los conceptos básicos de la Estadística Bayesiana y finalmente se expone una introducción a algu-

nos métodos de simulación estocástica.

En el Capítulo 2 se presenta el modelo Gamma univariado, el modelo Gamma multivariado y después se define un modelo de mezclas de distribuciones Gamma multivariadas. Posteriormente, se revisa el modelo proyectado sobre la esfera unitaria, que será el kernel para el modelo de mezclas. Finalmente, se presenta la propuesta de un modelo Bayesiano que describe datos composicionales basado en un modelo de mezclas de distribuciones Gamma proyectadas.

En el Capítulo 3 se desarrolla la propuesta de llevar a cabo inferencias bayesianas sobre el modelo propuesto. Para lo anterior, se derivan todas las distribuciones condicionales completas asociadas al modelo y se muestra la manera de simular de la distribución predictiva final del modelo de mezclas con kernel definido por distribuciones Gamma proyectadas.

En el Capítulo 4 se realizan diversas aplicaciones de la metodología propuesta. En los primeros ejemplos se analiza nuestra metodología usando datos simulados, con el objetivo de determinar el desempeño del modelo propuesto. En los ejemplos finales se aplica la metodología propuesta a datos reales derivados de los ejercicios de conteos rápidos realizados por el Instituto Nacional Electoral en las elecciones estatales de 2018.

Finalmente, en el último capítulo se presentan las conclusiones derivadas de este trabajo, se señalan algunas líneas de investigación futuras y se discuten las tareas pendientes.

Capítulo 1

Preliminares

En este capítulo se presentan los conceptos con los que se trabaja a lo largo de esta tesis. En primer lugar se describen algunas definiciones básicas para el análisis de datos composicionales, los cuales son el principal objetivo de estudio de este trabajo. Además, a través de una transformación se propondrá ir del simplex (datos composicionales) a la esfera unitaria (datos direccionales) y entonces utilizar algunos modelos que ya han sido propuestos para llevar a cabo inferencias en la esfera unitaria de dimensión p . Por esta razón, es necesario proporcionar algunas de las definiciones más relevantes sobre este tipo de datos y en este capítulo se podrán encontrar. Adicionalmente, el desarrollo de este trabajo se realiza desde un enfoque Bayesiano, por lo que se presentan algunas de las ideas más importantes de esta perspectiva de la Estadística, así como las herramientas que se utilizarán para llevar a cabo las inferencias en los modelos propuestos. Por lo anterior, se revisan algunos métodos de simulación como los métodos de Monte Carlo vía Cadenas de Markov (MCMC, por sus siglas en inglés).

1.1. Datos Composicionales

Los datos composicionales describen las partes de un todo y suelen representar vectores de proporciones, porcentajes, concentraciones o frecuencias, también pueden describirse como realizaciones de vectores aleatorios de suma constante. Con frecuencia aparecen ejemplos en diversas disciplinas que estudian con gran interés herramientas adecuadas para analizar este tipo de datos. Según Aitchison (1982) [2], los datos composicionales tienen características y operaciones propias las cuales se presentan a continuación.

Formalmente los datos composicionales se definen de la siguiente manera.

Definición 1.1 *Un dato composicional es un vector $\mathbf{x} = (x_1, x_2, \dots, x_D)$ constituido por D partes o componentes no negativas cuyo espacio muestral es el Símplex D -dimensional, S^D , definido por:*

$$S^D := \left\{ \mathbf{x} = (x_1, x_2, \dots, x_D) \mid x_1 \geq 0, \dots, x_D \geq 0; \sum_{i=1}^D x_i = k \right\}.$$

Los datos composicionales cuentan con operaciones que permiten trabajar con ellos, como el *operador clausura* \mathcal{C} , el cual se define a continuación.

Definición 1.2 *El operador clausura \mathcal{C} , es una transformación que hace corresponder a cada vector $\mathbf{x} = (x_1, x_2, \dots, x_D) \in \mathbb{R}_+^D$ con su dato composicional asociado. Es decir,*

$$\mathcal{C}(\mathbf{x}) = \left(\frac{x_1}{\sum_{i=1}^D x_i}, \dots, \frac{x_D}{\sum_{i=1}^D x_i} \right)$$

donde $\mathcal{C}(\mathbf{x}) \in S^D$.

Pero entonces, ¿en dónde se pueden encontrar ejemplos de Datos Composicionales? Es decir, ejemplos donde sea importante la relación entre las partes de un todo. A continuación se presentan algunos de ellos.

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?

Estudio de la composición de la leche

En un intento por mejorar la leche, se evaluó la leche de treinta vacas mediante la composición de la dieta antes y después de un régimen dietético y hormonal estrictamente controlado durante un periodo de ocho semanas. Se decidió tener un grupo de control de treinta vacas adicionales mantenidas en las mismas condiciones. Las sesenta vacas fueron asignadas para control y tratamiento de grupos al azar. Se registró la proporción de proteína, grasa láctea, carbohidratos, calcio, sodio y potasio del contenido total de la dieta. La pregunta esencial para ese problema es ¿el nuevo régimen dietético ha producido algún cambio en la composición de la leche?

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. Por ejemplo, si se tiene la composición de 5 minerales de una muestra de 25 especímenes de rocas de tipo hongite, algunas preguntas de interés pueden ser: ¿De qué manera podría describirse la variabilidad en la composición de estas rocas?, ¿Existe alguna composición central alrededor de la cual pueda describirse esta variabilidad?.

1.1.1. Operaciones con Datos Composicionales

Aitchison (1982) [2] define operaciones adecuadas que al aplicarlas a los datos composicionales los mantienen en el Simplex. Algunas de estas operaciones son la *perturbación* y la *potenciación*, las cuales resultan equivalentes a la suma y producto en \mathbb{R} , respectivamente.

Definición 1.3 *La operación perturbación entre $\mathbf{x} = (x_1, x_2, \dots, x_D) \in \mathcal{S}^D$ y*

$\mathbf{y} = (y_1, y_2, \dots, y_D) \in \mathcal{S}^D$ esta definida por

$$\mathbf{x} \oplus \mathbf{y} = \mathcal{C}(x_1 y_1, \dots, x_D y_D)$$

donde \mathcal{C} es el operador clausura.

El conjunto de perturbaciones forma un grupo el cual tiene como elemento neutro la composición $\mathbb{E} = \left(\frac{1}{D}, \dots, \frac{1}{D}\right)$ donde D es la dimensión del Símplex correspondiente. Este elemento neutral juega el papel del vector cero en los datos composicionales. La composición *inversa* de \mathbf{x} , denotada por $\oplus \mathbf{x}$, está dada por $\oplus \mathbf{x} = \mathcal{C}\left(\frac{1}{x_1}, \dots, \frac{1}{x_D}\right)$, donde \mathcal{C} es la clausura mencionada en la Definición 1.2.

Definición 1.4 La operación potenciación de $\mathbf{x} = (x_1, \dots, x_D) \in \mathcal{S}^D$ por un número $\alpha \in \mathbb{R}^+$ se define de la siguiente manera:

$$\alpha \odot \mathbf{x} = \mathcal{C}(x_1^\alpha, \dots, x_D^\alpha)$$

donde \mathcal{C} es el operador clausura.

Otros elementos de gran ayuda para definir líneas, ángulos, ortogonalidad, etc., en el simplex son la norma, el producto interior y la distancia, los cuales se definen a continuación.

Definición 1.5 Sean $\mathbf{x} = (x_1, x_2, \dots, x_D)$ y $\mathbf{y} = (y_1, y_2, \dots, y_D)$ dos elementos de \mathcal{S}^D . El producto interior de Aitchison está definido como

$$\langle \mathbf{x}, \mathbf{y} \rangle_A = \frac{1}{2D} \sum_{i=1}^D \sum_{j=1}^D \ln \frac{x_i}{x_j} \cdot \ln \frac{y_i}{y_j}.$$

Definición 1.6 Sea $\mathbf{x} = (x_1, x_2, \dots, x_D) \in \mathcal{S}^D$. La norma de Aitchison de \mathbf{x} se define como

$$\|\mathbf{x}\|_A = \sqrt{\frac{1}{2D} \sum_{i=1}^D \sum_{j=1}^D \left(\ln \frac{x_i}{x_j} \right)^2}.$$

Definición 1.7 La distancia de Aitchison entre $\mathbf{x} = (x_1, x_2, \dots, x_D)$ y $\mathbf{y} = (y_1, y_2, \dots, y_D)$ se define como

$$d_A(\mathbf{x}, \mathbf{y}) = \sqrt{\frac{1}{2D} \sum_{i=1}^D \sum_{j=1}^D \left(\ln \frac{x_i}{x_j} - \ln \frac{y_i}{y_j} \right)^2}.$$

Se puede notar que, dadas las definiciones anteriores, el simplex, $(\mathcal{S}^D, \oplus, \odot)$ es un *Espacio Vectorial* (Billheimer et al. (2001) [6]) y al estar dotado de un producto interior, el conjunto $(\mathcal{S}^D, \oplus, \odot, \langle \cdot \rangle)$, conforma un *Espacio Euclídeo*.

1.1.2. Algunas particularidades de los Datos Composicionales

En ocasiones el estudio de algún fenómeno requiere tomar un subconjunto de las componentes de un dato composicional en el Símplex D -dimensional y en consecuencia tener que trabajar con un nuevo espacio S -dimensional en el Símplex con $S < D$. Lo anterior lleva al concepto de subcomposición.

Definición 1.8 (*Subcomposición*) Si S es un subconjunto de las partes $(1, 2, \dots, D)$ de una composición \mathbf{x} , entonces $\mathcal{C}(\mathbf{x}_s)$ es una subcomposición de S -partes, con \mathcal{C} el operador clausura.

Por ejemplo, sea $\mathbf{x} = (x_1, x_2, x_3, x_4, x_5)$ la composición de una roca hongite, se puede estar interesado solo en la subcomposición química de 3 elementos (x_1, x_3, x_5) .

Si bien la formación de una subcomposición es en esencia una transformación de S^D a un Símplex de dimensión inferior, obsérvese que se tiene la propiedad de conservar la magnitud relativa entre las partes. Con la definición anterior y la propuesta de Aitchison para datos composicionales, se obtiene el siguiente resultado.

Propiedad 1.1 El cociente de 2 componentes de una subcomposición es igual al cociente de los correspondientes 2 componentes en la composición completa. Es decir, si $S = \mathcal{C}(\mathbf{x}_S) = \mathcal{C}(x_1, x_2, \dots, x_p)$ entonces

$$\frac{s_i}{s_j} = \frac{x_i}{x_j}$$

con $1 \leq i, j \leq p$ y $p < D$.

Como puede observarse de la Propiedad 1.1, una subcomposición conserva la relación de proporción existente en la relación de los datos en la composición total. Sin embargo bajo la propuesta de Aitchison, se dificulta entender la geometría del símplex \mathcal{S}^D , razón por la cual propone tomar logaritmos de

los cocientes entre las partes, así los datos son mapeados a \mathbb{R}^k y bajo esta transformación es posible utilizar métodos de estadística para variables definidas en \mathbb{R}^k .

Transformaciones de los Datos Composicionales

Para el análisis de Datos Composicionales, históricamente se han propuesto diversas transformaciones. A continuación se presenta cada una de ellas.

Definición 1.9 *La transformación log-cociente centrada (clr) se define como*

$$\text{clr}(\mathbf{x}) = \left(\ln \frac{x_1}{g(\mathbf{x})}, \ln \frac{x_2}{g(\mathbf{x})}, \dots, \ln \frac{x_D}{g(\mathbf{x})} \right)$$

con $g(\mathbf{x}) = \sqrt[D]{x_1 \cdot x_2 \cdots x_D}$.

La inversa de la transformación clr es sencilla: si $\mathbf{x}^* = \text{clr}(\mathbf{x})$, entonces $\mathbf{x} = \mathcal{C}(\exp[(\mathbf{x}^*)])$

El conjunto de composiciones junto con las operaciones de perturbación \oplus , potencia \odot y el producto escalar de Aitchison ¹ $\langle \cdot, \cdot \rangle_A$ construyen una estructura de *espacio Euclidiano* ($D - 1$) dimensional en el simplex. Esto significa que podemos 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 preserva los ángulos y las distancias.

El hecho de que el simplex sea un espacio euclideano tiene algunas implicaciones en la forma en que se aplican, interpretan y representan la mayoría de las estadísticas obtenidas en \mathbb{R}^k . La idea básica de llevar a cabo esta transformación ilr, se resume en que es posible:

- Hacer el cálculo de las coordenadas de los datos con respecto a una base ortonormal.

¹Definición 1.5

- Analizar esas coordenadas en una forma sencilla con el método deseado en \mathbb{R}^k .
- Aplicar los resultados que describen los objetos geométricos a la base ortonormal utilizada y regresarlos al Simplex, S^D . Los resultados finales no dependerán de la base elegida.

La transformación *log-cociente isométrica* se construye por la representación que resulte de una base dimensional de $(D-1)$ partes del espacio imagen \mathbb{H} de la transformación *clr*. Esta se construye tomando una base ortonormal de vectores de \mathbb{R}^D incluyendo el vector $\mathbf{v}_D = [1, \dots, 1]$, es decir, algunos $(D-1)$ vectores $\{\mathbf{v}_1^*, \dots, \mathbf{v}_{D-1}^*\} \in \mathbb{H}$ linealmente independientes. Entonces el conjunto de vectores definidos como $\mathbf{v}_i = \text{clr}^{-1}(v_i^*)$ forman una base de S^D . En términos computacionales se pueden organizar los vectores $\{v_j^*\}$ por columnas en una matriz de $D \times (D-1)$ elementos, denotado por \mathbf{V} .

Un ejemplo de estas transformaciones es el siguiente. Sea $\mathbf{a} = (1, 2, 4)$, aplicando la operación clausura se obtiene el dato composicional

$$\mathbf{a} = (0.1428571, 0.2857143, 0.5714286),$$

aplicando la transformación *ilr* a \mathbf{a} se obtiene el siguiente vector en \mathbb{R}^2 $\text{ilr}(\mathbf{a}) = (0.4901291, 0.8489285)$.

Otra forma de obtener ese vector es aplicar primero la transformación *clr* a \mathbf{a} y posteriormente aplicar la función *ilr*, es decir, multiplicar por una base ortonormal, haciendo esto se obtiene como resultado el vector $\text{ilr}(\text{clr}(\mathbf{a})) = (0.4901291, 0.8489285)$.

La base que transforma el vector \mathbf{a} a \mathbb{R}^2 es

$$V_{D \times (D-1)} = \begin{pmatrix} -0.7071068 & -0.4082483 \\ 0.0701068 & -0.4082483 \\ 0.0000000 & 0.8164966 \end{pmatrix}$$

Así, multiplicando $\mathbf{a} \cdot V_{D \times (D-1)}$ se obtiene $(0.4901291, 0.8489285)$.

La transformación *alr* es el enfoque en el que se basó Aitchison inicialmente para realizar el análisis de datos composicionales y se define de la siguiente

forma

$$\text{alr}(\mathbf{x}) = (\ln(x_1/x_D), \dots, \ln(x_{D-1}/x_D)).$$

Existen diversas opiniones respecto a cuál de ellas es mejor. Sin embargo, algunas de las razones por las que existen diversas transformaciones log-cocientes 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 produce vectores en \mathbb{R}^{D-1} , lo que ocasiona que en este espacio sea prácticamente imposible una interpretación correcta de la información original.
- La transformación ilr es isométrica y sus valores transformados producen matrices de covarianza de rango completo por lo que datos bajo esta transformación se pueden analizar de forma adecuada.

A continuación se listan algunas propiedades de las tres transformaciones mencionadas en esta sección.

■ Transformación Log-cociente Aditiva

$$\text{alr}(\mathbf{x} \oplus \mathbf{y}) = \text{alr}(\mathbf{x}) + \text{alr}(\mathbf{y})$$

$$\text{alr}(\lambda \odot \mathbf{x}) = \lambda \odot \text{alr}(\mathbf{x})$$

$$\langle \mathbf{x}, \mathbf{y} \rangle_A \neq \text{alr}(\mathbf{x}) \cdot \text{alr}'(\mathbf{y})$$

■ Transformación Log-cociente Centrada

$$\text{clr}(\mathbf{x} \oplus \mathbf{y}) = \text{clr}(\mathbf{x}) + \text{clr}(\mathbf{y})$$

$$\text{clr}(\lambda \odot \mathbf{x}) = \lambda \odot \text{clr}(\mathbf{x})$$

$$\langle \mathbf{x}, \mathbf{y} \rangle_A = \langle \text{clr}(\mathbf{x}), \text{clr}(\mathbf{y}) \rangle$$

■ **Transformación Log-cociente Isométrica**

$$\text{ilr}(\mathbf{x} \oplus \mathbf{y}) = \text{ilr}(\mathbf{x}) + \text{ilr}(\mathbf{y})$$

$$\text{ilr}(\lambda \odot \mathbf{x}) = \lambda \odot \text{ilr}(\mathbf{x})$$

$$\langle \mathbf{x}, \mathbf{y} \rangle_A = \text{ilr}(\mathbf{x}) \cdot \text{ilr}'(\mathbf{y})$$

$$\langle \mathbf{x}, \mathbf{y} \rangle_A = \langle \text{ilr}(\mathbf{x}), \text{ilr}(\mathbf{y}) \rangle$$

donde (w') representa la transpuesta del vector w .

1.1.3. Problemas en el análisis estadístico de los Datos Composicionales

Como se señala en Mateu et al. (2003) [20], para los datos composicionales, “la restricción de la suma constante ha sido considerada la fuente de todos los problema pues impide la aplicación de los procedimientos estadísticos habituales que se utilizan para datos que no presentan esta restricción”. Es importante hacer notar, por ejemplo, que el cambio en una de las partes induce un cambio en como mínimo una de las otras componentes. Además de esta problemática, se presentan algunas otras dificultades que deben tomarse en cuenta a la hora de estudiar los datos composicionales. A continuación se describen algunas de ellas.

Invarianza de Escala

Considérese un vector como una composición siempre que sus componentes representen el peso relativo o la importancia de un conjunto de partes que forman un todo. En este caso, el tamaño o el peso total de ese conjunto es irrelevante. Entonces se puede eliminar esa influencia aparentemente forzando a los vectores de datos a compartir la misma suma total k con la *operación clausura*. Este supuesto tiene consecuencias de largo alcance.

Para ilustrar este argumento se describe el siguiente ejemplo. Sean dos vectores de muestra

$$w = (1.6, 2.4, 4.0) \quad \text{y} \quad W = (3.0, 4.5, 7.5)$$

en \mathbb{R}_+^3 los cuales representan los pesos de tres componentes (a, b, c) de dos especímenes de peso total 8g y 15g, respectivamente. Si se está interesado en resolver un problema composicional, se puede notar que ambos vectores representan la misma composición, teniendo como diferencia la relación de escala $W = (15/8)w$.

Por otro lado, si se tuviera solo el dato composicional resulta imposible decidir de cual vector, W o w , proviene esa composición, es decir, en ausencia de información no composicional no hay manera de distinguir entre w y W .

Incoherencia subcomposicional

Como se menciona en sección anterior, las *subcomposiciones* representan subespacios de dimensiones más bajas donde los datos pueden ser proyectados para su análisis. Esto tiene implicaciones como la llamada *Coherencia Subcomposicional*, la cual se describe a partir de las siguientes características:

- Si se mide la distancia entre dos composiciones de D-partes, ésta debe ser mayor cuando se mide con todas las D componentes que cuando se mide en la subcomposición.
- La dispersión total de un conjunto de datos composicionales de D-partes debe ser mayor que la dispersión de cualquier subcomposición.
- Si se ajusta un modelo a un conjunto de datos composicionales de D-partes el resultado no debe cambiar si se incluye una nueva componente (aleatoria) y se trabaja con la composición de (D+1) partes.

Entonces el principio de *coherencia subcomposicional* establece que cuando se examina un subconjunto de las partes de una composición, se requiere que los resultados del análisis no sean contradictorios con los obtenidos por la composición original.

Aitchison (2003) [2] muestra en su publicación, este problema mediante un sencillo ejemplo. Se consideran dos científicos A y B que analizan muestras de tierra. Para cada una de las muestras, el científico A calcula un dato composicional de 4 partes (animal, vegetal, mineral, agua). El científico B elimina el agua de las muestras y calcula datos composicionales de tres partes

(animal, vegetal, mineral). Se puede notar que los datos del científico B son subcomposiciones de los datos del científico A. Los datos obtenidos son:

$(x_1; x_2; x_3; x_4)$	$(s_1; s_2; s_3)$
$(0.1; 0.2; 0.1; 0.6)$	$(0.25; 0.50; 0.25)$
$(0.2; 0.1; 0.1; 0.6)$	$(0.50; 0.25; 0.25)$
$(0.3; 0.3; 0.2; 0.2)$	$(0.375; 0.375; 0.25)$

Tabla 1.1: Composición de muestras de tierra para dos científicos.

Al calcular la correlación de los datos composicionales entre las partes animal y vegetal se obtiene que para el científico A $corr(x_1, x_2) = 0.5$ mientras que para el científico B es de $corr(s_1, s_2) = -1$. Así se puede observar que existe un problema de coherencia subcomposicional. Sin embargo, este problema podría resolverse notando que la relación de dos componentes permanece inalterada cuando se pasa de la composición completa a la subcomposición

$$\frac{x_i}{x_j} = \frac{s_i}{s_j}$$

de modo que mientras se trabaje con las composiciones en términos de proporciones, se tendrá *coherencia subcomposicional*.

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 (1995) [3].
- **El sesgo negativo:** Aparece porque cada renglón o columna en la matriz de covarianza de una composición cerrada suman cero, dado que las varianzas son positivas, esto implica que algunas covarianzas son forzadas a tomar valores negativos.

