PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE
Facultad de Matemáticas
Departamento de Estadística

ANÁLISIS BAYESIANO NO PARAMÉTRICO UTILIZANDO PROCESOS SKEW DIRICHLET

YASNA ELIZABETH ORELLANA ZAPATA

Tesis presentada al Departamento de Estadística de la Facultad de Matemáticas como requisito para obtener el Grado de Doctor en Estadística

Profesor Guía:
Dra. Pilar Iglesias Zuazola — Pontificia Universidad Católica de Chile.

Comité Científico:
Dr. Reinaldo Arellano — Pontificia Universidad Católica de Chile.
Dr. Fernando Quintana Q. — Pontificia Universidad Católica de Chile.
Dr. Manuel Galea — Universidad de Valparaíso.

Santiago de Chile, Marzo 2007

PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE

Facultad de Matemáticas

Departamento de Estadística

"Analisis Bayesiano No paramétrico utilizando Procesos Skew Dirichlet "

Tesis profesional presentada por

Yasna Elizabeth Orellana Zapata

como requisito para

obtener el grado de Doctor

en Estadística

Santiago de Chile. Verano 2007

PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE

Facultad de Matemáticas

Departamento de Estadística

"Analisis Bayesiano No paramétrico utilizando Procesos Skew Dirichlet "

Tesis profesional presentada por

Yasna Elizabeth Orellana Zapata

como requisito para
obtener el grado de Doctor
en Estadística

Santiago de Chile. Enero 2007

Agradecimientos

Quisiera agradecer a cada una de las personas que me han apoyado en mis estudios durante el transcurso de mi vida. En primer lugar a mis profesoras de la niñez, Srta. Elba Alfaro y Trinidad Baeza, su muestras de cariño para con una niña que gustaba del estudio, hizo nacer en mí el espíritu de búsqueda y de conocimiento. A mi profesora de enseñanza media Sra Mónica Hayashi, quien siempre me alento a seguir profundizando en lo que ya sabía y a continuar aprendiendo cada vez más. A mis profesores de licenciatura en la USACH, especialmente a Sergio Plaza y Rafael Labarca porque siempre confiaron en mis habilidades y me impulsaron a seguir estudiando. Con un sentimiento especial quiero agradecer al profesor, Renato Levin, quien me apoyó y me impulsó a replantear mis estudios de postgrado.Gracias a él pude conocer una nueva ciencia, "la estadística ".

No puedo dejar de mencionar a todos los excelentes profesores con los cuales me he formado, Reinaldo Arellano, Fernando Quintana, Guido del Pino, Ernesto San Martín, Guillermo Marshall, Victor Salinas.

En especial quiero dedicar unas palabras a mi tutora Pilar Iglesias Zuazola, quien desde sus primeras cátedras, me cautivo por su espíritu alegre y positivo, lo que daba un toque de humanidad a las, muchas veces, largas jornadas de aprendizaje dentro de la facultad. *"Pilar, fue un privilegio haberte conocido, fuiste y en mi corazón siempre serás una gran maestra y amiga"*

También quiero mencionar a todos mis compañeros, de distintas generaciones, Porfirio, Mauricio , Giovanna, Natalia, Paula, Mallen, Karla, con quienes tuve gratas conversaciones de sobremesa en muchos almuerzos. También quiero nombrar en forma especial

a Mauricio Castro y Carlos Navarrete quienes compartieron conmigo sus conocimientos computacionales en forma amable y desinteresada. Y como no referirme a mis compañeros, pilares en mi paso por el postgrado, Claudio Beltrán y Mario Tagle por nuestras largas jornadas de estudio, de discusión, de viajes a congresos y especialmente pues forman ya parte de mis confidentes y amigos.

Finalmente quiero agradecer a Conicyt por haberme beneficiado con una beca de Postgrado durante los años 2001 y 2004.