Por lo anterior, hay que tener cuidado en el análisis estadístico de datos composicionales.

Invarianza de la permutación

Los resultados de cualquier análisis no deben depender de la secuencia en la que se proporcionan los componentes de un conjunto de datos. Para el

log-cociente este principio es muy importante, por ejemplo cuando se usa la distancia euclídeana a los datos composicionales con la transformación **alr** no se cumple la invarianza de la 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.

1.1.4. Estadística Descriptiva

Un conjunto de datos composicionales puede representarse gráficamente de diversas formas, dependiendo del número de componentes. Por ejemplo, para dos componentes pueden ser los ya conocidos *Diagramas de Dispersión*, o su forma particular para log-cocientes. Para tres componentes, los más utilizados son los *Diagramas Ternarios*. Para ilustrar los ejemplos de cada tipo de grafico se tomarán los datos de *Geoquímica de Sedimentos Glaciares* de Tolsana-Delgado y von Eynatten (2010) [31].

Diagrama de Dispersión. El *Diagrama de Dispersión* convencional de dos componentes sin ninguna transformación, también llamado *Diagrama de Harker* por su uso en Geoquímica, resalta cualquier relación aditiva entre las variables trazadas. Por ejemplo, en la Figura 1.1 se visualizan dos componentes químicos que evolucionan en varias etapas y se observa la representación del balance de masas por etapa.

Desafortunadamente los diagramas de dispersión no son invariantes ante escala, ni perturbación y no tienen coherencia subcomposicional. Además de que no hay una garantía de que la gráfica de una subcomposición exhiba patrones similares o incluso compatibles con la composición original, incluso si las variables no incluidas eran irrelevantes para el proceso de estudio (correlación espuria), lo que ocasiona que una recta de regresión no sea confiable.

Diagrama de Dispersión Log-Cociente. Los *diagramas de dispersión de log-cociente* son gráficas de logaritmos de dos componentes versus el logaritmo de otras dos componentes (aunque a veces se pueden tener denominadores iguales).

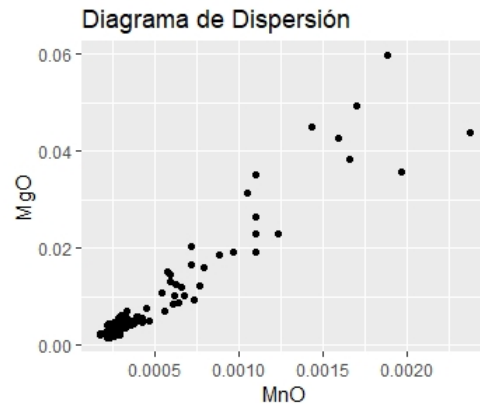


Figura 1.1: Diagrama de Dispersión de MnO (Óxido de Manganeso) *vs.* MgO (Óxido de Magnesio) de datos de Geoquímica de Sedimentos Glaciares.

Su representación es coherente con la descripción hecha anteriormente sobre los datos composicionales, pero no captura ninguna relación entre las partes representadas.

Algunas de las desventajas de esta forma de visualización es que no se garantiza que alguna componente de la composición este decayendo o creciendo con respecto a otra. Sin embargo este tipo de diagrama puede describir relaciones entre los cocientes.

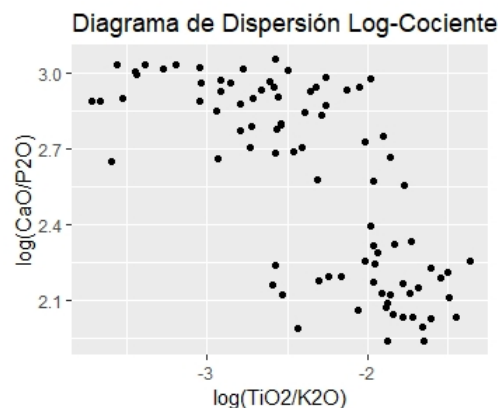


Figura 1.2: Diagrama de Dispersión Log-Cociente para componentes de datos Geoquímicos de Sedimentos Glaciares.

Diagrama Ternario. Para el caso $D = 3$, el simplex \mathcal{S}^3 suele representarse por un *diagrama ternario*, el cual tiene la propiedad de contar con segmentos ortogonales que unen un punto con los tres lados del triángulo equilátero. La

longitud de cada segmento es tomada como la proporción de una componente dada. Tienen la propiedad de representar los datos en sus características: *composicionales y relativos*.

Diagrama Ternario

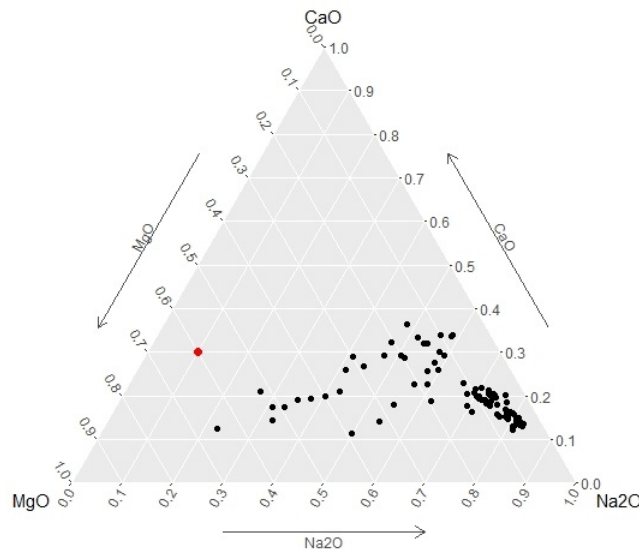


Figura 1.3: Diagrama de Ternario de tres componentes de datos Geoquímicos de Sedimentos Glaciares.

Existe una correspondencia biunívoca entre los datos composicionales de tres partes y los puntos del diagrama ternario. Un dato composicional $\mathbf{x} = (x_1, x_2, x_3)' = (\text{MgO}, \text{CaO}, \text{Na}_2\text{O})$ se corresponde con el punto que dista x_1 , x_2 y x_3 , respectivamente, de los lados opuestos a los vértices MgO, CaO y Na₂O, como puede observarse en la Figura 1.3. Por ejemplo, si $\mathbf{x} = (0.6, 0.3, 0.1)$ entonces el dato composicional se encuentra a una distancia 0.6 del lado opuesto a MgO, a 0.3 del lado opuesto a CaO y a 0.1 del lado opuesto a Na₂O, este punto se distingue en color rojo en la Figura 1.3.

Medidas Descriptivas

En su formulación estadística, los datos composicionales son realizaciones de una composición, es decir, son vectores aleatorios cuyo soporte está en S^D . En estos casos se plantea la necesidad de aplicar técnicas estadísticas adecuadas para el estudio e interpretación de este tipo de datos, que contemplen la

topología de S^D .

A continuación se presentan algunas de las medidas estadísticas numéricas que son útiles para la descripción de los datos composicionales.

En \mathbb{R} una medida estadística empleada con mayor frecuencia es una medida de ubicación. En el caso de datos composicionales, una medida de ubicación se define a través de la *media geométrica*.

Definición 1.10 Sea $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{iD}) \in \mathcal{S}^D$, la *media geométrica composicional* de \mathbf{x}_i , $i = 1, 2, \dots, n$ se define como

$$g(\mathbf{x}_1, \dots, \mathbf{x}_i, \dots, \mathbf{x}_n) = \mathcal{C}(g_1, g_2, \dots, g_D),$$

donde

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

$j = 1, 2, \dots, D$ y \mathcal{C} es el operador clausura.

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

$$\begin{aligned} \text{Vartot}[\mathbf{x}] &= \frac{1}{D} \sum_{i=1}^{D-1} \sum_{j=i+1}^D \text{Var}[\log(x_i/x_j)] \\ &= \sum_{j=1}^{D-1} \text{Var}[\text{ilr}(\mathbf{x})] \end{aligned}$$

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 (Pawlowsky-Glahn y Egozcue (2001), Pawlowsky-Glahn et al. (2015) [24]). El análisis de la variabilidad de una muestra se realiza utilizando la matriz de variación composicional de todos los log-cocientes simples, llamada en Aitchison (1982) *matriz de variación composicional*, como se define a continuación.

Definición 1.11 Para una composición \mathbf{X} de D -partes y n observaciones

$(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)$ la matriz de variación composicional esta dado por

$$\begin{array}{ccccccc}
 & 1 & 2 & 3 & \cdots & D-1 & D \\
 1 & \cdot & \tau_{12} & \tau_{13} & \cdots & \tau_{1(D-1)} & \tau_{1D} \\
 2 & \xi_{12} & \cdot & \tau_{23} & \cdots & \tau_{2(D-1)} & \tau_{2D} \\
 3 & \xi_{13} & \xi_{23} & \cdot & \cdots & \tau_{3(D-1)} & \tau_{3D} \\
 \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
 D-1 & \xi_{1(D-1)} & \xi_{2(D-1)} & \xi_{3(D-1)} & \cdots & \cdot & \tau_{(D-1)D} \\
 D & \xi_{1D} & \xi_{2D} & \xi_{3D} & \cdots & \xi_{(D-1)D} & \cdot
 \end{array}$$

donde $\xi_{ij} = E\{\log(x_i/x_j)\}$ y $\tau_{ij} = var\{\log(x_i/x_j)\}$.

Esta matriz de variación se puede estimar mediante $\hat{\xi}_{ij}$ y $\hat{\tau}_{ij}$ que son los estimadores de la media de los log-cocientes ξ_{ij} y la varianza de los log-cocientes τ_{ij} respectivamente, dados por: $n\hat{\xi}_{ij} = \sum_{r=1}^n \log(x_{ri}/x_{rj})$ y $(n-1)\hat{\tau}_{ij} = \sum_{r=1}^n \{\log(x_{ri}/x_{rj})\}^2 - n(\hat{\xi}_{ij})^2$. La matriz de variación composicional se puede interpretar, según Aitchison (1982) [2] como sigue. Si $\hat{\xi}_{ij}$ es positivo indica que el porcentaje de x_i en la composición tiende a tener mayor peso, que el de x_j . Sin embargo, si $\hat{\xi}_{ij} < \sqrt{\hat{\tau}_{ij}}$, entonces para un número sustancial de replicas $\log(\frac{x_i}{x_j})$ es negativo con un porcentaje de x_j excediendo al de la componente x_i . Por otro lado, un valor de $\hat{\tau}_{ij}$ pequeño, muestra que hay poca variabilidad relativa entre las componentes x_i y x_j . Además, si se tiene que $\hat{\xi}_{ij} > \sqrt{\hat{\tau}_{ij}}$, esto indica que la proporción de la componente x_i es apreciablemente más grande que la proporción de la componente x_j .

Aunque esta *matriz de variación* es teórica, se enfatiza en que es posible la estimación de $\hat{\xi}_{ij}$ y $\hat{\tau}_{ij}$ y como Aitchison (1982) [2] lo menciona, es posible interpretarla.

Una medida que explica la relación existente entre dos variables es la covarianza. A continuación se presentan algunos resultados importantes en el contexto de datos composicionales.

Un concepto importante en el análisis estadístico de datos es la asociación, para los datos composicionales su análisis es relevante para el manejo correcto de este tipo de datos. Como se ha mencionado anteriormente, una composición \mathbf{x} puede ser completamente determinada por D cocientes tales

como $\frac{x_i}{x_D}$, ($i = 1, \dots, D$).

De acuerdo a la propuesta de Aitchison, los datos composicionales se trabajan en términos de magnitudes relativas. Por esta razón la covarianza queda definida como la covarianza de las magnitudes relativas del dato composicional, es decir,

$$Cov\left(\frac{x_i}{x_k}, \frac{x_j}{x_l}\right)$$

con i, j, k, l valores en $\{1, \dots, D\}$.

Aún así se tiene la dificultad de trabajar en el espacio de los cocientes (ortante positivo \mathbb{R}_+^{D-1}). Para trabajar en todo el espacio \mathbb{R}^{D-1} , Aitchison propone tomar logaritmos de los cocientes entre las partes. Así,

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

Formalizando las ideas anteriores, se tienen las siguientes definiciones.

Definición 1.12 *La estructura de covarianza de una composición \mathbf{x} de D -partes es el conjunto*

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

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

Definición 1.13 *Para dos elementos i y j de una composición \mathbf{x} de D -partes, la varianza log-cociente se define como*

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

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 1.14 (*Matriz de covarianza log-cociente*). Para una composición \mathbf{x} de D -partes la matriz de $D \times D$

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

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 D -partes es reemplazando el denominador por la media geométrica $g(\mathbf{x})$.

Definición 1.15 (*Matriz de Covarianza log-cociente Centrada*). Para una composición \mathbf{x} de D -partes la matriz $d \times d$

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

con $i, j = 1, \dots, D$ 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 \mathcal{S}^D para describir patrones de variabilidad en el Simplex. A continuación se revisan algunos de estos modelos.

1.1.5. Distribuciones de Probabilidad en el Simplex

La Distribución de Dirichlet

La *Distribución de Dirichlet*, $Dir(\alpha)$, es una familia de distribuciones de probabilidad multivariable, continua y parametrizada por un vector $\alpha = (\alpha_1, \dots, \alpha_K)$ real de términos positivos. Esta distribución es la generalización multivariable de la distribución beta. La distribución de Dirichlet de orden $K \geq 2$ con parámetros $\alpha_1, \dots, \alpha_K > 0$ tiene una función de densidad de probabilidad en el espacio euclidiano \mathbb{R}^{K-1} dada por:

$$f(x_1, \dots, x_K | \alpha_1, \dots, \alpha_K) = \frac{1}{B(\alpha)} \prod_{i=1}^K x_i^{\alpha_i - 1}$$

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 es una distribución definida en el Simplex abierto de $(K - 1)$ -dimensional definido por:

$$\begin{aligned} x_1, \dots, x_{K-1} &> 0 \\ x_1 + \dots + x_{K-1} &< 1 \\ x_K &= 1 - x_1 - \dots - x_{K-1} \end{aligned}$$

y cero en otro caso.

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

$$B(\alpha) = \frac{\prod_{i=1}^K \Gamma(\alpha_i)}{\Gamma(\sum_{i=1}^K \alpha_i)}$$

con $\alpha = (\alpha_1, \dots, \alpha_K)'$.

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

La Distribución Normal Logística Aditiva

Una composición aleatoria \mathbf{X} tiene una *distribución normal en el Simplex* (o *Normal Logística Aditiva*) con un vector de medias μ y matriz de varianza Σ , denotada como $\mathcal{N}^D(\mu, \Sigma)$, 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 $\langle \mu, u \rangle_A$ y varianza $clr(u) \cdot \Sigma \cdot clr^t(u)$. 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

$$f(\mathbf{x}|\mu_{\mathbf{V}}, \Sigma_{\mathbf{V}}) = (2\pi)^{-\frac{(D-1)}{2}} |\Sigma_{\mathbf{V}}|^{-\frac{1}{2}} \exp \left[-\frac{1}{2} (ilr(\mathbf{x} - \mu_{\mathbf{V}}) \cdot \Sigma_{\mathbf{V}}^{-1} \cdot (ilr(\mathbf{x}) - \mu_{\mathbf{V}})^t \right]$$

con $\mu_{\mathbf{V}}$ y $\Sigma_{\mathbf{V}}$, respectivamente, sus vectores de medias y matriz de varianza.

1.2. Datos Direccionales

Los datos direccionales tienen que ver con observaciones de vectores unitarios en la esfera q -dimensional. Cuando $q = 2$ los datos direccionales se denominan datos circulares. Cuando $q = 3$ se denominan datos esféricos, y cuando $q > 3$, en general se denominan datos direccionales.

Los datos direccionales aparecen en varias áreas de la ciencia, tal como en Meteorología (análisis de dirección del viento), Biología (dirección de navegación de animales), Psicología (estudio de mapas mentales), Geología (estimación de rotaciones relativas de las placas tectónicas), Astronomía (distancia entre cuerpos celestes), entre otras. Los datos direccionales se pueden representar de diversas maneras, una de ellas es a través de puntos sobre la esfera unitaria $\mathbb{C}^q := \{u \in \mathbb{R}^q : u'u = 1\}$. Se debe notar que como $u \in \mathbb{C}^q$ es un vector unitario q -dimensional, éste también puede ser definido utilizando $q - 1$ ángulos. Así, en el caso general, un dato direccional se puede especificar utilizando coordenadas hiper-esféricas definidas por

$$\mathbf{u} = \begin{bmatrix} \cos(\theta_1) \\ \sin(\theta_1) \cos(\theta_2) \\ \sin(\theta_1) \sin(\theta_2) \cos(\theta_3) \\ \vdots \\ \sin(\theta_1) \cdots \sin(\theta_{q-2}) \cos(\theta_{q-1}) \\ \sin(\theta_1) \cdots \sin(\theta_{q-1}) \end{bmatrix} \quad (1.1)$$

donde $\theta_i \in [0, \pi]$ para $i = 1, \dots, q - 2$ y $\theta_{q-1} \in [0, 2\pi]$. En particular, cuando $q = 2$, las coordenadas hiper-esféricas se denominan coordenadas polares definidas por $\mathbf{u} = (\cos(\theta), \sin(\theta))'$, con $\theta \in [0, 2\pi]$. Se debe notar que para el caso de coordenadas esféricas $\theta = \theta_1$ y $\phi = \theta_2$ en la ecuación (1.1).

Dado $\mathbf{u} = (u_1, \dots, u_q) \in \mathbb{C}^q$ se pueden obtener los ángulos $\boldsymbol{\theta} = (\theta_1, \dots, \theta_{q-1})$ que lo definen bajo la siguiente transformación

$$\boldsymbol{\theta} = \begin{bmatrix} \tan^{-1}(\sqrt{\sum_{i=2}^q u_i^2}/u_1) \\ \tan^{-1}(\sqrt{\sum_{i=3}^q u_i^2}/u_2) \\ \tan^{-1}(\sqrt{\sum_{i=4}^q u_i^2}/u_3) \\ \vdots \\ \tan^{-1}(\sqrt{\sum_{i=q-1}^q u_i^2}/u_{q-2}) \\ \tan^{-1}(u_q/u_{q-1}) \end{bmatrix} \quad (1.2)$$

donde $\tan^{-1}(\cdot)$ toma valores en $(-\frac{\pi}{2}, \frac{\pi}{2})$.

1.2.1. Problemas en el análisis de Datos Direccionales

De la sección anterior, se definen formalmente los datos direccionales como:

Definición 1.16 *Un dato direccional es un vector $\mathbf{u} = (u_1, u_2, \dots, u_q)$ de componentes no negativas cuyo espacio muestral es la esfera unitaria \mathbb{C}^q definido como:*

$$\mathbb{C}^q := \left\{ \mathbf{u} = (u_1, u_2, \dots, u_q) \mid u_1 \geq 0, \dots, u_q \geq 0; \sum_{i=1}^q u_i^2 = 1 \right\}$$

Debido a que los datos direccionales presentan características especiales y diferentes a los datos definidos en la recta real, es necesario establecer distribuciones en la esfera unitaria que consideren estas propiedades. En particular,

una manera de especificar modelos de probabilidad es por medio de funciones de distribución. De esta manera si θ es un ángulo aleatorio, su función de distribución F se define como una función que cumple

$$F(x) = P(0 < \theta \leq x),$$

donde $0 \leq x \leq 2\pi$ y además

$$F(x + 2\pi) - F(x) = 1,$$

si $-\infty < x < \infty$. Nótese que esta última condición indica que cualquier arco de longitud 2π en el círculo unitario tiene probabilidad 1.

Ahora bien, por definición se tiene que

$$F(0) = P(0 < \theta \leq 0) = 0 \quad \text{y} \quad F(2\pi) = 1,$$

y si $\alpha \leq \beta \leq \alpha + 2\pi$, entonces

$$P(\alpha < \theta \leq \beta) = F(\beta) - F(\alpha).$$

Por otro lado, a diferencia de las funciones de distribución sobre la recta real, las funciones sobre el círculo unitario cumplen que:

$$\lim_{x \rightarrow \infty} F(x) = \infty, \quad \lim_{x \rightarrow -\infty} F(x) = -\infty.$$

En términos generales los modelos de probabilidad para describir datos direccionales se pueden clasificar en tres grandes categorías:

1. Modelos *wrapped* o envueltos:

- Normal envuelta
- Poisson envuelta

2. Modelos von Mises-Fisher:

- von Mises
- Fisher

3. Modelos generados por Proyección Radial:

- Normal Proyectada
- Gamma Proyectada

Los modelos *wrapped* destacan por ser modelos construidos a partir de envolver distribuciones escalares al rededor de la esfera unitaria. Los modelos tipo von Mises tienen como distribución principal a la distribución von Mises-Fisher, es una de las más utilizadas en el análisis inferencial de este tipo de datos. Por último, se tienen distribuciones generadas por proyecciones, dentro de las cuales la más importante es la Normal proyectada. Estas distribuciones se generan a partir de la proyección radial de un vector aleatorio multivariado. Una característica importante de estos modelos es su versatilidad ya que con éstos es posible modelar comportamientos simétricos, asimétricos, unimodales y multimodales. En este trabajo de tesis el enfoque será un modelo generado por la proyección radial de una distribución Gamma multivariada.

A pesar de tener diversos modelos, el interés por el estudio de datos direccionales ha sido poco explorado y es necesario contar con estructuras más completas, ya que una de las características principales de este tipo de datos es la periodicidad.

Para una revisión más detallada sobre las propiedades de las funciones de distribución y modelos de probabilidad sobre la esfera unitaria, se sugiere al lector consultar Mardia y Jupp (2000) [19].

1.2.2. Medidas Descriptivas

Sean $\mathbf{u}_1, \dots, \mathbf{u}_n$ puntos sobre la esfera unitaria \mathbb{C}^q . Entonces estos puntos pueden ser resumidos a través de su media muestral $\bar{\mathbf{u}} \in \mathbb{R}^{q+1}$, dada por

$$\bar{\mathbf{u}} = \frac{1}{n} \sum_{i=1}^n u_i$$

La *dirección media* de un conjunto de datos direccionales se define en términos de sus ángulos de la siguiente manera.

Definición 1.17 La *dirección media* $\bar{\boldsymbol{\theta}}$ de $\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_n$ se define como la dirección de la resultante $\mathbf{u}_1 + \dots + \mathbf{u}_n$ de los vectores $\mathbf{u}_1, \dots, \mathbf{u}_n$. La cual es también la dirección del centro de masa, $\bar{\mathbf{u}}$, de $\mathbf{u}_1, \dots, \mathbf{u}_n$.

Por ejemplo, en el caso de datos circulares, las coordenadas cartesianas de \mathbf{u}_i son $(\cos(\theta_i), \sin(\theta_i))$ para $i = 1, \dots, n$. Entonces las coordenadas cartesianas del centro de masa de \mathbf{u} resultan ser (\bar{u}_1, \bar{u}_2) donde

$$\bar{u}_1 = \frac{1}{n} \sum_{i=1}^n \cos(\theta_i) \quad \bar{u}_2 = \frac{1}{n} \sum_{i=1}^n \sin(\theta_i) \quad (1.3)$$

Así, $\bar{\theta}$ es la solución a las ecuaciones

$$\bar{u}_1 = \bar{R} \cos(\bar{\theta}) \quad \bar{u}_2 = \bar{R} \sin(\bar{\theta}) \quad (1.4)$$

donde

$$\bar{R} = (\bar{u}_1^2 + \bar{u}_2^2)^{1/2}$$

es la *longitud media resultante*. Cabe mencionar que $0 \leq \bar{R} \leq 1$. Si las direcciones $\theta_1, \dots, \theta_n$ están fuertemente agrupadas entonces \bar{R} puede ser casi 1. Por otra parte, si $\theta_1, \dots, \theta_n$ están dispersas \bar{R} es casi 0. Si $\bar{R} = 0$ entonces $\bar{\theta}$ no está definida. Cuando $\bar{R} > 0$, $\bar{\theta}$ está dada explícitamente por

$$\bar{\theta} = \begin{cases} \tan^{-1}(\bar{u}_2/\bar{u}_1) & \text{si } \bar{u}_1 \geq 0 \\ \tan^{-1}(\bar{u}_2/\bar{u}_1) + \pi & \text{si } \bar{u}_1 < 0. \end{cases}$$

Para el caso de datos esféricos (ver Mardia et al.(2000) [19]), las coordenadas de \mathbf{u}_i resultan $(\cos(\theta), \sin(\theta) \cos(\phi), \sin(\theta) \sin(\phi))$ para $i = 1, \dots, n$, entonces las coordenadas cartesianas del centro de masa de \mathbf{u} son $\{\bar{u}_1, \bar{u}_2, \bar{u}_3\}$, donde

$$\bar{u}_1 = \frac{1}{n} \sum_{i=1}^n \cos(\theta_i) \quad \bar{u}_2 = \frac{1}{n} \sum_{i=1}^n \sin(\theta_i) \cos(\phi_i), \quad \bar{u}_3 = \frac{1}{n} \sum_{i=1}^n \sin(\theta_i) \sin(\phi_i) \quad (1.5)$$

Por lo tanto, $\bar{\theta}$ y $\bar{\phi}$ son las soluciones a las ecuaciones

$$\bar{u}_1 = \bar{R} \cos(\bar{\theta}) \quad \bar{u}_2 = \bar{R} \sin(\bar{\theta}) \cos(\bar{\phi}) \quad \bar{u}_3 = \bar{R} \sin(\bar{\theta}) \sin(\bar{\phi}) \quad (1.6)$$

Análogamente,

$$\bar{R} = (\bar{u}_1^2 + \bar{u}_2^2 + \bar{u}_3^2)^{1/2}$$

Cuando $\bar{R} > 0$, $\bar{\theta}$ y $\bar{\phi}$ están dadas explícitamente por

$$\bar{\theta} = \begin{cases} \tan^{-1}((\bar{u}_2 + \bar{u}_3)^{1/2}/\bar{u}_1) & \text{si } \bar{u}_1 \geq 0 \\ \tan^{-1}((\bar{u}_2 + \bar{u}_3)^{1/2}/\bar{u}_1) + \pi & \text{si } \bar{u}_1 < 0. \end{cases}$$

$$\bar{\phi} = \begin{cases} \tan^{-1}(\bar{u}_3/\bar{u}_2) & \text{si } \bar{u}_2 \geq 0 \\ \tan^{-1}(\bar{u}_3/\bar{u}_2) + \pi & \text{si } \bar{u}_2 < 0. \end{cases}$$

Para propósitos descriptivos e inferenciales, la longitud media resultante es más importante que cualquier medida de dispersión. Sin embargo, a veces es útil contar con medidas de dispersión para datos direccionales. La más simple de éstas es la *varianza muestral direccional*.

Definición 1.18 *La varianza muestral direccional se define como*

$$V = 1 - \bar{R}$$

con $0 \leq V \leq 1$.

Otra medida de gran ayuda es la *desviación estándar*.

Definición 1.19 *La desviación estándar direccional se encuentra dada por*

$$v = \{-2 \log(1 - V)\}^{1/2}$$

1.2.3. Métodos Gráficos

Gráficamente los datos circulares se representan como puntos en la circunferencia de un círculo unitario, a este tipo de gráficos se les llama *histogramas circulares*. En la Figura 1.4 Se muestran los datos de 75 tortugas después de depositar sus huevos. Los datos fueron tomados de la Tabla 1.5 de Mardia et al. (2000) [19].

Este tipo de gráfica presenta los datos de forma desagrupada; sin embargo, es posible representar los datos circulares agrupándolos de manera similar a los histogramas que se utilizan para los datos en \mathbb{R} .

El *histograma de rosas* es una versión muy útil del histograma para datos reales, en el cual se reemplazan las barras por sectores que parecen pétalos de rosas. El radio de cada sector debe ser proporcional a la frecuencia del grupo correspondiente. Para lograr esta relación entre el área y la frecuencia de los grupos se utilizan diferentes radios; una convención es que el radio sea la raíz cuadrada de la frecuencia del grupo. Como se puede observar el correspondiente diagrama de rosas, en la Figura 1.5, para los datos de 75 tortugas

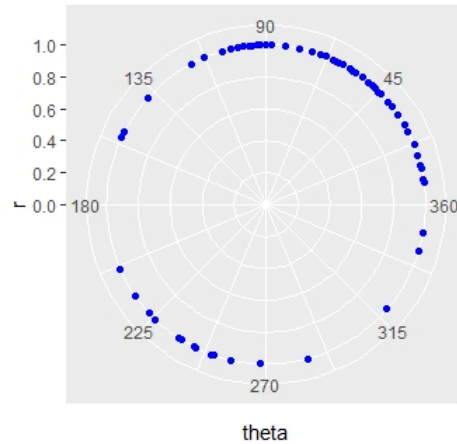


Figura 1.4: Diagrama Circular de 75 datos de tortugas.

donde el Radio (r) es la variable que contiene el radio de este conjunto de datos.

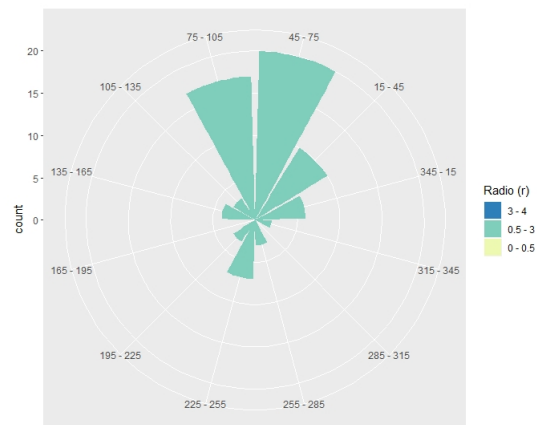


Figura 1.5: Diagrama de Rosa de 75 datos de tortugas.

Estas representaciones gráficas de los datos circulares capturan con mayor detalle la naturaleza direccional de las observaciones. Tanto el diagrama circular como el diagrama de rosas proporcionan una referencia visual intuitiva sobre las direcciones medidas individualmente, así como en grupo.

1.3. Estadística Bayesiana

En este trabajo de tesis el enfoque estadístico para llevar a cabo inferencias en los modelos propuestos será el enfoque Bayesiano de la estadística. Por este motivo, a continuación se presentan algunos de los conceptos más relevantes acerca de este tema. Existe una gran variedad de referencias sobre la aparición y desarrollo de la estadística Bayesiana, la profundidad de los aspectos que se relacionan a ella pueden consultarse en Box y Tiao (1973)[7], Bernardo (1981)[4], Bernardo y Smith (2000)[5], Gelman et al. (2004) [12], Hoff (2009) [15], entre otros.

El trabajo de Thomas Bayes, publicado de forma póstuma en 1763, impactó de manera importante el modo en que se hace inferencia estadística: ofrece las bases para combinar el conocimiento inicial (*a priori*) de un fenómeno, con el nuevo conocimiento adquirido de los datos y nuevas mediciones, obteniendo información final (*a posteriori*). Es decir, el conocimiento inicial se renueva con la información obtenida de una muestra aleatoria en el conocimiento final y dicho proceso de actualización es de naturaleza secuencial.

Desde la perspectiva de la Estadística Bayesiana, se considera la información extra como información subjetiva, la cual se refiere a toda aquella información inicial que se tiene en relación al fenómeno aleatorio de interés, anterior a la recolección o la información hasta ese momento sobre él, y esto incluye: datos históricos, teorías, opiniones y conjeturas de expertos, conclusiones basadas en estudios previos, etc. Como primer paso la inferencia bayesiana traduce todo lo anterior en una distribución inicial (*a priori*) para tener como segundo paso la recolección de mediciones que actualizan esa distribución inicial y obtiene, mediante la Regla de Bayes, una distribución de probabilidad final (*a posteriori*).

El enfoque frecuentista por su parte limita en su análisis, información extra disponible y se apoya únicamente en los datos muestrales observados. Si hay pocos datos, la estadística frecuentista presenta problemas puesto que muchos de sus métodos se apoyan en resultados asintóticos, como la Ley de los Grandes Números y el Teorema Central del Límite.

Así, la estadística Bayesiana toma como elemento central el conocido Teorema de Bayes, el cual se presenta a continuación para el caso de eventos.

Teorema 1.1 *Sea $(\Omega, \mathcal{A}, \mathbb{P})$ un espacio de probabilidad. Sea $\{A_1, A_2, \dots, A_n\}$ una partición de Ω , tales que $P(A_j) > 0 \forall j = 1, \dots, n$. Sea B un evento en \mathcal{A} . Entonces,*

$$\begin{aligned} P(A_j|B) &= \frac{P(B|A_j)P(A_j)}{\sum_{k=1}^n P(B|A_k)P(A_k)} \\ &= \frac{P(B|A_j)P(A_j)}{P(B)}, \quad \forall j = 1, \dots, n. \end{aligned}$$

En el caso de inferencias para un parámetro ϕ asociado a un modelo probabilístico se tiene lo siguiente. Sea $p(x|\phi)$ un modelo probabilístico paramétrico donde la función $p(\cdot|\phi)$ puede ser una función de masa de probabilidad de una variable (o vector) aleatoria (v.a) discreta o bien una función de densidad de una v.a. continua. Para hacer referencia a una muestra aleatoria se utilizará la notación $\mathbf{X} := (X_1, \dots, X_n)$ y para referirnos a una observación muestral utilizaremos $\mathbf{x} := (x_1, \dots, x_n)$. Así, el espacio paramétrico será el conjunto Φ de todos los valores que puede tomar ϕ , es decir la familia paramétrica asociada se define como $\mathcal{P} = \{p(x|\phi) : \phi \in \Phi\}$.

La estadística Bayesiana modela la incertidumbre sobre ϕ probabilísticamente, esto es, se considera una distribución de probabilidad para describir la incertidumbre sobre ϕ . En este proceso, se define un elemento $p(\phi)$ denominado **distribución inicial** o **a priori**. En términos generales, esta distribución inicial describe la experiencia previa (experiencia de especialistas, datos históricos, análisis previos, etc.) antes de obtener datos muestrales nuevos. A continuación se procede a observar los nuevos datos (obtención de la muestra) \mathbf{x} y se combina esta información con la distribución a priori mediante la Regla de Bayes, lo que resulta en una **distribución de probabilidad a posteriori** (o **final**):

$$p(\phi|\mathbf{x}) = \frac{p(\mathbf{x}, \phi)}{p(\mathbf{x})} = \frac{p(\mathbf{x}|\phi)p(\phi)}{\int_{\Phi} p(\mathbf{x}|\phi)p(\phi)d\phi} \quad (1.7)$$

Se tiene que $p(\phi|\mathbf{x})$ es una distribución de probabilidad de ϕ y a diferencia de la distribución a priori, $p(\phi)$, toma en cuenta tanto la información contemplada en $p(\phi)$ así como la contenida en los datos observados \mathbf{x} . Desde el

punto de vista Bayesiano de la estadística, la distribución a posteriori es la base para hacer inferencias sobre ϕ .

Es importante notar que $p(\mathbf{x})$ es constante respecto a ϕ , por lo que el Teorema de Bayes se suele escribir como

$$p(\phi|\mathbf{x}) \propto p(\mathbf{x}|\phi)p(\phi) \quad (1.8)$$

De la ecuación anterior se tiene que $p(\mathbf{x}|\phi) = p(x_1, \dots, x_n|\phi)$ se trata de la probabilidad conjunta de la muestra condicional en ϕ llamada también **verosimilitud**.

Por otro lado, frecuentemente se está interesado en el comportamiento de observaciones futuras del fenómeno aleatorio en cuestión, esto es, hacer una *predicción*. Dado un valor de ϕ , la distribución que describe el comportamiento de la observación futura X es $p(x|\phi)$. El problema es que por lo general el valor de ϕ es desconocido. En la estadística frecuentista se aborda este problema estimando puntualmente a ϕ con base en la muestra observada y dicho estimador $\hat{\phi}$ es sustituido en $p(x|\phi)$, es decir, utilizan $p(x|\hat{\phi})$. Desde la perspectiva bayesiana el modelo $p(x|\phi)$ junto con la distribución a priori $p(\phi)$ inducen una distribución conjunta para el vector aleatorio (X, ϕ) mediante el concepto de probabilidad condicional:

$$p(x, \phi) = p(x|\phi)p(\phi). \quad (1.9)$$

Así, marginalizando esta distribución de probabilidad conjunta se obtiene:

$$p(x) = \int_{\Phi} p(x, \phi)d\phi = \int_{\Phi} p(x|\phi)p(\phi)d\phi. \quad (1.10)$$

la cual se le denomina **distribucion predictiva a priori (o inicial)**, y describe el conocimiento acerca de una observación futura X basado únicamente en la información contenida en $p(\phi)$. Nótese que $p(x)$ no depende ya de ϕ .

Una vez obtenida la muestra, el modelo $p(x|\phi)$ y la distribución a posteriori $p(\phi|\mathbf{x})$ inducen una distribución conjunta para (X, ϕ) condicional en los

valores observados \mathbf{x} :

$$\begin{aligned} p(x, \phi | \mathbf{x}) &= \frac{p(x, \phi, \mathbf{x})}{p(\mathbf{x})} \\ &= \frac{p(x|\phi, \mathbf{x})p(\phi, \mathbf{x})}{p(\mathbf{x})} \\ &= p(x|\phi, \mathbf{x})p(\phi|\mathbf{x}) \\ &= p(x|\phi)p(\phi|\mathbf{x}) \end{aligned}$$

La última igualdad se da puesto que $p(x|\phi, \mathbf{x}) = p(x|\phi)$ por la independencia condicional de X y \mathbf{X} dado ϕ . Marginalizando la distribución conjunta:

$$\begin{aligned} p(x|\mathbf{x}) &= \int_{\Phi} p(x, \phi|\mathbf{x})d\phi \\ &= \int_{\Theta} p(x|\phi)p(\phi|\mathbf{x})d\phi. \end{aligned}$$

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 $p(\phi)$ como en la información muestral \mathbf{x} . Hay que notar nuevamente que $p(x|\mathbf{x})$ no depende de ϕ .

1.4. Métodos de Simulación

Como se menciona en la sección anterior, la estadística Bayesiana requiere del cálculo de integrales complejas de funciones multidimensionales. La resolución de estas integrales es necesaria, entre otras cosas, para obtener (la constante de proporcionalidad de la distribución final de los parámetros de interés) las distribuciones marginales, la distribución predictiva, así como la esperanza y varianza a posteriori, según se requiera.

En muchos casos no es posible evaluar explícitamente estas integrales, por lo que es necesario formular integraciones numéricas o emplear métodos de aproximación analítica. Aplicar ésta serie de métodos puede llegar a ser complicado por la dimensión del espacio paramétrico o incluso por la complejidad de las expresiones a las que se puede llegar. Por lo anterior, se han desarrollado una serie de técnicas llamadas *Métodos de Monte Carlo vía Cadenas de Markov* (MCMC), los cuales son otra forma de simulación estocástica que proporcionan algoritmos para muestrear de una distribución final de manera

aleatoria e indirecta y con ello obtener estimaciones de las cantidades de interés deseadas.

Los métodos de simulación se basan en la generación directa o indirecta de variables aleatorias distribuidas de acuerdo a una distribución f , que no necesariamente es explícitamente conocida. Algunos de estos métodos son el método de la transformación inversa, el algoritmo Box-Müller, el muestreo por aceptación y rechazo, el muestreo de rechazo adaptativo, entre otros. (Ver Robert, et al. (2004) [28] y Gamerman & Freitas (2006) [11] para más información).

1.4.1. Muestreo por aceptación y rechazo

En algunos casos, no se dispone de una forma accesible de simulación para la distribución de la cual se quieren generar variables aleatorias. Es por esto que en ocasiones se hace uso de otros métodos que solo requieren del conocimiento de la forma funcional de la densidad de interés f multiplicada por una constante. En el muestreo por aceptación y rechazo se generan realizaciones de una densidad g de la cual la simulación es factible realizar.

Sea $X \sim f$ y sea $g(x)$ una función de densidad que satisface que $f(x) \leq Mg(x)$ para alguna constante $M \geq 1$. Entonces, para simular $X \sim f$, es suficiente generar $Y \sim g$ y $U|Y = y \sim \mathcal{U}(0, Mg(y))$ hasta que $0 < u < f(y)$. (Ver Gamerman et. al (2006) [11]). Esta forma de simular trae consigo dos consecuencias importantes. La primera es que provee de un método genérico para simular de cualquier densidad f que es conocida excepto tal vez, por una constante, ya que el método solo requiere del cociente f/M , el cual no depende de esa constante. Esta propiedad es particularmente importante en los cálculos de la Estadística Bayesiana, donde una cantidad de interés es la distribución posterior $p(\phi|x)$,

$$p(\phi|x) \propto p(\phi)f(x|\phi).$$

ya que es especificada salvo una constante de normalización, y al usar el método de aceptación y rechazo se observa que no se necesita conocer esa constante. Por supuesto, queda la tarea de encontrar una densidad g que satisfaga $f \leq Mg$, una cota que no necesita ser justa, en el sentido de que el método de aceptación y rechazo es válido cuando M es cualquier constante

grande.

Una segunda consecuencia es que la probabilidad de aceptación es exactamente $1/M$ cuando se evalúa la densidad correctamente normalizada y el número esperado de intentos hasta aceptar una realización es M .

Así, el algoritmo para el método de Aceptación y Rechazo queda dado por los siguientes pasos:

1. Generar $X \sim g$, $U \sim \mathcal{U}[0, 1]$.
2. Hacer $Y = X$ si $U \leq f(X)/Mg(X)$.
3. Regresar a 1 en otro caso.

1.4.2. Muestreo por rechazo adaptativo

El muestreo por rechazo adaptativo es una técnica que permite generar muestras de cualquier función de densidad de probabilidad log-cóncava $f(x)$ y se basa en el muestreo por rechazo, el cual no requiere determinar $f(x)$ explícitamente. El término *adaptable* refiere que la función envolvente y la de compresión (las cuales forman los límites superior e inferior en este método, respectivamente) convergen a la densidad $f(x)$ a medida que el muestreo avanza. Una de las ventajas de esta técnica de muestreo es que se pueden extraer muestras con pocas evaluaciones de $f(x)$, lo que resulta útil cuando su evaluación es costosa.

El muestreo de rechazo adaptativo (ARS, por sus siglas en inglés) reduce el número de evaluaciones de $g(x)$ en dos formas. Primero, al suponer la log-concavidad de $f(x)$ se evita la necesidad de ubicar el supremo de $g(x)$ en D (donde D denota el soporte de $f(x)$, es decir el conjunto de x tales que $f(x) > 0$). En segundo lugar, después de cada rechazo, la probabilidad de necesitar evaluar cada rechazo en $g(x)$ se reduce al actualizar las funciones envolvente y de compresión para incorporar la información más reciente sobre $g(x)$.

Por lo tanto el método consiste en lo siguiente. Se supone que D es conexo, que $g(x)$ es continua y diferenciable en todas partes y que $h(x) = \ln(g(x))$

es cóncavo en todas partes en D . Esta definición de log-concavidad admite tanto segmentos de recta en $h(x)$ como discontinuidades en $h'(x)$. La curva continua en la Figura 1.6 ejemplifica una $h(x)$ en un dominio D y sus envolventes superior e inferior en líneas punteadas.