A mi marido, Cristian y a mis hijos,

Alvaro y Elizabeth

Índice general

1. Introducción 1

1.1. Estructura general de la tesis 4

1.2. Preliminares 7

1.3. Medidas de probabilidad sobre un espacio finito dimensional 7

1.3.1. Esquema para un modelo Beta-Binomial 7

1.3.2. Esquema para un modelo Dirichlet-Multinomial 9

1.4. Modelo de Urna asociado a una distribución de Dirichlet 12

2. Procesos de Dirichlet 14

2.1. Procesos de Dirichlet (DP) 14

2.1.1. Medidas de Probabilidad Aleatorias 15

2.1.2. Medidas de probabilidad sobre $\mathcal{M}(\mathcal{X})$ 16

2.1.3. Procesos de Dirichlet 17

2.1.4. Esquema de Urna De Polya 20

2.2. Algunos procesos construidos a partir de los procesos de Dirichlet . . . 22

2.2.1. Procesos de Dirichlet Invariantes 24

2.2.2. Otras medidas de probabilidad aleatorias de interés 25

3. Procesos Skew Dirichlet 27

3.1. Definiciones y algunas propiedades 28

3.1.1. Algunas propiedades de los procesos Skew Dirichlet 29
3.1.2. Otra representación de los procesos Skew Dirichlet 30
3.1.3. Representación de SDP en esquema de urna 32

4. Regresión Bayesiana No paramétrica con SDP **39**
4.1. Introducción 39
4.1.1. Análisis de Regresión utilizando mezclas de SDP 42
4.1.2. Aplicación : Retorno de acciones del mercado chileno 46
4.1.3. Algunos resultados obtenidos en la modelación 48

5. Estimación de una densidad **52**
5.1. Introducción 52
5.1.1. Modelo de mezclas de normales con DP 54
5.1.2. Estimación de una densidad utilizando SDP 55

6. Notas Finales **59**
6.1. Conclusiones 59
6.2. Trabajo Futuro 61

A. Modelo de Regresión **62**
A.1. Algoritmo de estimación en el Modelo de Regresión SDP 62
A.1.1. Distribuciones Condicionales de los parámetros 64

Índice de Tablas

4.1. Algunos resúmenes de las posterioris para el ejemplo de regresión. . . . 51

Índice de figuras

4.1. *Gráfico de dispersión retorno de acciones de Concha y Toro versus el IPSA.* 47
4.2. *Distribuciones a posteriori de los parámetros del modelo* 48
4.3. *HPD's de 95 % para los residuos de la regresión.* 49
4.4. *Distribución a posteriori para algunos residuos en el modelo.* 50

5.1. *Histograma y Densidad Estimada* 57
5.2. *densidad a posteriori condicional* $p(\theta \mid y_1, \ldots, y_n, \theta \neq 1/2)$. 57

Resumen

Los procesos de Dirichlet representan una herramienta natural para definir medidas de probabilidad aleatorias cuyas distribuciones actúan como prioris no paramétricas para la inferencia bayesiana.

En este trabajo se introduce una nueva clase de medidas de probabilidad aleatorias que se derivan como una mezcla convexa de dos procesos de Dirichlet (DP) con soportes ortogonales. Los pesos asignados a cada una de las componentes de esta mezcla están determinados por un único parámetro, este parámetro tiene la particularidad que al tomar un valor específico se obtiene un proceso de Dirichlet simétrico que coincide con el proceso obtenido por Dalal (1979). Es decir en este nuevo proceso se incorpora un parámetro de asimetría donde un proceso simétrico es obtenido como un caso particular a partir de la construcción general.

Como es sabido los DP tiene varias caracterizaciones, una de ellas debida a Sethuraman (1994), el cual describe este proceso como una suma infinita de masas puntuales mezcladas y ponderadas por pesos aleatorios ordenados. Otra caracterización es debida a Blackwell y MacQueen (1973), que obtienen el proceso de Dirichlet como proceso límite de una sucesión de variables aleatorias que conforman una secuencia de Polya. En este trabajo ambas caracterizaciones son imitadas, aplicadas y adecuadas a los procesos Skew Dirichlet lo que se traduce en un manejo computacional del proceso heredado de los procesos de Dirichlet, en el sentido de que ellos pueden ser adecuados agregando un

paso para el parámetro de asimetría, dentro de la implementación computacional.