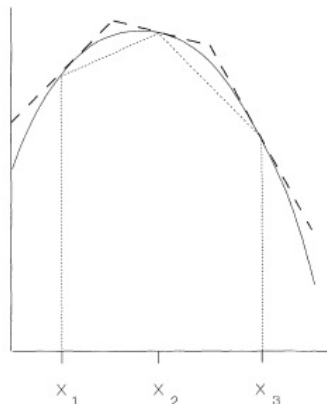


Figura 1.6: Una log-densidad $h(x)$ (línea continua), delimitada a la izquierda, junto con sus envolventes superior (línea punteada) e inferior (línea punteada) basados en 3 abscisas.

Supóngase que $h(x)$ y $h'(x)$ han sido evaluadas en k abscisas sobre D , $x_1 \leq x_2 \leq \dots \leq x_k$. Sea $T_k = \{x_i : i = 1, \dots, k\}$. Defínase el rechazo envolvente sobre T_k como $\exp(u_k(x))$ donde $u_k(x)$ es un casco superior lineal por tramos formado desde las tangentes de $h(x)$ en las abscisas en T_k como se puede observar en la línea punteada superior de la Figura 1.6. Para $j = 1, \dots, k - 1$ las tangentes en x_j y x_{j+1} intersectan en

$$z_j = \frac{h(x_{j+1}) - h(x_j) - x_{j+1}h'(x_{j+1}) + x_j h'(x_j)}{h'(x_j) - h'(x_{j+1})}.$$

Así para $x \in [z_{j-1}, z_j]$ con $j = 1, \dots, k$, se define

$$u_k(x) = h(x) + (x - x_j)h'(x)$$

donde z_0 es la cota inferior de D (ó $-\infty$ si D no está limitado por debajo) y z_k es la cota superior de D (ó $+\infty$ si D no está acotado por arriba). También se define

$$s_k = \frac{\exp(u_k(x))}{\int_D \exp(u_k(x')) dx'}.$$

Finalmente, se define la función de compresión sobre T_k como $\exp(l_k(x))$, donde $l_k(x)$ es un cascarón inferior de líneas a trozos de las abscisas adyacentes

en T_k como puede observarse en la Figura 1.6 por la línea punteada inferior. Por lo tanto para $x \in [x_j, x_{j+1}]$

$$l_k(x) = \frac{(x_{j+1} - x)h(x_j) + (x - x_j)h(x_{j+1})}{x_{j+1} - x_j}$$

para $j = 1, \dots, k - 1$. Para $x < x_k$ definimos $l_k = -\infty$. Así, la envolvente de rechazo y la función de compresión son funciones exponenciales por partes. La concavidad de $h(x)$ asegura que $l_k(x) \leq h(x) \leq u_k(x)$ para toda $x \in D$. (Ver Gilks y Wild (1992) [13]). Para muestrear n puntos independientes de $f(x)$ por muestreo de rechazo adaptativo deben realizarse los siguientes pasos y repetirlos hasta que se hayan aceptado los n puntos:

1. Paso de inicialización: Inicializar las abscisas en T_k . Si D no es acotado por la izquierda elegir x_1 tal que $h'(x) > 0$. Si D no es acotado por la derecha elegir x_k tal que $h'(x) < 0$. Teniendo definido k de inicio, calcular la función $u_k(x)$, $s_k(x)$ y $l_k(x)$.
2. Paso de muestreo: Muestrear un valor x^* de $s_k(x)$ y muestrear un valor independiente ω de la distribución uniforme $(0,1)$. Si

$$\omega \leq \exp\{l_k(x^*) - u_k(x^*)\}$$

entonces aceptar x^* . De otra manera evaluar $h(x^*)$ y $h'(x^*)$ y realizar el siguiente test de rechazo: Si

$$\omega \leq \exp\{h(x^*) - u_k(x^*)\}$$

entonces aceptar x^* ; De otra forma rechazar x^* .

3. Paso de actualización: Si $h(x^*)$ y $h'(x^*)$ fueron evaluados en el paso de muestreo, incluir x^* en T_k para formar T_{k+1} ; reetiquetar los elementos de T_{k+1} en orden ascendente, construir las funciones $u_{k+1}(x)$, $s_{k+1}(x)$ y $l_{k+1}(x)$ sobre la base de T_{k+1} ; incrementar k . Regresar al paso de muestreo si los n puntos aún no han sido aceptados.

1.4.3. Un acercamiento 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 1.20 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 1.21 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 1.22 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 1.23 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 1.24 Se dice que una distribución π 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 1.25 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, podemos enunciar el siguiente teorema importante y de gran utilidad en el desarrollo práctico de la estadística bayesiana.

Proposición 1.1 *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 $\pi(\cdot)$.

Definición 1.26 *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, uno puede proponer una infinidad de implementaciones prácticas. Algunos métodos MCMC son el Metrópolis Hastings, el Slice Sampler, el Gibbs Sampler, 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.

El algoritmo de Gibbs sampler fue propuesto en 1984 por Geman y Geman 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 (1990) señalaron que el Gibbs sampler podría servir para estudiar otras distribuciones como las distribuciones finales en estadística bayesiana.

El algoritmo del Gibbs sampler 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 $p_i(\theta_i|\theta_{-i}, \mathbf{y})$, $i = 1, \dots, d$, son totalmente conocidas y es posible simular de ellas. Entonces, el Gibbs sampler permite obtener una muestra de la distribución final (conjunta) de los parámetros dados los datos, $p(\boldsymbol{\theta}|\mathbf{y})$, a partir de las $p_i(\theta_i|\theta_{-i}, \mathbf{y})$. 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 método de muestreo de Gibbs 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 $p(\boldsymbol{\theta}|\mathbf{y})$, lo anterior puede describirse de la siguiente manera:

Dado $\boldsymbol{\theta}^{(0)} = (\theta_1^{(0)}, \theta_2^{(0)}, \dots, \theta_d^{(0)})'$,

1. Inicializar el contador $t = 0$
2. Generar $\boldsymbol{\theta}^{(t+1)} = (\theta_1^{(t+1)}, \theta_2^{(t+1)}, \dots, \theta_d^{(t+1)})'$ de la siguiente manera:
 - a) Generar $\theta_1^{(t+1)}$ de $\pi(\theta_1|\theta_2^{(t)}, \dots, \theta_d^{(t)})$
 - b) Generar $\theta_2^{(t+1)}$ de $\pi(\theta_2|\theta_1^{(t+1)}, \theta_3^{(t)}, \dots, \theta_d^{(t)})$
 - c) Generar $\theta_3^{(t+1)}$ de $\pi(\theta_3|\theta_1^{(t+1)}, \theta_2^{(t+1)}, \theta_4^{(t)}, \dots, \theta_d^{(t)})$
 - ⋮
 - d) Generar $\theta_d^{(t+1)}$ de $\pi(\theta_d|\theta_1^{(t+1)}, \theta_2^{(t+1)}, \dots, \theta_{d-1}^{(t+1)})$
3. Actualizar el contador a $t = t + 1$ y regresar al paso 2.

El Gibbs sampler 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.

1.4.4. Evaluación de Convergencia

Un elemento φ 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 et al. (2006) [11] 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 por Gelfand y Smith (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. A continuación 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

Un modelo Bayesiano para Datos Composicionales

En este capítulo se describe el modelo de mezclas, el cual es la propuesta de esta tesis, para la modelación de datos en el Simplex D -dimensional. Comienza con la descripción del modelo Gamma como una breve introducción a este tipo de distribuciones. En una siguiente sección se retoma el modelo Gamma proyectado estudiado por Núñez-Antonio y Geneyro (2019) [22] como un modelo para datos direccionales. Por último, se explica el modelo de mezclas que se utilizará como propuesta y cómo a partir de un proceso Dirichlet, siendo discreto, es posible llegar a una representación del modelo de mezclas infinitas.

2.1. El Modelo Gamma

Una variable aleatoria Y se dice que se distribuye Gamma con parámetros de forma y escala $\alpha > 0$ y $\beta > 0$, respectivamente, si su función de densidad de probabilidad está dada por

$$f(y|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} y^{\alpha-1} e^{-\beta y} \quad ; \quad y > 0, \quad (2.1)$$

donde $\Gamma(\alpha)$ es la función Gamma definida como

$$\Gamma(\alpha) = \int_0^\infty y^{\alpha-1} e^{-y} dy.$$

Bajo esta parametrización, los momentos de Y pueden expresarse como

$$E(Y) = \frac{\alpha}{\beta} \quad \text{y} \quad V(Y) = \frac{\alpha}{\beta^2}.$$

Así, si se tiene una muestra aleatoria $\{Y_1, \dots, Y_n\}$ independiente e idénticamente distribuida del modelo $\text{Gamma}(\alpha, \beta)$, entonces la función de verosimilitud esta dada por

$$\begin{aligned} L(\alpha, \beta|Y) &= \prod_{i=1}^n f(y_i|\alpha, \beta) \\ &= \prod_{i=1}^n \frac{\beta^\alpha}{\Gamma(\alpha)} y_i^{\alpha-1} e^{-\beta y_i} \\ &= \frac{\beta^{n\alpha}}{[\Gamma(\alpha)]^n} \left(\prod_{i=1}^n y_i \right)^{\alpha-1} e^{-\beta \sum_{i=1}^n y_i}, \end{aligned}$$

donde $\prod_{i=1}^n y_i$ y $\sum_{i=1}^n y_i$ son estadísticos suficientes¹ para α y β .

En este trabajo el interés es describir variables que se encuentran restringidas al primer ortante con un modelo Gamma multivariado. Este modelo es lo suficientemente flexible y trata las observaciones como provenientes de una distribución multivariada la cual es generada como un producto de k distribuciones Gamma univariadas independientes.

El objetivo principal es entonces, hacer inferencia para los parámetros α y β , a partir de la distribución posterior

$$\begin{aligned} \pi(\alpha, \beta|Y) &= \frac{L(\alpha, \beta|Y) \cdot \pi(\alpha, \beta)}{\int L(\alpha, \beta|Y) \cdot \pi(\alpha, \beta) d\alpha d\beta} \\ &\propto L(\alpha, \beta|Y) \cdot \pi(\alpha, \beta). \end{aligned}$$

En secciones posteriores se mostrará la forma de hacer dicha inferencia para esta distribución posterior del modelo $\text{Gamma}(\alpha, \beta)$.

2.2. El Modelo de la Gamma Projectado

Una de las formas más simples para generar distribuciones en la esfera unitaria, S^k , es proyectar radialmente distribuciones de probabilidad origi-

¹Sea Y_1, \dots, Y_n una muestra aleatoria de una distribución de probabilidad con parámetro desconocido ϕ . Entonces se dice que el estadístico $U = g(Y_1, \dots, Y_n)$ es *suficiente* para ϕ si la distribución condicional de Y_1, \dots, Y_n , dada U , no depende de ϕ . Es decir, los estadísticos suficientes en cierto sentido resumen *toda* la información de una muestra acerca de un parámetro objetivo. Wackerly, et al. (2008) [32]

nalmente definidas en \mathbb{R}^k . Como se menciona en Núñez-Antonio y Geneyro (2019) [22], en su artículo sobre el análisis de la distribución Gamma Proyectada para datos direccionales, el estudio de modelos generados por proyecciones, en general, se ha pasado por alto. Pero dado los vastos ejemplos en donde pueden encontrarse este tipo de datos, es conveniente tener metodologías definidas únicamente en espacios muestrales restringidos de S^k . En consecuencia el modelo propuesto para este fin es el modelo Gamma proyectado multivariado el cual se define a continuación.

Dado un vector aleatorio \mathbf{Y} tal que $P(\mathbf{Y} = \mathbf{0}) = 0$, es posible hacer una proyección radial sobre la esfera unitaria. La proyección de \mathbf{Y} sobre la esfera unitaria esta dada por el vector direccional $\mathbf{U} = \|\mathbf{Y}\|^{-1}\mathbf{Y}$. Esto se sigue de la definición de distribuciones direccionales definidas por proyección radial de distribuciones multivariadas definidas en \mathbb{R}^k . Esta técnica puede usarse para generar distribuciones direccionales definidas solamente en subconjuntos de \mathbb{C}^k . Siguiendo la idea anterior, un modelo direccional definido solamente en el octante positivo de \mathbb{C}^k es el siguiente.

Sea $\mathbf{Y} = (y_1, y_2, \dots, y_k)^t$ un vector multivariado Gamma de dimensión k con la densidad conjunta definida por

$$f_{Y_1, \dots, Y_k}(\mathbf{y} | \boldsymbol{\alpha}, \boldsymbol{\beta}) = \prod_{i=1}^k Ga(y_i | \alpha_i, \beta_i), \quad (2.2)$$

donde $Ga(y_i | \alpha_i, \beta_i)$ es la densidad dada por una Gamma como en la ecuación (2.1). Aplicando la transformación a coordenadas esféricas k -dimensional

$\mathbf{Y} \rightarrow (R, \Theta)$, definida como

$$\begin{aligned}
Y_1 &= R \cos \Theta_1, \\
Y_2 &= R \sin \Theta_1 \cos \Theta_2, \\
Y_{k-2} &= R \sin \Theta_1 \sin \Theta_2 \cdots \sin \Theta_{k-3} \cos \Theta_{k-2} \\
&= R \cos \Theta_{k-2} \prod_{j=1}^{k-3} (\sin \Theta_j), \\
Y_{k-1} &= R \sin \Theta_1 \sin \Theta_2 \cdots \sin \Theta_{k-3} \sin \Theta_{k-2} \cos \Theta_{k-1} \\
&= R \cos \Theta_{k-1} \prod_{j=1}^{k-2} (\sin \Theta_j), \\
Y_k &= R \sin \Theta_1 \sin \Theta_2 \cdots \sin \Theta_{k-3} \sin \Theta_{k-2} \sin \Theta_{k-1} \\
&= R \prod_{j=1}^{k-1} (\sin \Theta_j),
\end{aligned}$$

donde $R = \|\mathbf{Y}\|$, se puede obtener la correspondiente densidad conjunta de R y $\Theta = (\Theta_1, \dots, \Theta_{k-1})$, denotada por $f_{R, \Theta}(r, \boldsymbol{\theta} | \boldsymbol{\alpha}, \boldsymbol{\beta})$, es decir,

$$f_{R, \Theta}(r, \boldsymbol{\theta} | \boldsymbol{\alpha}, \boldsymbol{\beta}) = f_{Y_1, \dots, Y_k}(r, \boldsymbol{\theta} | \boldsymbol{\alpha}, \boldsymbol{\beta}) \times |\mathbf{J}(r, \boldsymbol{\theta})|$$

donde $|\mathbf{J}(r, \boldsymbol{\theta})| = r^{k-1} \sin^{k-2} \theta_1 \sin^{k-3} \theta_2 \cdots \sin \theta_{k-2}$. Así,

$$\begin{aligned}
f_{R, \Theta}(r, \boldsymbol{\theta} | \boldsymbol{\alpha}, \boldsymbol{\beta}) &= Ga(r \cos \theta_1 | \alpha_1, \beta_1) \times \prod_{i=2}^{k-1} \left(Ga(r \cos \theta_i \prod_{j=1}^{i-1} (\sin \theta_j) | \alpha_i, \beta_i) \right) \\
&\quad \times Ga(r \prod_{j=1}^{k-1} (\sin \theta_j) | \alpha_k, \beta_k) \times |\mathbf{J}(r, \boldsymbol{\theta})| \\
&= \frac{\beta_k^{ak}}{\Gamma(\alpha_k)} \times \left(\prod_{j=1}^{k-1} \frac{\beta_j^{\alpha_j}}{\Gamma(\alpha_j)} (\cos \theta_j)^{\alpha_j-1} (\sin \theta_j)^{(\sum_{h=j+1}^k \alpha_h)-1} \right) \\
&\quad \times \exp \left\{ -r \left(\beta_1 \cos \theta_1 + \sum_{i=2}^{k-1} \left(\beta_i \cos \theta_i \prod_{j=1}^{i-1} \sin \theta_j \right) + \beta_k \prod_{j=1}^{k-1} \sin \theta_j \right) \right\}.
\end{aligned}$$

Entonces, resulta que Θ tiene una distribución definida solamente sobre el ortante positivo de \mathbb{C}^k y se le llamará *distribución Gamma Multivariada Proyectada*, con densidad dada por

$$PG_{k-1}(\boldsymbol{\theta} | \boldsymbol{\alpha}, \boldsymbol{\beta}) = \frac{\Gamma(A) \beta_k^{\alpha_k}}{(B)^A \Gamma(\alpha_k)} \left(\prod_{j=1}^{k-1} \frac{\beta_j^{\alpha_j}}{\Gamma(\alpha_j)} (\cos \theta_j)^{\alpha_j-1} (\sin \theta_j)^{(\sum_{h=j+1}^k \alpha_h)-1} \right) I_{(0, \pi/2)^{k-1}}(\boldsymbol{\theta}),$$

donde

$$A = \sum_{i=1}^k \alpha_i \quad \text{y} \quad B = \beta_1 \cos \theta_1 + \sum_{i=2}^{k-1} \left(\beta_i \cos \theta_i \prod_{j=1}^{i-1} \operatorname{sen} \theta_j \right) + \beta_k \prod_{j=1}^{k-1} \operatorname{sen} \theta_j.$$

Los parámetros α, β deben estimarse, pero no son identificables, ya que para cualquier $a > 0$, tomando $\beta^* = a\beta$ no se altera la distribución de las direcciones observadas. Una manera de resolver esta dificultad (propuesta en Nuñez-Antonio y Geneyro (2019) [22]) es tomar $\beta = (1, \beta_2, \beta_3, \dots, \beta_k)^t$. Modelos con estructuras más generales para β podrían ser más útiles. Sin embargo, este modelo es lo suficientemente flexible, puesto que puede describir comportamientos simétricos, asimétricos, y/o bimodales.

2.3. El Modelo Propuesto

Los modelos de mezclas ofrecen una gran versatilidad en la descripción de datos, lo que los convierte en herramientas ampliamente utilizadas en Estadística. Desde un punto de vista pragmático, una mezcla puede visualizarse como la combinación convexa de componentes que son, en si mismas, distribuciones de probabilidad; el número de estos componentes pueden ser finito o infinito numerable.

Bajo un enfoque Bayesiano, el primero en contextualizar de manera general los modelos de mezclas infinitas fue Lo(1984) [17] con expresiones para los pesos utilizando el proceso Dirichlet. Sin embargo, hasta que Escobar y West (1995) [9] propusieron una implementación computacional, se generó un aumento en el interés por este tipo de modelos. Algunas de las aplicaciones de estos modelos son la estimación de la densidad y el análisis de conglomerados.

Las funciones de probabilidad (aleatorias) generadas por un proceso Dirichlet son discretas casi seguramente (c.s.), esto es, se pueden escribir como

$$P(\cdot) = \sum_{j=1}^{\infty} \omega_j \delta_{\phi_j}(\cdot), \quad (2.3)$$

donde δ_{ϕ} denota la medida con una masa 1 en ϕ y los pesos ω_j son aleatorios que suman uno c.s. y ϕ_j son localizaciones aleatorias (independientes de los

pesos) distribuidas de acuerdo a una medida de probabilidad P_0 no atómica.

De esta manera, para un kernel K , el modelo de Lo(1984) [17] se obtiene mezclando los parámetros del kernel con la distribución P , esto es,

$$f_P(y) = \int K(y|\phi)dP(\phi) ,$$

y donde P se distribuye de acuerdo a un proceso Dirichlet. De acuerdo a la ecuación (2.3), el modelo se puede representar como

$$f(y|\omega, \phi) = \sum_{j=1}^{\infty} \omega_j K(y|\phi_j). \quad (2.4)$$

Una de las principales ventajas de un modelo de mezclas como éste es que se obtienen funciones de distribución absolutamente continuas y permiten modelar cualquier conjunto de datos que tomen valores en el mismo espacio que el soporte del kernel K . Nótese además que, aunque los parámetros asociados al modelo son la colección infinita de ω 's y de ϕ 's, en el fondo, el parámetro de interés es el modelo subyacente a los datos, es decir, el modelo de mezclas f_P .

Con la relación a los pesos aleatorios, existen diferentes metodologías tanto para construir sucesiones infinitas de éstos como para realizar inferencias. El procedimiento stick-breaking (ver Sethuraman y Tiwari (1982)[30]) ha sido ampliamente utilizado ya que permite realizar ambas tareas. Bajo este procedimiento, los pesos se construyen de la siguiente manera.

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

donde ν_j son variables aleatorias independientes que toman valores en el intervalo $(0, 1)$. En particular, para un proceso Dirichlet con parámetro de escala $M > 0$ y medida de probabilidad base P_0 , denotado por $DP(M, P_0)$, la distribución de las ν_j es una *Beta* con parámetros $(1, M)$.

El algoritmo de muestreo

Para llevar a cabo inferencias sobre el modelo (2.4), Walker (2007) [33] utiliza un muestreo de Gibbs, para lo cual se necesitan especificar todas las condicionales finales completas de los parámetros involucrados. En principio se deberían muestrear, en cada paso del muestreo de Gibbs, toda la colección infinita de las variables $\{\nu_j\}$ y $\{\phi_j\}$. Sin embargo, como se verá a continuación, solo es necesario muestrear un conjunto finito de ellas.

Para estimar el modelo vía un muestreador de Gibbs, se introduce una variable latente u tal que la densidad conjunta de (y, u) dado (ω, ν) está dada por

$$f(y, u|\omega, \nu) = \sum_{j=1}^{\infty} I_A(u < \omega_j) K(y|\phi_j), \quad (2.5)$$

donde $I_A(\cdot)$ es la función indicadora sobre el conjunto A . 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.