Finalmente se utiliza el SDP en un modelo de regresión, con datos reales. En este modelo se utiliza una mezcla del proceso en un Kernel normal en el espíritu de los MDP. Se compara la simetría y la asimetría del proceso según un factor de Bayes, el que se puede obtener en el output de las simulaciones. Luego se utiliza el proceso en el problema de estimación de una densidad , donde se obtiene que el modelo captura las características principales de los datos como son asimetría y multimodalidad.

Capítulo 1

Introducción

Motivados por las estructuras que aparecen al analizar datos provenientes de distintos contextos, tales como, medio ambiente, finanzas, medicina entre otros y junto con el avance de la computación estadística, diversos autores han centrado su atención y esfuerzos en construir familias paramétricas que contengan una amplia gama de distribuciones. Por ejemplo que contemplen colas más o menos pesadas que la distribución normal, que sean asimétricas o multimodales.

Las primeras extensiones utilizadas son aquellas que son simétricas, pero no son normales por ejemplo las distribuciones t-student, exponencial potencia, logística etc... El libro *Symmetric Multivariate and Related distributions* de Fang y otros (1990) contiene una completa revisión de estas distribuciones y los trabajos de Osiewalski y Steel (1993), Geweke y Terui (1993) , Arellano-Valle y otros (2000), Arellano-Valle y otros (2003), entre otros, tratan con el análisis bayesiano de los mismos en distintos contextos.

A posterior una extensión más general como es la construcción de distribuciones asimétricas que extienden a la distribución normal fue presentada en el trabajo de Azzalini (1986) a partir del cual comenzaron a desarrollarse una serie de maneras de introdu-

cir asimetría como en los trabajos de Fernández y Steel (1998), Mudholkar y Hutson (2000), Arellano-Valle y otros (2005) entre otros.

El libro de Genton (2004), *Skew Elliptical distributions and their applications*, contiene una revisión de las diferentes familias paramétricas que han sido propuestas en la literatura, conocidas bajo el nombre *Skew distributions* y en el trabajo de Arellano-Valle y Genton (2005) se discute las propiedades que satisfacen las distintas familias y los vínculos existentes entre ellas.

La idea es construir familias más flexibles, sin aumentar demasiado la complejidad, ya sea por el número de parámetros, la interpretación o la implementación computacional, esto ha llevado a adoptar el enfoque bayesiano no paramétrico o semiparamétrico, es decir considerar una distribución sobre el espacio de distribuciones. Esta visión ha tenido un creciente desarrollo en los últimos años, de hecho la medida de probabilidad aleatoria más utilizada como priori en este enfoque es el proceso de Dirichlet (DP) presentado por Ferguson (1973) y que tiene la propiedad de ser cerrado bajo muestras aleatorias (es decir la posteriori es también un DP).

Los DP han sido caracterizados de diversas maneras; a través de la representación en esquema de urna de Polya debida a Blackwell y MacQueen (1973) y como una mezcla infinita de masas puntuales ponderadas, con ponderaciones estocásticas ordenadas debida a Sethuraman (1994). Estos procesos tienen dos características muy importantes como son su gran soporte y el hecho de que la inferencia a posteriori sea analíticamente manejable.

Ahora bien, el hecho de que los procesos de Dirichlet sean casi seguramente discretos los tornan inapropiados para algunas aplicaciones. Para mitigar este problema, Antoniak (1974), extiende los procesos de Dirichlet obteniendo una medida aleatoria, mezclando

procesos de Dirichlet con una distribución mezcla para un parámetro que determina la distribución desde donde provienen las observaciones. Si la distribución mezcla es continua se obtienen una medida de probabilidad aleatoria continua.