Algunas observaciones que se pueden hacer en este momento son:

1. Claramente, integrando sobre u se recupera la densidad $f(y|\omega, \phi)$ dada por (2.4)
2. Se puede ver que

$$f(y, u|\omega, \phi) = \sum_{j=1}^{\infty} \omega_j U(u|0, \omega_j) K(y|\phi_j)$$

3. Así, con probabilidad ω_j , u y y son independientes. Además,

$$f(u|\omega, \phi) = \sum_{j=1}^{\infty} \omega_j U(u|0, \omega_j) = \sum_{j=1}^{\infty} I(u < \omega_j).$$

Por otro lado, si $A(u|\omega)$ es el conjunto definido por $A(u|\omega) = \{j : \omega_j > u\}$, entonces la ecuación (2.5) se puede escribir como

$$f(y, u|\omega, \phi) = \sum_{j \in A(u|\omega)} K(y|\phi_j).$$

Por lo tanto,

$$f(y|u, \omega, \phi) = \frac{f(y, u|\omega, \phi)}{f(u|\omega, \phi)} = \frac{\sum_{j \in A(u|\omega)} K(y|\phi_j)}{\sum_{j=1}^{\infty} I(u < \omega_j)} = \frac{\sum_{j \in A(u|\omega)} K(y|\phi_j)}{f(u|\omega, \phi)}$$

Así, dada la variable u se tiene un modelo de mezcla infinita con pesos todos iguales a $1/f(u|\omega, \phi)$.

Ahora, como es común en el contexto de mezclas, se pueden introducir $\forall i = 1, \dots, n$, variables indicadoras z_i . Así, dada z_i la variable y_i sigue una distribución $K(y|\phi_{z_i})$. Es decir, z_i indica a cuál componente de la mezcla pertenece la observación y_i .

Por lo tanto, consideremos la siguiente densidad conjunta:

$$f(y, z = k, u|\omega, \phi) = K(y|\phi_k)I(k \in A(u|\omega)). \quad (2.6)$$

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

$$f(\{y_i, u_i, z_i = k_i\}_{i=1}^n|\omega, \phi) = \prod_{i=1}^n K(y_i|\phi_{k_i})I(u_i < \omega_{k_i}). \quad (2.7)$$

Densidades condicionales finales completas:

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

$$u_i|\omega, \phi, z = k_i \sim U(0, \omega_{k_i})$$

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

$$\phi_j \sim \pi(\phi_j|\cdot) \propto P_0(\phi_j) \prod_{k_i=j} K(y_i|\phi_j).$$

Si no existiera un k_i igual a j entonces ϕ_j se debe simular de su respectiva distribución inicial $P_0(\phi_j)$.

3. De la verosimilitud (2.7) se tiene

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

Recuérdese 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 $\{\nu\}$ solo las $\nu_s \leq dmax$, donde $dmax = max\{k_1, \dots, kn\}$ serán afectadas, y sólo para esos índices se muestrean de la distribución condicional final completa, las cuales están dadas por:

$$\pi(\nu_j) \propto Beta(a_j, b_j)$$

donde

$$a_j = 1 + \sum_{i=1}^n I(z_i = j)$$

$$b_j = \alpha + \sum_{i=1}^n I(z_i > j)$$

Sin embargo, para los ν_j con $j > dmax$, $\nu_j \sim Beta(1, c)$.

4. De la densidad (2.7) se tiene

$$P(z_i = k | \dots) \propto K(y_i | \phi_k) I(k \in A(u_i | \omega))$$

Los valores que toma cada z_i es sobre un conjunto finito, el cual es de la forma $\{k : \omega > u_i\}$. Por lo que sólo se necesitan muestrear tantos ω_k 's hasta asegurar que se tienen todos los ω_k para toda $i = 1, \dots, n$. Para encontrar todos los ω_k en cada paso del algoritmo, Walker (2007) [33] propone muestrear s^* pesos ω_s hasta que se cumpla que

$$\sum_{s=1}^{s^*} \omega_s > 1 - u^*$$

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

Aunque en este trabajo de tesis se implementa el algoritmo de Kalli (2010) [16], se pospone su descripción en la sección de Inferencia para el modelo propuesto del siguiente capítulo, puesto que su algoritmo retoma el trabajo realizado por Walker en 2007.

Capítulo 3

Inferencia Bayesiana

A continuación se discute la manera de realizar inferencia en los modelos que se propusieron en el Capítulo 2, es decir, para el Modelo Gamma, el Modelo Gamma Proyectado y para el Modelo de Mezclas infinitas. Particularmente, revisando la estructura algebraica de las expresiones que tienen las distribuciones condicionales completas utilizadas en los algoritmos que serán empleados.

3.1. Inferencia para el modelo Gamma

Como se menciona en la Sección 2.1, el objetivo principal es hacer inferencia para los parámetros α y β en el modelo Gamma. Para esto, se quiere encontrar la distribución posterior

$$\begin{aligned}\pi(\alpha, \beta | x_1, \dots, x_n) &= \frac{\prod_{i=1}^n \text{Gamma}(x_i, \alpha, \beta) \cdot \pi(\alpha, \beta)}{\int \prod_{i=1}^n \text{Gamma}(x_i, \alpha, \beta) \cdot \pi(\alpha, \beta) d\alpha d\beta} \\ &\propto \prod_{i=1}^n \text{Gamma}(x_i, \alpha, \beta) \cdot \pi(\alpha, \beta)\end{aligned}$$

donde $\prod_{i=1}^n \text{Gamma}(x_i, \alpha, \beta)$ es la verosimilitud.

Ahora bien, para definir la inicial $\pi(\alpha, \beta)$ se propone tomar $\pi(\beta) = \pi_1(\beta) \sim \text{Gamma}(a, b)$ y una densidad $\pi(\alpha|\beta)$ que sea log-cóncava e independiente a $\pi_1(\beta)$, es decir, $\pi(\alpha|\beta) = \pi(\alpha) = \pi_2(\alpha)$. Así, si $\underline{x}_n = (x_1, \dots, x_n)$, la distribu-

ción posterior a encontrar queda como

$$\begin{aligned}\pi(\alpha, \beta | \underline{x}_n) &\propto \prod_{i=1}^n \text{Gamma}(x_i, \dots, x_n) \pi(\alpha, \beta) \\ &= \prod_{i=1}^n \text{Gamma}(x_i, \dots, x_n) \pi_1(\beta) \pi_2(\alpha),\end{aligned}$$

donde

$$\pi_1(\beta) = \frac{b^a}{\Gamma(a)} \beta^{a-1} e^{-b\beta} \quad ; \quad \beta > 0.$$

Así,

$$\begin{aligned}\pi(\alpha, \beta | \underline{x}_n) &\propto \left[\frac{\beta^{n\alpha}}{(\Gamma(\alpha))^n} \left(\prod_{i=1}^n x_i \right)^{\alpha-1} e^{-\beta \sum_{i=1}^n x_i} \right] \left[\frac{b^a}{\Gamma(\alpha)} \beta^{a-1} e^{-b\beta} \right] \pi_2(\alpha) \\ &\propto \frac{\beta^{a+n\alpha-1}}{(\Gamma(\alpha))^n} e^{-\beta(b+\sum_{i=1}^n x_i)} \left(\prod_{i=1}^n x_i \right)^{\alpha-1} \pi_2(\alpha).\end{aligned}$$

De lo anterior, la distribución final completa para β está dada por

$$\pi(\beta | \alpha, \underline{x}_n) = \frac{\pi(\beta, \alpha | \underline{x}_n)}{\pi(\alpha | \underline{x}_n)}$$

donde $\pi(\alpha | \underline{x}_n)$ esta dada por

$$\begin{aligned}\pi(\alpha | \underline{x}_n) &\propto \pi(\underline{x}_n | \alpha) \pi(\alpha) \\ &= \pi(\alpha) \int \pi(\underline{x}_n, \beta, | \alpha) d\beta \\ &= \pi(\alpha) \int \pi(\underline{x}_n | \beta, \alpha) \pi(\beta | \alpha) d\beta \\ &= \pi_2(\alpha) \int L(\beta, \alpha | \underline{x}_n) \pi_1(\beta) d\beta \\ &= \pi_2(\alpha) \int \left[\frac{\beta^{n\alpha}}{(\Gamma(\alpha))^n} \left(\prod_{i=1}^n x_i \right)^{\alpha-1} e^{-\beta \sum_{i=1}^n x_i} \right] \left[\frac{b^a}{\Gamma(\alpha)} \beta^{a-1} e^{-\beta \sum_{i=1}^n x_i} \right] d\beta \\ &\propto \frac{(\prod_{i=1}^n x_i)^{\alpha-1}}{[\Gamma(\alpha)]^n} \frac{\Gamma(a+n\alpha)}{[b+\sum_{i=1}^n x_i]^{a+n\alpha}} \pi_2(\alpha) \int \frac{[b+\sum_{i=1}^n x_i]^{a+n\alpha}}{\Gamma(a+n\alpha)} \beta^{(a+n\alpha)-1} e^{-\beta(b+\sum_{i=1}^n x_i)} d\beta.\end{aligned}$$

La expresión dentro de la integral es una $\text{Gamma}(a+n\alpha, b+\sum_{i=1}^n x_i)$, ya que la integral corre dentro del soporte de β en $(0, \infty)$, dicha integral vale uno y se tiene que

$$\pi(\alpha | \underline{x}_n) \propto \frac{\Gamma(a+n\alpha)}{[\Gamma(\alpha)]^n} \frac{(\prod_{i=1}^n x_i)^{\alpha-1}}{(b+\sum_{i=1}^n x_i)^{a+n\alpha}} \pi_2(\alpha). \quad (3.1)$$

Por otro lado,

$$\begin{aligned}
\pi(\beta|\underline{x}_n, \alpha) &= \frac{\pi(\alpha, \beta|\underline{x}_n)}{\pi(\alpha|\underline{x}_n)} \\
&= \frac{\frac{\beta^{a+n\alpha-1}}{[\Gamma(\alpha)]^n} e^{-\beta(b+\sum_{i=1}^n x_i)} (\prod_{i=1}^n x_i)^{\alpha-1} \pi_2(\alpha)}{\frac{\Gamma(a+n\alpha)}{[\Gamma(\alpha)]^n} \frac{(\prod_{i=1}^n x_i)^{\alpha-1}}{(b+\sum_{i=1}^n x_i)^{a+n\alpha}} \pi_1(\alpha)} \\
&= \frac{\beta^{a+n\alpha-1}}{\Gamma(a+n\alpha)} (b + \sum_{i=1}^n x_i)^{a+n\alpha} e^{-\beta(b+\sum_{i=1}^n x_i)} \quad (3.2)
\end{aligned}$$

lo cual es una distribución $Gamma(a+n\alpha, b+\sum_{i=1}^n x_i)$.

Una vez que se tienen las densidades condicionales completas $\pi(\beta|\alpha, \underline{x}_n)$ y $\pi(\alpha|\underline{x}_n)$, éstas se pueden emplear con un Gibbs Sampler para obtener muestras de la distribución final conjunta de α y β y así, estar en condiciones de llevar a cabo inferencia para los parámetros α y β .

Se debe mencionar que para simular de la densidad condicional de α (3.1) se empleó un ARS (revisado en la Sección 1.4.2), puesto que $\pi_2(\alpha)$ se tomó como la densidad Gamma (2.1), lo que hace que la densidad condicional de α sea log-cóncava (ver Pradhan y Kundu (2011) [25]). Para mayores detalles sobre implementación del ARS, el lector se puede referir a Gilks y Wild (1992) [13].

3.2. Inferencia para el modelo Gamma Proyectado

Dada una muestra aleatoria $\{\theta_1, \dots, \theta_n\}$ de una distribución Gamma Proyectada $GP_{k-1}(\theta|\alpha, \beta)$, donde $\theta_i = (\theta_{i1}, \theta_{i2}, \dots, \theta_{ik-1})$ para $i = 1, 2, \dots, n$, el objetivo es hacer inferencia sobre los vectores de parámetros $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_k)$ y $\beta = (1, \beta_2, \dots, \beta_k)$. Dada la estructura del modelo, un enfoque se basa en la introducción de variables latentes $R_i, i = 1, \dots, n$, tal que el vector aleatorio $\mathbf{Y}_i \rightarrow (R_i, \Theta_i)$ tiene una distribución Gamma multivariada de dimensión k , como se definió en la ecuación (2.2). De hecho este enfoque es usado por Presnell et al.(1998) [26] y Núñez-Antonio y Gutierrez Peña (2005) [23] en el caso normal multivariado.

Entonces, para los datos completos $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n$ la verosimilitud esta dada por

$$f(\{\mathbf{y}_i\}_{i=1}^n | \boldsymbol{\alpha}, \boldsymbol{\beta}) = \prod_{i=1}^n \left(\prod_{j=1}^k Ga(y_{ij} | \alpha_j \beta_j) \right), \quad \text{donde } \beta_1 = 1.$$

Usando la especificación inicial dada por

$$f(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \prod_{j=1}^k Ga(\alpha_j | c_j, d_j) \prod_{l=2}^k Ga(\beta_l | a_l, b_l)$$

Entonces la distribución posterior esta dada, salvo una constante de proporcionalidad, por

$$\pi(\boldsymbol{\alpha}, \boldsymbol{\beta} | \{\theta_i, r_i\}_{i=1}^n) \propto \prod_{i=1}^n f(\{\mathbf{y}_i\}_{i=1}^n | \boldsymbol{\alpha}, \boldsymbol{\beta}) \prod_{j=1}^k Ga(\alpha_j | c_j, d_j) \prod_{l=1}^k Ga(\beta_l | a_l, b_l) \quad (3.3)$$

dada la estructura de la transformación esférica, si $\mathbf{D}_n = \{(\theta_i, r_i)\}_{i=1}^n$, el resultado son distribuciones condicionales compuestas por partes para los vectores $\boldsymbol{\alpha}, \boldsymbol{\beta}$ que están dadas a continuación.

De (3.3) la distribución final de α está dada por

$$\pi(\alpha_1 | \mathbf{D}_n) \propto \frac{\prod_{i=1}^n (r_i \cos \theta_{i1})^{\alpha_1 - 1}}{\Gamma^n(\alpha_1)} \times Ga(\alpha_1 | c_1, d_1).$$

Así, para $p = 2, \dots, k - 1$,

$$\pi(\alpha_p | \mathbf{D}_n) \propto \frac{\Gamma(C_p)(D_p)^{-C_p}}{\Gamma^n(\alpha_1)} \times Ga(\alpha_p | c_p, d_p) \quad \text{y}$$

$$\pi(\beta_p | \mathbf{D}_n), \boldsymbol{\alpha} = Ga(C_p, D_p),$$

donde

$$\begin{aligned} C_p &= C(p) = n\alpha_p, \text{ y} \\ D_p &= \sum_{i=1}^n (r_i \cos \theta_{ip} \prod_{l=1}^{p-1} \sin \theta_{il}) + b_p, \end{aligned}$$

Finalmente,

$$\begin{aligned} \pi(\alpha_k | \mathbf{D}_n) &\propto \frac{\Gamma(C_k)(D_k)^{-C_k}}{\Gamma^n(\alpha_k)} Ga(\alpha_k | c_k, d_k) \prod_{i=1}^n (r_i \prod_{j=1}^k (\sin \theta_{ij}))^{\alpha_k - 1} \\ \pi(\beta_k | \mathbf{D}_n, \boldsymbol{\alpha}) &= Ga(C_k, D_k), \end{aligned}$$

donde

$$C_k = C(k) = n\alpha_k + a_k, \quad y$$

$$D_k = \sum_{i=1}^k (r_i \prod_{d=1}^k \text{sen}\theta_d) + b_k.$$

Para llevar a cabo inferencias para el modelo Gamma proyectado mediante un muestreo de Gibbs, esencialmente se requiere simular de las densidades condicionales de las variables latentes R_i , lo cual se define a continuación. De $\pi(\boldsymbol{\alpha}, \boldsymbol{\beta} | \{\theta_i, r_i\}_{i=1}^n)$, la densidad condicional posterior para R_i , $\forall i = 1, \dots, n$ está dada por

$$f(r_i | \theta_i, \boldsymbol{\alpha}, \boldsymbol{\beta}) = Ga \left(\sum_{j=1}^k \alpha_j, \cos\theta_1 + \sum_{i=2}^{k-1} \left(\beta_i \cos\theta_i \prod_{j=1}^{i-1} \text{sen}\theta_j \right) + \beta_k \prod_{j=1}^{k-1} \text{sen}\theta_j \right). \quad (3.4)$$

Es importante señalar que el muestreo de las densidades condicionales para las β 's y r_i 's no son difíciles ya que son distribuciones estándar, mientras que el muestreo de las densidades condicionales $\pi(\alpha_j | \cdot)$ no es sencillo. Existen varias opciones para el muestreo de esas densidades condicionales. Como se mencionó anteriormente, en este trabajo el muestreo de esas densidades condicionales se realiza a través de un muestreo ARS (muestreo de Aceptación-Rechazo). Particularmente, se utiliza la función `ars` del paquete ARS de R (R Core Team (2019) [27]) el cual se basa en el algoritmo propuesto por Gilks y Wild (1992) [13].

Usando todas las condicionales completas que se presentaron anteriormente en un muestreo de Gibbs se obtienen realizaciones de la densidad conjunta posterior $\pi(\boldsymbol{\alpha}, \boldsymbol{\beta} | \{\theta_i, r_i\}_{i=1}^n)$. Así, el algoritmo de Gibbs queda especificado de la siguiente manera:

1. Dar un conjunto inicial de valores de $\boldsymbol{\alpha}$ y $\boldsymbol{\beta}$.
2. Para $i = 1, \dots, n$ obtener r_i de $f(r_i | \theta_i, \boldsymbol{\alpha}, \boldsymbol{\beta})$ como en la Ec. (3.4).
3. Obtener α_1 de la densidad $\pi(\alpha_1 | \{\theta_m, r_m\}_{m=1}^n)$.
4. Para $j = 2, \dots, k$, obtener una muestra de α_j y β_j de la densidad conjunta correspondiente a través del siguiente muestreo en bloques:
 - Obtener α_j de la densidad $\pi(\alpha_j | \{\theta, r_m\}_{m=1}^n)$

- Usando el valor de α_j obtenida en el paso anterior, obtener β_j de la densidad condicional $\pi(\beta_j|\{\theta_m, r_m\}_{m=1}^n, \boldsymbol{\alpha})$.

3.2.1. Dos casos especiales importantes

En esta sección se discute sobre el caso de datos circulares (Distribución de la Gamma Proyectada) así como el caso de datos esféricos (Distribución de la Gama Proyectada Bivariada).

El caso circular

Los vectores en un espacio 2-dimensional son vectores en el círculo unitario y por lo tanto cada observación puede ser representado por un ángulo. De esta forma, se dice que un ángulo Θ tiene una distribución Gamma Proyectada si su función de probabilidad esta dada por

$$GP(\theta|\boldsymbol{\alpha}, \boldsymbol{\beta}) = \frac{\beta_2^{\alpha_2} (\cos\theta)^{\alpha_1-1} (\sen\theta)^{\alpha_2-1}}{B(\alpha_1, \alpha_2) (\cos\theta + \beta_2 \sen\theta)^{(\alpha_1+\alpha_2)}} I_{(0, \pi/2)}(\theta)$$

con $\boldsymbol{\alpha} = (\alpha_1, \alpha_2)$ y $\boldsymbol{\beta} = (1, \beta_2)$, donde $B(\cdot, \cdot)$ representa la función beta definida por $B(\alpha_1, \alpha_2) = \frac{\Gamma(\alpha_1+\alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)}$. En este caso la transformación $\mathbf{Y} \rightarrow (R, \Theta)$ esta definida por $\mathbf{Y} = (R \cos\Theta, R \sen\Theta)^t$ y la densidad conjunta de R y Θ está dada por

$$f(r, \theta|\boldsymbol{\alpha}, \beta_2) = \frac{\beta_2^{\alpha_2} (\cos\theta)^{\alpha_1-1} (\sen\theta)^{\alpha_2-1} r^{(\alpha_1+\alpha_2)-1}}{\Gamma(\alpha_1)\Gamma(\alpha_2)} \exp\{-r(\cos\theta + \beta_2 \sen\theta)\}.$$

Entonces, si $\boldsymbol{\theta} = \{\theta_1, \dots, \theta_n\}$ es una muestra de $GP(\theta|\boldsymbol{\alpha}, \beta_2)$ la verosimilitud para los datos completos esta dada por

$$f(\{r_i, \theta_i\}_{i=1}^n|\boldsymbol{\alpha}, \boldsymbol{\beta}) = \frac{\beta_2^{n\alpha_2}}{\Gamma^n(\alpha_1)\Gamma^n(\alpha_2)} \prod_{i=1}^n r_i^n r_i^{(\alpha_1+\alpha_2)-1} (\cos\theta_i)^{\alpha_1-1} (\sen\theta_i)^{\alpha_2-1} \\ \times \exp\left\{-r \left(\sum_{i=1}^n \cos\theta_i + \beta_2 \sum_{i=1}^n \sen\theta_i\right)\right\}.$$

Usando la inicial dada por

$$\pi(\alpha_1, \alpha_2, \beta_2) = Ga(\alpha_1|c_1, d_1)Ga(\alpha_2|c_2, d_2)Ga(\beta_2|a, b),$$

las correspondientes densidades condicionales completas están dadas por

$$\begin{aligned}\pi(\alpha_1|\boldsymbol{\theta}, \mathbf{r}) &\propto \frac{\prod_{i=1}^n (r_i \cos\theta_i)^{\alpha_1-1}}{\Gamma^n(\alpha_1)} Ga(\alpha_1|c_1, d_1) \\ \pi(\alpha_2|\boldsymbol{\theta}, \mathbf{r}) &\propto \frac{\Gamma(n\alpha_2 + a) \prod_{i=1}^n (r_i \sen\theta_i)^{\alpha_2-1}}{\Gamma^n(\alpha_2)(b + \sum_{i=1}^n (r_i \sen\theta_i))^{n\alpha_2+a}} Ga(\alpha_2|c_2, d_2) \\ \pi(\beta_2|\boldsymbol{\theta}, \mathbf{r}, \boldsymbol{\alpha}) &= Ga(n\alpha_2 + a, b + \sum_{i=1}^n r_i \sen\theta_i) \\ \pi(r_i|\boldsymbol{\theta}, \mathbf{r}, \beta_2) &= Ga(\alpha_1 + \alpha_2, \cos\theta_i + \beta_2 \sen\theta_i), \quad i = 1, \dots, n.\end{aligned}$$

Caso esférico

Otro caso importante es el relacionado a los datos esféricos. Los datos esféricos son vectores en un espacio 3-dimensional y pueden ser representados por vectores 2-dimensionales angulares. Esto quiere decir que un vector angular $\boldsymbol{\Theta} = (\Theta_1, \Theta_2)$ tiene una distribución Gama Proyectada Bivariada si su función de probabilidad esta dada por

$$GP_2(\boldsymbol{\theta}|\boldsymbol{\alpha}, \boldsymbol{\beta}) = \frac{\Gamma(\sum_{i=1}^3)\beta_2^{\alpha_2}\beta_3^{\alpha_3}(\cos\theta_1)^{\alpha_1-1}(\sen\theta_1)^{(\alpha_2+\alpha_3)-1}(\sen\theta_2)^{\alpha_3-1}}{\Gamma(\alpha_1)\Gamma(\alpha_2)\Gamma(\alpha_3)(\cos\theta_1 + \beta_2\cos\theta_1\sen\theta_2 + \beta_3\sen\theta_1\sen\theta_2)^{(\alpha_1+\alpha_2+\alpha_3)}} I_{(0,\pi/2)^2}(\boldsymbol{\theta})$$

donde $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \alpha_3)$ y $\boldsymbol{\beta} = (1, \beta_2, \beta_3)$.

Sea $\boldsymbol{\theta} = \{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \dots, \boldsymbol{\theta}_n\}$ una muestra de tamaño n de $GP_2(\boldsymbol{\theta}|\boldsymbol{\alpha}, \boldsymbol{\beta})$ y una inicial dada por

$$\pi(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \prod_{i=1}^3 Ga(\alpha_i|c_i, d_i) \prod_{j=2}^3 Ga(\beta_j|a_j, b_j),$$

entonces, la distribución posterior esta definida por

$$\pi(\boldsymbol{\alpha}, \boldsymbol{\beta}|\{\theta_i, r_i\}_{i=1}^n) \propto \prod_{i=1}^n f(\theta_i, r_i|\boldsymbol{\alpha}\boldsymbol{\beta}) \prod_{i=1}^3 Ga(\alpha_i|c_i, d_i) \prod_{j=2}^3 Ga(\beta_j|a_j, b_j)$$

De lo anterior, las densidades condicionales completas estan dadas por

$$\begin{aligned}
\pi(\alpha_1|\{r_i, \theta_i\}_{i=1}^n) &\propto \frac{\prod_{i=1}^n (r_i \cos \theta_{i1})^{\alpha_1-1}}{\Gamma^n(\alpha_1)} Ga(\alpha_1|c_1, d_1) \\
\pi(\alpha_2|\{r_i, \theta_i\}_{i=1}^n) &\propto \frac{\Gamma(n\alpha_2 + a_2) \prod_{i=1}^n (r_i \cos \theta_{i2} \text{sen} \theta_{i2})^{\alpha_2-1}}{\Gamma^n(\alpha_2) (b_2 + \sum_{i=1}^n (r_i \cos \theta_{i2} \text{sen} \theta_{i2}))^{n\alpha_2+a_2}} Ga(\alpha_2|c_2, d_2) \\
\pi(\alpha_3|\{r_i, \theta_i\}_{i=1}^n) &\propto \frac{\Gamma(n\alpha_3 + a_3) \prod_{i=1}^n (r_i \text{sen} \theta_{i1} \text{sen} \theta_{i2})^{\alpha_3-1}}{\Gamma^n(\alpha_3) (b_3 + \sum_{i=1}^n (r_i \text{sen} \theta_{i1} \text{sen} \theta_{i2}))^{n\alpha_3+a_3}} Ga(\alpha_3|c_3, d_3) \\
\pi(\beta_2|\{r_i, \theta_i\}_{i=1}^n, \boldsymbol{\alpha}) &= Ga((n\alpha_2 + a_2), b_2 + \sum_{i=1}^n (r_i \text{sen} \theta_{i1} \cos \theta_{i2})) \\
\pi(\beta_3|\{r_i, \theta_i\}_{i=1}^n, \boldsymbol{\alpha}) &= Ga((n\alpha_3 + a_3), b_3 + \sum_{i=1}^n (r_i \text{sen} \theta_{i1} \text{sen} \theta_{i2})) \\
\pi(r_i|\{\theta_i\}_{i=1}^n, \boldsymbol{\alpha}, \boldsymbol{\beta}) &= Ga((\alpha_1 + \alpha_2 + \alpha_3), \cos \theta_{i1} + \beta_2 \cos \theta_{i1} \text{sen} \theta_{i2} + \beta_3 \text{sen} \theta_{i1} \text{sen} \theta_{i2}).
\end{aligned}$$

3.3. Inferencia para el Modelo Propuesto

Retomando el modelo de mezclas del capítulo anterior, Kalli et al. (2010) [16] señala que si se considera la densidad conjunta, dada por la ecuación:

$$f(y, z = k, u|\omega, \phi) = K(y|\phi_k)I(k \in A(u|\omega))$$

se presentan varios problemas con el muestreo debidos en parte a la correlación entre las variables u y ω . Adicionalmente señala que la actualización de u puede llevar a simular más pesos ω de los necesarios. Por lo anterior proponen considerar un Slice Sampler más general definiendo la siguiente densidad conjunta:

$$f(y, z = k, u|\omega, \phi) = \xi_z^{-1} I(u < \xi_z) \omega_z K(y|\phi_z)$$

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 $\xi_j = e^{-j}$.

Para propósitos de esta tesis se toma $K(y|\phi_k) = GP_2(\boldsymbol{\theta}|\boldsymbol{\alpha}, \boldsymbol{\beta})$, es decir, una Gamma Proyectada Bivariada. Así, la verosimilitud para los datos completos basada sobre una muestra de tamaño n está dada por:

$$f(\{\boldsymbol{\theta}_i, u_i, z_i = k_i\}_{i=1}^n|\xi, \omega, \boldsymbol{\alpha}, \boldsymbol{\beta}) = \prod_{i=1}^n I(u_i < \xi_{z_i}) \omega_{z_i} / \xi_{z_i} GP_2(\boldsymbol{\theta}|\boldsymbol{\alpha}_{z_i}, \boldsymbol{\beta}_{z_i}) \quad (3.5)$$

Si ξ y ν son condicionalmente independientes entonces las condicionales completas para la implementación de un Gibbs Sampler, puedan ser especificadas de la siguiente manera.

1. $\pi(\phi|\dots) \propto Ga(\alpha_0)Ga(\beta_0) \prod_{d_i=j} GP_2(\theta_i|\alpha_j, \beta_j)$.
2. $\pi(\nu) \propto Be(\nu_j|a_j, b_j)$, donde

$$a_j = 1 + \sum_{i=1}^n I(d_i = j) \quad \text{y} \quad b_k = M + \sum_{i=1}^n I(d_i > j)$$

3. $\pi(u_i|cdots) \propto I(0 < u_i < \xi_{d_i})$
4. $P(d_i = k)|\dots \propto I(k : \xi_k > u_i)\omega_k/\xi_k GP_2(\theta_i|\alpha_j, \beta_j)$

Es obvio que no se pueden muestrear todos los $(\alpha_j, \beta_j, \nu_j)$. Pero no se requiere para proceder con la cadena. Solo se necesita muestrear hasta el 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 's requeridos, ya que será del tipo $\{1, \dots, N\}$ donde $N = \max_i\{N_i\}$ 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. (2010) [16].

Capítulo 4

Aplicaciones

Para ilustrar el potencial que puede tener el Modelo de Mezclas de Gammas Proyectadas en la descripción de datos composicionales, en esta sección se presentan algunos ejemplos. Primero se analizan datos simulados, con la finalidad de evaluar la eficiencia del modelo, es decir, lo que se pretende es observar si se recupera el comportamiento de los datos con características que en un principio son conocidas, para así poder determinar si son correctos o no los resultados obtenidos con la metodología propuesta. Después de realizar esto para varios escenarios, se ejemplifica la aplicación del modelo empleando un conjunto de datos reales. Para este último caso, se eligieron los datos asociados al Conteo Rápido para la Gobernatura del Estado de Morelos en 2018. En este ejemplo real, además de ofrecer la distribución predictiva como una descripción del comportamiento de los datos, se ofrece un estimador puntual y un intervalo de probabilidad para la proporción de tres componentes electorales asociadas al proceso electoral de 2018 en México.

4.1. Datos Simulados

A continuación se presentan dos de las simulaciones que se realizaron para analizar el desempeño de la metodología propuesta en este trabajo. En cada uno de los ejemplos se analizan datos con diferentes comportamientos y se espera que la metodología propuesta obtenga inferencias adecuadas sobre el modelo que generó cada conjunto de datos.

Como primer ejemplo se simuló una muestra de 300 observaciones de un modelo de mezclas de Gammas proyectadas de dimensión 3 con 3 componen-

tes. Para especificar el modelo, se tomaron 3 densidades Gamma proyectadas con vector de pesos $(0.33, 0.33, 0.33)$, los valores de los parámetros que se utilizaron en cada densidad fueron:

Componente			
	1	2	3
α_1	10	90	45
α_2	15	95	42
α_3	18	90	40
β_1	1	1	1
β_2	3	1	5
β_3	1	3	5

Se aplicó la metodología descrita en capítulos anteriores para generar una muestra de la densidad predictiva final de tamaño 1000. Para esto, se corrió el algoritmo con un número de iteraciones de calentamiento de 20,000. Posteriormente, se tomó una observación cada 40 iteraciones. Es decir, se corrió el algoritmo 60,000 iteraciones en total. La decisión de tomar este número de iteraciones es porque como se puede observar en la Fig. 4.1 (a), a las 15,000 iteraciones y un espacio de 20 entre cada observación, se nota que para el monitoreo de la variable de número de grupos, la convergencia tiene aún un ligero descenso, es por ello que se decide tomar un mayor número de ellos. En la Fig. 4.1 (b) el número de iteraciones es de 20,000 con una separación de 40 lo que mejora la convergencia y como se puede observar la correlación es todavía menor.

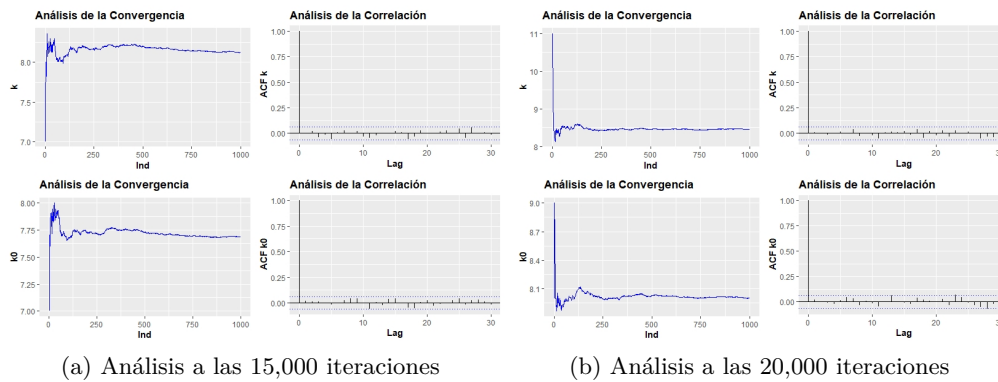


Figura 4.1: Análisis de correlación y convergencia para el primer ejemplo de datos simulados.

Por otro lado, los valores de los hiperparámetros de la distribución ini-

cial para cada parámetro $\alpha_h \sim \text{Gamma}(\alpha_{0h}, \beta_{0h})$, con $h = (1, 2, 3)$ y $\beta_i \sim \text{Gamma}(\alpha_{0i}, \beta_{0i})$, con $i = (2, 3)$ se tomaron como:

α_1	α_2	α_3	β_2	β_3
$\alpha_{10} = 10$	$\alpha_{20} = 10$	$\alpha_{30} = 10$	$\alpha_{20} = 10$	$\alpha_{30} = 10$
$\beta_{10} = 0.1$	$\beta_{20} = 0.1$	$\beta_{30} = 0.1$	$\beta_{20} = 10$	$\beta_{30} = 10$

Se puede ver en la Figura 4.2 (a) la muestra simulada en donde se visualizan tres conglomerados. En la Figura 4.2 (b) se presenta la región de credibilidad de la predictiva en color rojo junto con la correspondiente muestra simulada original. En este caso se puede ver que esta región de credibilidad recupera el comportamiento de los 3 grupos que se simularon. En las Tablas 4.1, 4.2 y 4.3 se muestran los intervalos correspondientes asociados a la población de la cual se obtuvo la muestra, así como a la muestra obtenida de la distribución predictiva final. Los valores se encuentran dados marginalmente para cada componente. Como puede observarse, para cada componente los intervalos contienen los valores verdaderos dado que el método funciona, es decir, los resultados anteriores reflejan que el modelo propuesto ofrece inferencias adecuadas para describir conglomerados de datos composicionales.

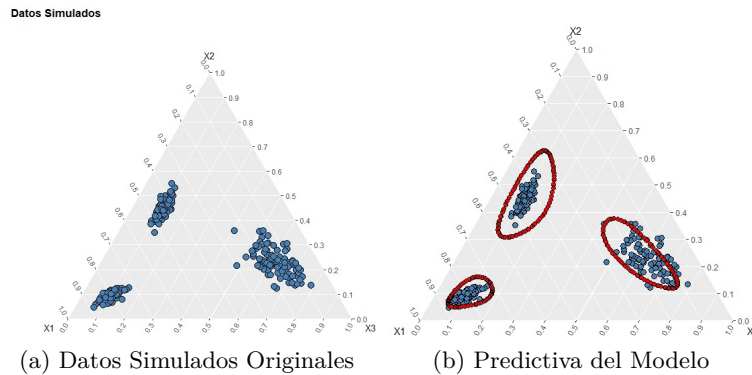


Figura 4.2: Primer ejemplo de Datos Simulados

Componente 1	Intervalos de Probabilidad	
X_1	Inferior	Superior
Poblacional	0.1045390	0.8386196
Distribución Predictiva	0.09338012	0.86045253

Tabla 4.1: Componente 1

Componente 2	Intervalos de Probabilidad	
X_2	Inferior	Superior
Poblacional	0.0721203	0.5077929
Distribución Predictiva	0.06500767	0.57722913

Tabla 4.2: Componente 2

Componente 3	Intervalos de Probabilidad	
X_3	Inferior	Superior
Poblacional	0.07765358	0.69957535
Distribución Predictiva	0.02142546	0.70180667

Tabla 4.3: Componente 3

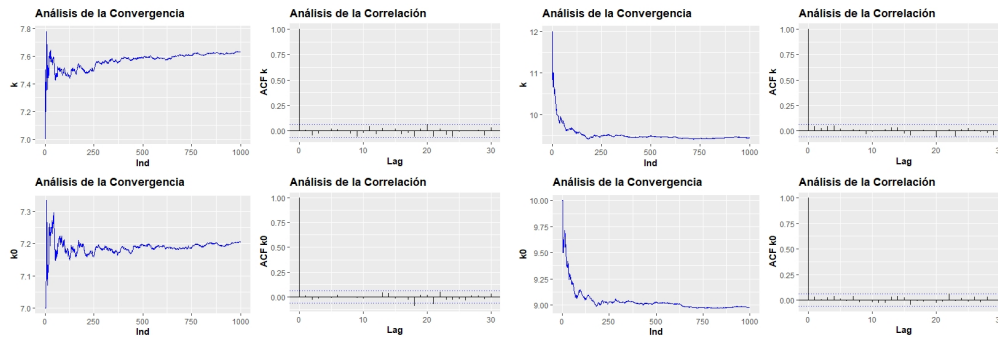
Como segundo ejemplo se analiza el desempeño de la metodología propuesta en un escenario extremo, es decir, donde se tienen datos composicionales con valores muy pequeños. Para este ejemplo, se simula una muestra de 300 observaciones de un modelo de mezcla de Gammas Proyectadas de dimensión 3 con 2 componentes. Se tomaron 2 densidades Gamma Proyectada con vector de pesos $(0.45, 0.55)$ y los parámetros que se encuentran en la Tabla 4.4.

Componente		
	1	2
α_1	5	5
α_2	2	3
α_3	0.2	0.1
β_1	1	1
β_2	1	0.05
β_3	1	0.01

Tabla 4.4: Parámetros para el segundo ejemplo

En este ejemplo se evalúa el alcance que tiene el modelo a través del análisis de la media de los datos poblacionales como parámetro de interés. Para este fin se llevó a cabo una simulación de tamaño 1000, es decir, una simulación de una “población” de 1000 valores, y a éstos se les calculó su media geométrica, la cual se considera como el valor “real” de la media poblacional, la cual resulta ser $(0.9214, 0.0784, 0.0002)$.

En la Figura 4.4 (a) se presenta la muestra simulada. Para generar una muestra de tamaño 1,000 de la distribución predictiva final, se corrió el algoritmo con 20,000 iteraciones como periodo de calentamiento, posteriormente se tomó la muestra de 1,000, una cada 20 iteraciones. Así, el algoritmo completo se corrió 40,000 iteraciones. Al igual que para el primer ejemplo se realizó un análisis gráfico de correlación y convergencia, el cual se puede ver en la Figura 4.3.



(a) Análisis a las 10,000 iteraciones con espacio de 20. (b) Análisis a las 20,000 iteraciones con espacio de 20.

Figura 4.3: Análisis de correlación y convergencia para el segundo ejemplo de datos simulados.

Por otro lado, los valores de la distribución inicial para cada parámetro fueron los mismos que para el primer ejemplo. Es importante notar que para todos los ejemplos simulados se tomó como valor de arranque numérico para el proceso Dirichlet $c = 10$ y como número de grupos $k_{start} = 10$.

Para este caso no solo se evalúa el desempeño del modelo para la predictiva, que resulta muy buena pues se recupera el comportamiento de los datos al igual que en el primer caso, sino que también se procede a calcular la media y mediana de esa predictiva como primeras aproximaciones a un estimador puntual ver Figura 4.4 (b).

Los valores de la media y mediana que se obtuvieron fueron

Valores reales	(0.9214, 0.0784, 0.0002)
Media	(0.9234, 0.0763, 0.0003)
Mediana	(0.8981, 0.0998, 0.0021)

Como puede observarse, el valor de la media es muy cercana a la media real, es decir, con la metodología propuesta, aún en escenarios complicados,

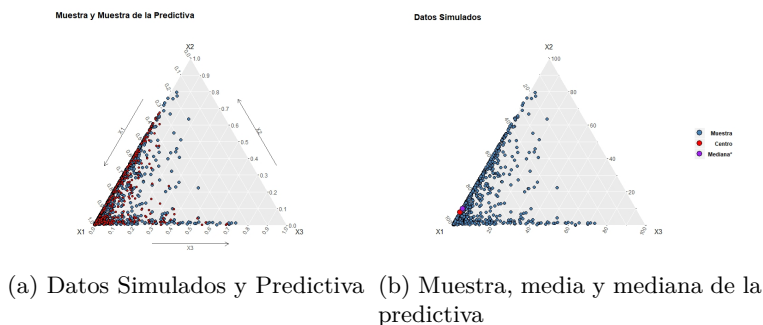


Figura 4.4: Segundo ejemplo de Datos Simulados

se pueden obtener inferencias adecuadas.

4.2. Datos Reales

En México se llevan a cabo procesos electorales de manera regular (gubernamentales, presidenciales, etc.). Sin embargo para algunos de ellos, además del proceso normal de la emisión y conteo de votos, se realiza un procedimiento estadístico que pretende estimar las tendencias de los resultados finales de la elección. Este proceso es llamado Conteo Rápido y se lleva a cabo la noche de la elección, comienza eligiendo, previamente, una muestra aleatoria del universo de casillas aprobadas para la Jornada electoral, para después, por un proceso matemático y estadístico que suele realizarse con diferentes métodos de estimación y aprobado con anterioridad por la autoridad electoral, estimar las tendencias de los resultados.

Las estimaciones que resultan del Conteo Rápido se comunican a la población la misma noche de la jornada electoral por medio de intervalos de confianza, además se incluye la estimación del porcentaje de participación ciudadana en la elección. El Reglamento de Elecciones (Art. 373) establece que la muestra debe ser al 95 % de confianza.

El primer Conteo Rápido se llevó a cabo en 1994 por el entonces Instituto Federal electoral (IFE) para la elección presidencial de ese año. En ese ejercicio, los resultados fueron altamente coincidentes con los del cómputo final (Cómputo Distrital).

En el año 2003, el reto fue mayúsculo ya que se pretendía que el Conteo Rápido determinara la composición de los 500 diputados que conformarían la Cámara. El reto era complejo puesto que la forma en que se tomaría la muestra debería tener en cuenta los 300 distritos electorales. Además, de que para la asignación de los lugares se debían seguir reglas propias de no-sobrerrepresentación. En este caso, el IFE decidió emplear sus propios recursos sin convocar a ninguna empresa y fueron los Capacitadores y Asistentes Electorales quienes se encargaron de transmitir la información de las diferentes casillas de la muestra. De nuevo los resultados coincidieron con los previstos por el Conteo Rápido lo que ayudó a que este ejercicio fuera considerado como un proceso recurrente y confiable.

Por su parte la elección presidencial de 2006 fue una de las más competidas en la historia de México, con un margen de diferencia muy pequeño entre la primera y segunda posición. En este caso se utilizaron tres métodos estadísticos: clásico, bayesiano y robusto. Al corte se había recibido el 95.12% de la muestra total, observando una diferencia menor al 0.5% en los intervalos de confianza para las estimaciones de los dos candidatos punteros. El entonces Consejero Presidente del IFE (Luis Carlos Ugalde), anunció que los resultados del procedimiento del Conteo Rápido no permitían anunciar a un ganador y por esta razón no se darían a conocer las estimaciones de los dos candidatos punteros. Esto, generó un ambiente de incertidumbre y duda sobre la certeza de la elección. A continuación se decidió que el PREP sería el procedimiento que se utilizaría para mostrar los resultados conforme se computaran las casillas. Para el año 2009 se llevaron a cabo elecciones intermedias, aunque en esta ocasión no se realizó un Conteo Rápido. Sin embargo, para la elección presidencial de 2012 fue posible realizarlo, el Consejero General especificó que se darían a conocer los rangos de votación por candidato cualesquiera que fueran las diferencias entre ellos.

En el año 2014 se promulga la nueva Ley General de Instituciones y Procedimientos Electorales (LEGIPE), en donde se definen y describen con rigor los programas de resultados preliminares. En el Reglamento de Elecciones (Art. 357) se establece la obligación de realizar conteos rápidos para gobernadores y jefe de gobierno de la Ciudad de México, bajo este nuevo régimen,

se llevaron a cabo elecciones en 2015 y 2017 donde los problemas de las diferencias horarias, el porcentaje de no respuesta de parte de los CAEs que apoyan al levantamiento de los resultados para la muestra se puntualizaron como problemas a resolver.

Teniendo en cuenta los porcentajes de no respuesta en 2017, el COTECORA, ahora conformado por científicos y estadísticos de diversas universidades, puso una restricción para el tamaño de muestra para las elecciones de 2018. La estrategia fue utilizar tamaños de muestra no tan grandes, principalmente en las elecciones locales. Se buscó que los CAEs no tuvieran sobrecarga y que el margen de error no fuera mayor al 1%. Con esta estrategia el porcentaje de no respuesta se estimó entre el 20% y el 42%. Nuevamente, el Conteo Rápido logró tener una alta coincidencia entre sus resultados y los Cómputos Distritales para el primer lugar. Para el segundo lugar, esta coincidencia fue casi total excepto en Tabasco. Además hubo un traslape entre los intervalos reportados en el Estado de Puebla. A pesar de estos detalles, en general las estimaciones del Conteo Rápido de 2018 reflejaron correctamente las tendencias en todas las votaciones (Gubernamentales, en la CDMX y Presidenciales).

A continuación se muestran las Estimaciones ofrecidas por el COTECORA-2018, del Conteo Rápido para el Estado de Morelos en 2018.

Estimación de los resultados de la votación para la elección de la Gubernatura del Estado de Morelos 2018

Con datos recibidos a las 22:40 hora del centro, del 01 de julio de 2018, el Comité Técnico Asesor informa lo siguiente:

1. De las 200 casillas que integran la muestra, se recibió información de 160 casillas, las cuales representan el 80.0 % de la muestra total.
2. De los 12 estratos considerados para definir el diseño muestral, se contó con información de 12 estratos.

Con la información recibida y con un nivel de confianza de al menos 95 % se estima lo siguiente:

3. La participación ciudadana se encuentra entre 63.7 % y 67.6 %.
4. El porcentaje de votos para cada candidatura a la Gubernatura del estado de Morelos se presenta a continuación:

NOMBRE	PARTIDO / COALICIÓN / INDEPENDIENTE	INTERVALO %	
		LÍMITE INFERIOR	LÍMITE SUPERIOR
Victor Manuel Cabello Solano	PAN_MC	13.4	16.1
Jorge Armando Meade Ocaranza	PRI	5.6	7.0
Rodrigo Gayosso Cepeda	PRD_PSD de Morelos	10.2	12.1
Nadia Luz María Lara Chávez	PVEM	3.3	4.0
Cuauhtémoc Blanco Bravo	PES_PT_MORENA	51.0	53.8
Alejandro Vera Jiménez	NA	2.2	2.7
Mario Rojas Alba	PH de Morelos	1.2	2.0
Fidel Demédecis Hidalgo	INDEPENDIENTE	3.8	4.9

Atentamente
Comité Técnico Asesor del Conteo Rápido

Dr. Alberto Alonso y Coria

Dra. Michelle Anzarut Chacalo

Dr. Carlos Hernández Garciadiego

Dr. Manuel Mendoza Ramírez

Dr. Luis Enrique Nieto Barajas

Dra. Gabriel Núñez Antonio

Dr. Carlos Erwin Rodríguez
Hernández-Vela

Mtra. Patricia Isabel Romero Mares

Dr. Paul Rueda Díaz del Campo

Para efectos de esta tesis, en esta sección se aplica la metodología desarrollada en los capítulos anteriores para el análisis del Conteo Rápido del Estado de Morelos en 2018, siendo de las elecciones más recientes.

Estado de Morelos

Se tomó el Conteo Rápido para el Estado de Morelos 2018, el cual, como puede leerse en la página anterior, se realizó con una muestra de 160 casillas, tomada por el Comité Técnico Asesor del Conteo Rápido.

Después de un tratamiento a la base original de los datos, se colapsaron las variables de interés en datos composicionales de dimensión 3. Aglomerando los partidos/coaliciones/candidatos independientes en 3 grupos, se llevó a cabo el análisis con la metodología propuesta. De esta manera se cuenta con las variables:

- Coalición Morena: Incluye Morena, Partido Encuentro Social, Partido del Trabajo, y sus coaliciones.
- Coalición PAN: Incluye Partido Acción Nacional, Movimiento Ciudadano, y sus coaliciones.
- Otros: Partido Revolucionario Institucional, Partido de la Revolución Democrática, Partido Social Demócrata de Morelos, Partido Verde Ecológico 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.

Con esto se realizó la transformación de los datos a variables direccionales, en específico a datos esféricos, obteniendo vectores bivariados de ángulos, los cuales son el insumo muestral en el algoritmo que realiza la inferencia del modelo.

Se debe señalar que la especificación de la distribución inicial es la misma que en el caso de los escenarios simulados, es decir, $\alpha_h \sim \text{Gamma}(\alpha_{0h}, \beta_{0h})$, con $h = (1, 2, 3)$, $\beta_i \sim \text{Gamma}(\alpha_{0i}, \beta_{0i})$, con $i = (2, 3)$, donde

α_1	α_2	α_3	β_2	β_3
$\alpha_{10} = 10$	$\alpha_{20} = 10$	$\alpha_{30} = 10$	$\alpha_{20} = 10$	$\alpha_{30} = 10$
$\beta_{10} = 0.1$	$\beta_{20} = 0.1$	$\beta_{30} = 0.1$	$\beta_{20} = 10$	$\beta_{30} = 10$

ya que se obtiene una distribución inicial no informativa, lo que quiere decir que el ejercicio se realiza como si no se tuviera la posibilidad de tener más información que la que se puede dar en general para cualquier conjunto de datos. Adicionalmente se utilizó un parámetro para el proceso Dirichlet de $c = 3$. Al terminar las iteraciones del algoritmo, lo que se hace es volver a transformar las realizaciones de la distribución predictiva final obtenida y llevarlas al Simplex para poder analizarlas e interpretarlas adecuadamente.

En la Figura 4.5 (a) se exhibe en color azul la muestra recolectada de las casillas por los CAEs y en la Figura 4.5 (b) en rojo, se observa la muestra de la correspondiente distribución predictiva de tamaño 2000, que se obtuvo después de 120,000 iteraciones del modelo con un periodo de calentamiento de 80,000 y un espacio de 20 entre cada toma de la muestra.

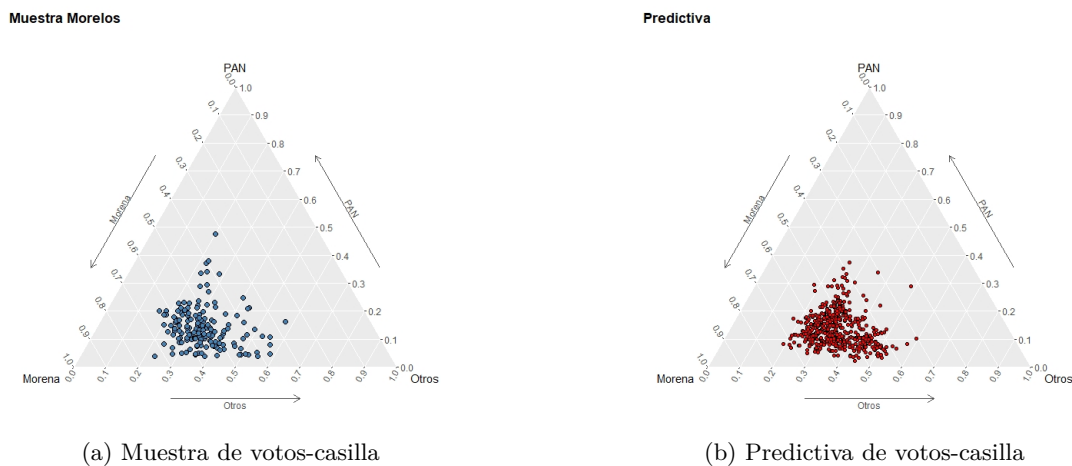


Figura 4.5: Conteo Rápido de Morelos

Es importante resaltar que al igual que en el caso de los datos simulados, en este ejercicio, el modelo propuesto reproduce razonablemente el comportamiento de los datos (ver Fig. 4.5), además con la metodología propuesta se puede ofrecer un estimador puntual que cumple con la restricción de que la

suma de sus componentes sea 1 como se puede observar en la Tabla 4.5.

	Cómputo Distrital	Media	Mediana
Coal. MORENA	0.5259	0.5345	0.5350
Coal.PAN	0.1405	0.1331	0.1312
OTROS	0.3336	0.3324	0.3338

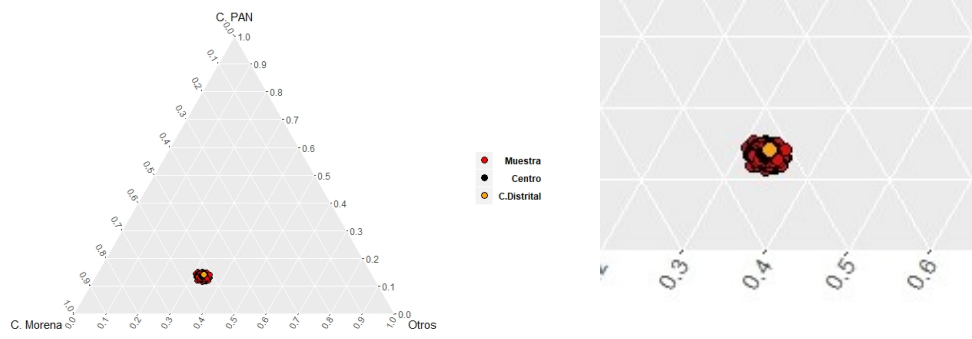
Tabla 4.5: Cómputo Distrital reportado por el INE y estimadores puntuales calculados. La media en este caso es la media geométrica composicional.

Nótese que en la Tabla 4.5, el valor de la media geométrica resulta un mejor estimador puntual que la mediana marginal. La media geométrica se calcula a partir de la muestra de la predictiva final como se describe en la Definición 1.10. Por otro lado, al no tener una definición de mediana en el Simplex, se optó por calcularla marginalmente para cada componente de la muestra de la predictiva final, lo que puede causar que éste no sea un buen estimador.

Para tener una idea de la variabilidad del estimador puntual propuesto, y en consecuencia estar en condiciones de calcular regiones/intervalos de probabilidad, se procedió de la siguiente manera. En cada iteración (en equilibrio) del algoritmo, se generó una muestra aleatoria de 500 realizaciones del modelo (vectores composicionales de dimensión tres) y para esta muestra se calculó la media geométrica. Así, se contaba con 2000 realizaciones del estimador puntual (la media geométrica vectorial). En la Figura 4.6 se muestra este conjunto de medias geométricas. Se puede apreciar, que el valor correspondiente al Cómputo Distrital (punto en color amarillo) se encuentra dentro de la “nube de medias”. Adicionalmente, también se puede construir una región de credibilidad del 95 %, ver Figura 4.7, la cual como se esperaba contiene al valor del Cómputo Distrital asociado a la elección a la Gubernatura del Estado de Morelos en 2018.

En la Tabla 4.6 se muestran los intervalos de credibilidad para la proporción de votos de cada una de las 3 variables. Se puede observar que comparando estos intervalos con el cómputo distrital que es (0.5259, 0.1405, 0.3336), estos lo contiene. Los resultados obtenidos nos permiten concluir que la metodología propuesta en este trabajo de tesis es capaz de obtener inferencias

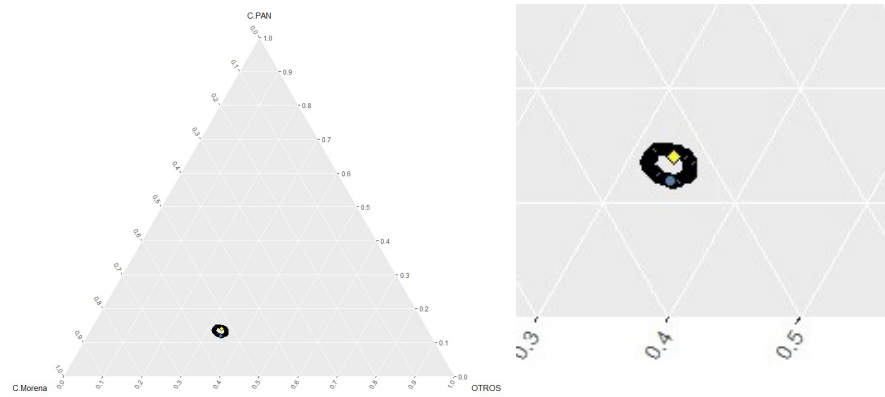
Muestra de Medias para el Conteo Rápido de Morelos



(a) Variabilidad del estimador puntual

(b) Zoom

Figura 4.6: Conteo Rápido de Morelos 2018



(a) Región de Credibilidad para la proporción de votos

(b) Zoom

Figura 4.7: Conteo Rápido de Morelos 2018

	Intervalo al 95 %
C. MORENA	(0.5199, 0.5457)
C. PAN	(0.1225, 0.1446)
Otros	(0.3188, 0.3491)

Tabla 4.6: Intervalos de Probabilidad al 95 % para cada Coalición

adecuadas en problemas reales como los Conteos Rápidos organizados por el Instituto Nacional Electoral.

Dado que este modelo se probó con algunos otros estados de la República Mexicana, para ampliar el ejercicio a continuación presentamos solo los resultados encontrados para dos estados más. El procedimiento fue muy similar al de Morelos cambiando la muestra, para el estado de Yucatán el COTECORA contó con 188 casillas y para la Ciudad de México con 647. Es importante señalar que los valores para las iniciales son los mismos que para Morelos. Mientras que los valores de arranque también fueron iguales a excepción del periodo de calentamiento que fue de 30,000 para la Ciudad de México y de 80,000 para Yucatán.

	Cómputo Distrital	Media	Mediana
Coal. PAN	0.3955	0.3999	0.3981
Coal. PRI	0.3616	0.3608	0.3526
Otros	0.2429	0.2393	0.2493

Tabla 4.7: Cómputo distrital reportado por el INE y estimadores puntuales calculados para el estado de Yucatán. La media es la media geométrica composicional.

	Intervalo al 95 %
Coal. PAN	(0.3762, 0.4120)
Coal. PRI	(0.3445, 0.3816)
Otros	(0.2228, 0.2642)

Tabla 4.8: Intervalo de credibilidad al 95 % para el estado de Yucatán.

Como se puede observar en las Tablas 4.7 y 4.9, tanto la media geométrica composicional como la mediana para los dos estados son similares al cómputo distrital. Por otro lado en las tablas 4.8 y 4.10, los intervalos de credibilidad contienen al cómputo distrital para ambos estados al igual que con el ejemplo descrito de Morelos.

	Cómputo Distrital	Media	Mediana
Coal. MORENA	0.4705	0.4707	0.4797
Coal. PAN	0.3102	0.3043	0.2994
Otros	0.2193	0.2250	0.2209

Tabla 4.9: Cómputo distrital reportado por el INE y estimadores puntuales calculados para la Ciudad de México. La media es la media geométrica composicional.

	Intervalo al 95 %
Coal. MORENA	(0.4551, 0.4782)
Coal. PAN	(0.2980, 0.3230)
Otros	(0.2127, 0.2340)

Tabla 4.10: Intervalo de credibilidad al 95 % para la Ciudad de México.

Como una forma de complementar este trabajo de tesis, presentamos algunos ejercicios adicionales realizados con diferentes tamaños de muestra para el estado de Morelos para observar los efectos que se tienen al aplicar este modelo.

Se tomaron las muestra recibidas la noche de la elección de Morelos a las 22:10 y a las 22:30 horas, con 129 y 153 casillas, respectivamente. Ambas muestras cuentan con la información de los 12 distritos locales al igual que la muestra final de 160. Llevando a cabo la misma metodología ya aplicada para la muestra que tomó el COTECORA, se procedió a introducir cada muestra en el modelo, lo que nos arrojó como resultado los siguientes intervalos de credibilidad al 95 %.

	Intervalo al 95 %
Coal. MORENA	(0.5199, 0.5507)
Coal. PAN	(0.1254, 0.1486)
Otros	(0.3133, 0.3410)

Tabla 4.11: Intervalo de credibilidad al 95 % para la muestra de 129 casillas del estado de Morelos.

	Intervalo al 95 %
Coal. MORENA	(0.5183, 0.5430)
Coal. PAN	(0.1231, 0.1457)
Otros	(0.3195, 0.3485)

Tabla 4.12: Intervalo de credibilidad al 95 % para la muestra de 153 casillas del estado de Morelos.

Como puede observarse en las Tablas 4.11 y 4.12, los intervalos de credibilidad para las componentes en general tienen mayor distancia para la muestra de 129 y de 153 que para la final de 160 como era de esperarse, sin embargo, las diferencias no sobrepasan los 3,1 puntos porcentuales.

Se debe señalar que el modelo propuesto tiene varias aplicaciones, particularmente, se puede emplear para datos como los de los Conteos Rápidos por su naturaleza estadística, sin embargo como ya se mencionó, los datos composicionales pueden encontrarse en diversos fenómenos reales.

Conclusiones y Perspectivas

En este trabajo de tesis se propone una metodología Bayesiana para datos composicionales que toma en cuenta la naturaleza topológica de los datos, es decir, nunca se pierde de vista que los datos viven en el simplex. Esto es muy importante por la correlación negativa entre las variables, lo cual cobra relevancia para el estudio de este tipo de datos, ya que cualquier perturbación en alguna de las componentes genera cambios directamente en las demás. Así, es de suma importancia notar que este aspecto fundamental se respeta en todo momento en la metodología que se propone. Además se está en condiciones de ofrecer, no sólo una muestra de la distribución predictiva final, sino también un estimador puntual que cumple con la restricción de que sus partes suman uno. Esto no ocurre si se aplican otros métodos estadísticos como los que se utilizan actualmente en, por ejemplo, los ejercicios estadísticos de Conteos Rápidos.

De manera justa, se debe señalar que puede parecer poca cosa dar un estimador puntual que cumpla con la restricción de suma constante, pero dar respuesta estadística y formal al análisis de datos como los del ejercicio del Conteo Rápido, es un gran logro metodológico.

Una vez que se mapean los datos composicionales, el modelo de mezclas basado en proyecciones radiales de distribuciones Gamma multivariadas y definido sobre la esfera unitaria, permite analizar los datos composicionales sin tener que usar transformaciones como la *ilr* o la *clr*. Adicionalmente, resulta muy flexible y permite describir conjuntos de datos con comportamientos grupales, asimétricos y/o simétricos. Además, inicialmente permite analizar datos composicionales con valores en sus componentes muy cercanas a cero.

A pesar de las bondades de la metodología propuesta en este trabajo, pa-

ra analizar datos composicionales, hay que señalar que para su desarrollo se tuvieron que revisar una gran variedad de temas como el enfoque bayesiano de la estadística, la modelación de datos direccionales (definidos en la esfera unitaria), la inferencia en modelos multivariados, métodos de simulación directa y estocástica, el análisis y comprensión de modelos de mezclas infinitas de distribuciones. Este último tema, sigue representando un tema complejo desde un enfoque de inferencia bayesiano. Lo anterior ya que, aunque existen algoritmos para implementarlo, se siguen haciendo propuestas metodológicas para mejorar los tiempos de convergencia. Además los modelos Bayesianos requieren siempre de un gran trabajo computacional, que también fue desarrollado en esta tesis. Gracias a todo esto el modelo propuesto es robusto y nada sencillo, ya que se modelan datos composicionales que a su vez pueden mostrar una estructura de agrupamiento.

Aunque quedan cosas por hacer, consideramos que la propuesta desarrollada en esta tesis sienta las bases para el análisis Bayesiano de datos composicionales.

Referencias

- [1] Aithison, John. (2003). *A Concise Guide to Compositional Data Analysis*. Honorary Senior Research Fellow. Department of Statistics University of Glasgow. United Kingdom.
- [2] Aitchison, John. (1982). *The statistical of compositional data (with discussion)*. Journal of the Royal Statistical Society, Vol. 44. pp 139-177.
- [3] Aldrich, J. (1995). *Correlations Genuine and Spurious in Person y Yule*. Statistical Science. Vol. 10, pp. 364-367.
- [4] Bernardo, J.M. (1981). *Bioestadística: Una perspectiva Bayesiana*. Disponible en <http://www.uv.es/bernardo/Bioestadistica.pdf>
- [5] Bernardo, J.M., Smith, A.F.M. (2000) *Bayesian Theory*. Wiley. Nueva York.
- [6] Billheimer, D., Guttorp, P., Fagan, W.F. (2001). *Statistical interpretation of species composition*. Journal of the American Statistical Association 83 (401), pp. 85-91.
- [7] Box, G., Tiao G. (1973) *Bayesian Inference in Statistical Analysis*. Wiley Classics Library.
- [8] Erdely, A., Gutiérrez, E. (2007). *Monografía de Estadística Bayesiana*. Recuperado el 21 de mayo de 2019, de <http://drive.google.com/file/d/0B2bCelAllmzINWRIdmIJQ2JXdjQ/view>
- [9] Escobar, M.D., West, M. (1995). *Bayesian Density Estimation and Inference Using Mixtures*. Journal of the American Statistical Association. Vol. 30 (430), pp. 577-588.
- [10] Ferguson, T. (1973). *A Bayesian Analysis of some Nonparametric Problems*. The Annals of Statistics. Vol. 1 No.2, pp. 209-230.

- [11] Gamerman, D., Freitas Lopes, H. (2006) *Markov Chain Monte Carlo*. (2nd ed.) Florida, USA. Ed. Chapman & Hall/CRC. Taylor & Francis Group.
- [12] Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2004). *Bayesian data analysis* (2nd ed.). Boca Raton: Chapman and Hall/CRC.
- [13] Gilks, W.R., Wild, P.(1992). *Adaptative Rejection Sampling for Gibbs Sampling*. Journal of the Royal Statistical Society. Series C (Applied Statistics). Vol 41. No.2, pp. 337-348.
- [14] Gilks, W.R., Richardson, S., Spiegelhalter, D.J., (1996). *Markov Chain Monte Carlo in Practice*. London, United Kingdom. Ed. Springer-Science+Businesss Media, B.V.
- [15] Hoff, P. (2009) *A First Course in Bayesian Statistical Methods*. Springer Texts in Statistics.
- [16] Kalli, M., Griffin, J.E., Walker, S.G. (2010). *Slice sampling mixture models*. Centre for Health Services Studies. pp. 23.
- [17] Lo, A.Y. (1984). *On a class of Bayesian nonparametric estimates: I. Density Estimates*. The Annals of Statistics. Vol. 12 (1). pp. 351-357.
- [18] Mardia, K.V.(1976). *Discussion on de ordering of multivariate data*. Journal of the Royal Statistical Society, pages 346-347.
- [19] Mardia, K.V., Jupp, P.E. (2000) *Directional Statistics*. Chischester: Wiley.
- [20] Mateu-Figueras, G., Martín-Fernández, J.A., Pawlosky-Glan, V., y Barceló-Vidal, C. (2003, abril). *El problema del análisis estadístico de Datos Composicionales* 27 Congreso Nacional de Estadística e Investigación Operativa. Departamento de Informática y Matemática Aplicada. Universidad de Girona. España.
- [21] Nuñez-Antonio, G., Concepción-Ausín M., Wiper M.P. (2015). *Bayesian nonparametric models of circular variables based on dirichlet process mixtures of normal distributions*. Journal of Agricultural Biological and Environmental Statistics,20(1),pp.47-64.

- [22] Nuñez-Antonio, G., Geneyro, E.,(2019). *A multivariate Projected Gamma Model for Directional Data*.
- [23] Nuñez-Antonio, G., Gutierrez-Peña, E. (2005) *A Bayesian analysis of directional data using the projected normal distribution*. Journal of Applied Statistics, Vol.32 (10), pp. 995-1001.
- [24] Pawlosky-Glahn, V., Egozcue, J. J., Tolosana-Delgado, R. (2015) *Modeling and Analysis of Compositional Data* . United Kingdom. John Wiley & Sons.
- [25] Pradhan, B., Kundu, D.(2011). *Bayes estimation and prediction of the two-parameter gamma distribution*. Journal of Statistical Computation and Simulation. Vol. 81, No. 9., pp. 1187-1198.
- [26] Presnell, B., Morrison, S., Littell, R.C. (1998) *Projected Multivariate Linear Models for Directional Data* Journal of the American Statistical Association, Vol. 93 (443), pp. 1068-1077.
- [27] R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.
- [28] Robert, C., Casella, G. (2004). *Monte Carlo Statistical Methods* (2nd ed.) New York, USA. Ed. Springer.
- [29] Scely J.L., Welsh A.H.(2011). *Regression for compositional data by using distribution defined on the hypersphere*. Journal of the Royal Statistical Society, pp. 197-203.
- [30] Sethuraman, J., Tiwari R.C., (1982) *Convergence of Dirichlet Measures and the Interpretation of Their Parameter*. Statistical Decision Theory and Related Topics, III. Academic Press.
- [31] Tolsana-Delgado,R., von Eynatten, H.(2010) *Simplifying compositional multiple regression: Application to gran size controls on sediment geometry*. Computers and Geosciences. Vol. 36, pp. 577-589.
- [32] Wackerly, D., Mendenhall III, W., Scheaffer, R. (2008). *Estadística matemática con aplicaciones*. (7ma. ed.) Traducción, CDMX, México. Ed. Cengage Learning.

- [33] Walker, S.G. (2007). *Sampling the Dirichlet Mixture Model with Slices*. Communication in statistics - Simulation and Computation. Vol. 36, No. 1, pp. 45-54.
- [34] Wang F., Gelfand A.E. (2007). *Directional data analysis under the general projected normal distribution*. Stat Methodol, pp. 113-117.

Apéndice A

Códigos en R

A.1. Funciones para la Inferencia de los parámetros de una Gamma univariada

```
#Librerias
library(ars)
##Funcion para los parametros gamma##
#Funcion que obtiene una muestra para alfa y beta de una densidad Gamma
(univariada) a partir de n observaciones (univariadas)
fungamma<- function(y,alpha01,beta01,alpha02,beta02, iters=4){
  #y es la muestra de n observaciones (univariadas) de una densidad gamma
  (univariada)
  r<-length(y)
  suma<-sum(y)
  #Funcion log-posterior para alfa
  f <- function(alfa , alpha01 , beta01 , r , y ,alpha02 , beta02){
    lgamma(alpha01 + r*alfa) - r*lgamma(alfa) + alfa*sum(log(y)) -
    (r*alfa)*log(beta01 + sum(y)) + (alpha02-1)*log(alfa) - (beta02*alfa)
  }
  #Derivada de la log-posterior de alfa
  fprima <- function(alfa , alpha01 , beta01 , y , r ,alpha02 , beta02){
    r*digamma(alpha01 + r*alfa) - r*digamma(alfa) + sum(log(y)) -
    r*log(beta01 + sum(y)) + ((alpha01-1)/(alfa)) - beta02
  }
  for (i in 1:iters+1) {
    p1<-ars(1,f,fprima,x=c(1,10),m=2,lb=TRUE,xlb=0.000000001,
            ub=TRUE,xub=90000000, alpha01=alpha01, beta01=beta01, y=y,
            r=r, alpha02=alpha02, beta02=beta02)
    p2<-rgamma(1,shape=(r*p1+alpha01),rate=(suma+beta01))
  }
  param<-c(p1,p2);
  return(param)
}
fungamma<- compiler::cmpfun(fungamma) #La funcion compiler ayuda con
#el tiempo de compilacion de funciones en R.

#Funcion que obtiene una muestra para alfa y beta fijo (beta=1) de una
#densidad Gamma (univariada) a partir de n observaciones (univariadas)
fungamma2<- function(y,beta02, iters=4){
  #y es la muestra de n observaciones (univariadas) de una densidad gamma (univariada)
  longitud<-length(y)
  suma<-sum(y)
  g<-function(x) # el parametro x se refiere a alfa como variable
  {
    (x-1)*sum(log(y)) -x*beta02 -longitud*lgamma(x)
  }
  gprima<-function(x)
```

```

{
  sum(log(y)) -beta02 -longitud*digamma(x)
}

for (i in 1:iters+1) {
  p1<-ars(1,g,gprima,x=c(1,10),m=2,lb=TRUE,xlb=0.0000000001,
        ub=TRUE,xub=90000000);
}
return(p1)
}
fungamma2 <- compiler::cmpfun(fungamma2)

```

A.2. Función para la Inferencia del Modelo Propuesto

```

infinite_mixture_GP3<- function(sizeMCMC, burn, freqS, y_star, y, alpha01,
beta01, alpha02, beta02, alpha03, beta03, alpha04, beta04, alpha05, beta05, cc,
k_start){

# -----
# -----
# Creacion de variables
# -----
# No de iteraciones totales
N<-burn+sizeMCMC*freqS

n <- length(y[,1])
# Almacenamiento
#mu <- array(NA, c(N, n))
alphap1<-rep(NA,n)
alphap2<-rep(NA,n)
alphap3<-rep(NA,n)
betap2 <- rep(NA,n)
betap3<-rep(NA,n)
w <- rep(NA,n)
r<-rep(1,n)
v <- rep(NA, n)
k <- rep(NA, N)
k0 <- rep(NA, N)
# Estadísticas básicas
nj <- rep(0, n)
#sj <- rep(0, n)
#s2j <- rep(0, n)
# Variables latentes
z <- rep(NA, n)
u <- rep(NA, n)
ki <- rep(NA, n)

# Codigo Gabriel: Mayo 25, 2019
# Contador para la predictiva final
mez.cont<-0
#f.pred<-matrix(0,sizeMCMC,length(y_star[,1]))

y.pred<-matrix(0,sizeMCMC,ncol(y_star))
medmed<-matrix(0,sizeMCMC,6)
# vectores de salida
k.sal<-rep(0,sizeMCMC)
k0.sal<-rep(0,sizeMCMC)
#r.sal<-rep(1,sizeMCMC)

#x1.sal<-rep(0,n)
#x2.sal<-rep(0,n)
#x3.sal<-rep(0,n)
#p<-rep(0,n)

```

```

# -----
# -----
# Inicializacion
# -----
#
# Asignaciones en la iteracion t=1
t <- 1
k[t] <- k_start
k0[t] <- k_start

z<-kmeans(y,k_start)$cluster
#z<-floor(runif(n)*k[t]) + 1
u<-exp(-z)*runif(n)
ki<-as.integer(floor(-log(u)))
for(i in 1:n){
  # Se pueden sacar del for, y estn justo en las lnes previas.
  # z[i] <- floor(runif(1)*k[t]) + 1
  # u[i] <- exp(-z[i])*runif(1)
  # ki[i] <- floor(-log(u[i]))
  nj[z[i]] <- nj[z[i]] + 1
  # sj[z[i]] <- sj[z[i]] + y[i]
  #s2j[z[i]] <- s2j[z[i]] + y[i]*y[i]
}

# -----
# Asignacion de todos los componentes j=1,...,K[t]

j <- 1

#if(0){ #Codigo de Carlos E
# mu[t, j] <- rnorm(1, mu0, 1/sqrt(tau0))
#tau[t, j] <- rgamma(1, alpha0, beta0)
#v[j] <- rbeta(1, 1, cc)
#w[t, j] <- v[j]
#for(j in 2:k[t]){
# mu[t, j] <- rnorm(1, mu0, 1/sqrt(tau0))
#tau[t, j] <- rgamma(1, alpha0, beta0)
# v[j] <- rbeta(1, 1, cc)
# w[t, j] <- w[t, j-1]*(v[j]/v[j-1])*(1-v[j-1])
#} }

#Codigo de Gabriel N
alphap1 <- rgamma(n, alpha03, beta03)
alphap2 <- rgamma(n, alpha04, beta04)
alphap3 <- rgamma(n, alpha05, beta05)
betap2 <- rgamma(n, alpha01, beta01)
betap3 <- rgamma(n, alpha02, beta02)
v <- rbeta(n, 1, cc)
w[j] <- v[j]
for(j in 2:k[t]){
  # Se pueden sacar del for, y estn justo en las lnes previas.
  # mu[t,j] <- rnorm(n, mu0, 1/sqrt(tau0))
  # tau[t,j] <- rgamma(n, alpha0, beta0)
  # v[j] <- rbeta(n, 1, cc)
  w[j] <- w[j-1]*(v[j]/v[j-1])*(1-v[j-1])
}

k[t+1] <- max(ki)
k0[t+1] <- sum(tabulate(ki, k[t+1]) > 0)

# Termina la asignacion en todas las componentes j=1,...,K[t]
# en la iteracion t=1
# -----
# -----
# Empieza el LOOP GRANDE (subsecuentes iteraciones)
#
for(t in 2:N){

```

```

xx1<-r*cos(y[,1])
xx2<-r*sin(y[,1])*cos(y[,2])
xx3<-r*sin(y[,1])*sin(y[,2])
tmp <- 0
for(j in 1:k[t]){
  if(nj[j] > 0){
    y1temp<-xx1[which(z==j)]
    y2temp<-xx2[which(z==j)]
    y3temp<-xx3[which(z==j)]
    alphap1[j]<-fungamma2(y1temp,beta03,5)
    parametros2<-fungamma(y2temp,alpha01,beta01,alpha04,beta04,5)
    alphap2[j]<-parametros2[1]
    betap2[j]<-parametros2[2]
    parametros3<-fungamma(y3temp,alpha02,beta02,alpha05,beta05,5)
    alphap3[j]<-parametros3[1]
    betap3[j]<-parametros3[2]
    tmp <- tmp + nj[j]
    v[j] <- rbeta(1, nj[j] + 1, n - tmp + cc)
  }else{
    alphap1[j] <- rgamma(1,alpha03,beta03)
    alphap2[j] <- rgamma(1,alpha04,beta04)
    alphap3[j] <- rgamma(1,alpha05,beta05)
    betap2[j] <- rgamma(1,alpha01,beta01)
    betap3[j] <- rgamma(1,alpha02,beta02)
    v[j] <- rbeta(1, 1, cc)
  }
  ifelse(j==1,w[j]<-v[j],w[j]<-w[j-1]*(v[j]/v[j-1])*(1-v[j-1]))
}

for(l in 1:n) {#Actualiza r para cada individuo de acuerdo a la componente
temp.r<-rgamma(1,alphap1[z[l]]+alphap2[z[l]]+alphap3[z[l]],
cos(y[l,1]+betap2[z[l]]*sin(y[l,1])*cos(y[l,2])*
+betap3[z[l]]*sin(y[l,1])*sin(y[l,2]))
if(length(temp.r)==0){
  r[l]<-1 #En caso de que no se pueda calcular el valor de R
}
else{
  r[l]<-temp.r
}
}

# Hago cero las nj, sj y s2j
nj <- nj*0
#sj <- sj*0
#s2j <- s2j*0
for(i in 1:n){
  p <- rep(0, ki[i])
  for(j in 1:ki[i]){
    p[j] <- exp(j)*w[j]*dgamma(xx1[i], shape=alphap1[j],
rate=1)*dgamma(xx2[i], shape=alphap2[j],
rate=betap2[j])*dgamma(xx3[i], shape=alphap3[j],
rate=betap3[j])
  }
  p <- replace(p, c(which(is.na(p)), which(is.infinite(p)),
which(is.nan(p))), 0)
if((length(p) == 1) && (p == 0)){p <- 1}
z[i] <- sample.int(ki[i], 1, prob = p/sum(p))
u[i] <- exp(-z[i])*runif(1)
ki[i] <- as.integer(floor(-log(u[i])))
# Actualizo nj y sj
nj[z[i]] <- nj[z[i]] + 1
#sj[z[i]] <- sj[z[i]] + y[i]
#s2j[z[i]] <- s2j[z[i]] + y[i]*y[i]
#p <- p*0
}

```

```

k[t+1] <- max(ki)
k0[t+1]<- sum(tabulate(ki, k[t+1]) > 0)

# Calculo de la predictiva y de los vectores de salida
if(t>burn && (t%FreqS)==0){
  #Codigo Gabriel: Mayo 25, 2019
  mez.cont<-mez.cont+1
  k.sal[mez.cont]<-k[t]
  k0.sal[mez.cont]<-k0[t]
  # Calculo de la predictiva
  # Para asegurar que la mezcla sea una densidad e integre 1
  #vw<-as.vector(w[1:k[t]]/sum(w[1:k[t]],na.rm=TRUE))
  ##Funcion para pesos
  ww<-as.vector(w[1:k[t]]/sum(w[1:k[t]],na.rm=TRUE))
  indices.pos<-function(uni){
    which(uni<cumsum(ww[1:k[t]]))[1]
  }

  ind.pred<-sapply(runif(500),indices.pos)
  m.modelo<-function(ind){
    if(is.na(ind)){
      alphap51 <- rgamma(1,alpha03,beta03)
      alphap52 <- rgamma(1,alpha04,beta04)
      alphap53 <- rgamma(1,alpha05,beta05)
      betap52 <- rgamma(1,alpha01,beta01)
      betap53 <- rgamma(1,alpha02,beta02)
      return(simPG(1, alphap51,alphap52,alphap53, betap52,betap53))
    }else{
      return(simPG(1, alphap1[ind],alphap2[ind],
        alphap3[ind],betap2[ind],betap3[ind]))
    }
  }
  y.mod<-sapply(ind.pred,m.modelo)
  y.mod<-array(c(y.mod[1, ], y.mod[2, ]),dim = c(500,2))
  y.mod.s3<-dirtocomp(y.mod)
  medmed[mez.cont, ]<-c(mean(acomp(y.mod.s3)),medianareal(y.mod.s3))
}
#if(t==N){
# x1.sal<-xx1
#x2.sal<-xx2
#x3.sal<-xx3
#p.sal<-p
#}
} # termina el loop grande

list(y.pred=y.pred, k=k.sal, k0=k0.sal,medmed=medmed, y.mod=y.mod)
}

infinite_mixture_GP3 <- compiler::cmpfun(infinite_mixture_GP3)

```

A.3. Funciones Auxiliares para la Inferencia del Modelo Propuesto

```

#Densidad de la Gamma Proyectada Bivariada
#x: vector de angulos de dimension 2
#los parametros son parejas (a,b) correspondientes a los alfas(a) y betas(b)
#de cada una de las 3 Gammas en R
densidadPG<-function(x, a1, b1, a2, b2, a3, b3){
  suma<-sum(a1, a2, a3)
  d<-((lgamma(suma)+(a2*log(b2))+(a3*log(b3))+(a1-1)*log(cos(x[,1]))+
    (a2-1)*log(cos(x[,2]))+((a2+a3)-1)*(sin(x[,1]))+(a3-1)*(sin(x[,2])))-
    (a1+a2+a3)*log(cos(x[,1])+b2*cos(x[,1])*sin(x[,2])+b3*sin(x[,1])*
    sin(x[,2]))) - (lgamma(a1)+lgamma(a2)+lgamma(a3)))
  exp(d)
}

```

```

densidadPG <- compiler::cmpfun(densidadPG)

##Siguiendo el ejemplo de Emiliano Geneyro para la generacion de muestras
de Gamma Proyectadas GP(alfa1 , alfa2 , beta2 , alfa3 , beta3)
simPG<-function(n.dat , alp1 , alp2 , alp3 , be2 , be3)
{
  y<-matrix(NA,n.dat,3)
  angulo<-matrix(0,n.dat,2)
  #—— Simulacion de Componentes ——
  for(i in 1:n.dat)
  {
    y[i,1]<- rgamma(1,shape = alp1 , rate = 1)
    y[i,2]<- rgamma(1,shape = alp2 , rate = be2)
    y[i,3]<- rgamma(1,shape = alp3 , rate = be3)

    angulo[i,1]<-atan(sqrt(y[i,2]^2+y[i,3]^2)/y[i,1]) %%2*pi)
    angulo[i,2]<-atan(y[i,3]/y[i,2]) %%2*pi)
  }
  angulo
}
simPG <- compiler::cmpfun(simPG)

##Funcion para transformar una muestra de datos composicionales a datos
direccionales
comptodir<-function(y){
  n<-length(y[,1])
  angulo<-matrix(0,n,2)
  for (i in 1:n) {
    angulo[i,1]<-atan(sqrt(y[i,2]^2+y[i,3]^2)/y[i,1]) %%2*pi)
    angulo[i,2]<-atan(y[i,3]/y[i,2]) %%2*pi)
  }
  angulo
}

##Funcion que transforma los datos direccionales a composicionales
dirtocomp<-function(x){
  ##Conversion a datos circulares
  theta1<-x[,1]
  theta2<-x[,2]

  y<-cbind(cos(theta1) , sin(theta1)*cos(theta2) , sin(theta1)*sin(theta2))

  ##Conversion a Datos composicionales
  v<-clo(y)
  v
}
dirtocomp<- compiler::cmpfun(dirtocomp)

##Funcion que calcula la mediana como en los reales a partir de una
#muestra de datos composicionales
medianareal<-function(x){
  y<-ilr(x)
  med1<-median(y[,1])
  med2<-median(y[,2])
  medianare<-c(med1,med2)
  medianas3<-ilrInv(medianare)
  medianas3
}
medianareal<-compiler::cmpfun(medianareal)

```