Otra forma de obtener medidas aleatorias continuas son los modelos de mezcla que consisten en mezclar un kernel continuo utilizando una medida de mezcla aleatoria a la cual se le da como priori un procesos de Dirichlet. Escobar y West (1995a) introducen este tipo de mezcla utilizando un kernel Normal. Trabajos como los de Dey y Sinha (1998), Ramgopal y otros (1994), Damien y Walker (2002) y Müller y Quintana (2004) presentan una revisión, implementación computacional y aplicaciones utilizando estos modelos de mezcla.

Es importante destacar que los DP concentran su masa sobre distribuciones asimétricas y que excluye del soporte a las familias de distribuciones simétricas.

En este sentido es importante el trabajo debido a Dalal (1979) quien introduce los procesos de Dirichlet Invariantes, éstas medidas de probabilidad aleatoria dan probabilidad uno a las distribuciones simétricas y a estos procesos, se les llama en la literatura, procesos de Dirichlet Simetrizados.

En esta línea Doss (1984), considera un modelo donde el parámetro infinito dimensional tiene distribución simétrica alrededor de cero y como caso particular utiliza los procesos de Dirichlet simetrizados como priori. Brunner (1992), Brunner (1995a) trabaja con la familia entera de distribuciones simétricas unimodales, usando el hecho que éstas admiten una representación como mezclas en el parámetro de escala de distribuciones uniformes simétricas (Feller (1971)) y considera un DP para la distribución mezcla.

Según la revisión previa, vemos que estos trabajos presentan construcciones que

contienen clases de distribuciones simétricas o asimétricas, excluyentemente. Recientemente Kottas y Gelfand (2001) construyen familias de distribuciones con mediana cero, que contiene tanto a las distribuciones unimodales simétricas como las asimétricas, su construcción se basa en la mezcla de distribuciones uniformes.

Es en este aspecto en donde este trabajo presenta una solución, ya que en él se construye una clase de medidas de probabilidad aleatoria en cuyo soporte se encuentran tanto las distribuciones simétricas como asimétricas, unimodales y multimodales, determinadas por una distribución basal , un parámetro de precisión y un parametro de asimetría que permite obtener simetría como caso especial. Llamaremos a esta nueva clase de medidas aleatorias, Procesos Skew Dirichlet (SDP). Veremos que ellos preservan muchas propiedades que posee el proceso de Dirichlet, como la representación en esquema de urna de Polya y como una mezcla infinita de masas puntuales con pesos estocásticamente ordenados.

Esta última caracterización implica que los SDP concentran su masa en distribuciones discretas, sin embargo al igual que los DP es denso en el espacio de todas las distribuciones con el soporte de la distribución basal.

1.1. Estructura general de la tesis

Este trabajo está estructurado en 5 capítulos, los que siguen un orden natural, de tal manera que el lector parta desde los conceptos básicos de medidas de probabilidad aleatorias sobre espacios finito dimensionales, siguiendo con la construcción en espacio infinito dimensionales, específicamente la construcción de procesos de Dirichlet, para luego definir y caracterizar los nuevos procesos llamados Skew Dirichlet y para finalizar con aplicación de éstos en regresión y estimación de densidades.

Algunos cálculos de distribuciones predictivas y posteriores necesarias para la implementación computacional están dadas en el apéndice del trabajo.

En el primer capítulo de esta tesis se presenta una completa introducción la que contempla la motivación , los primeros estudios en la literatura paramétrica y los estudios recientes en literatura no paramétrica para la construcción de familias de distribuciones en espacios de distribuciones. Poniendo especial énfasis en los procesos de Dirichlet y procesos construidos a partir de éste.

A continuación se presenta un conjunto de preliminares, cuyo objetivo es dar en principio una visión básica de medidas de probabilidad aleatorias en espacio finito dimensionales de tal manera que aparezca de forma natural la extesión de aquellas a espacios infinito dimensionales.

En el capítulo 2, se presenta la definición de una medida de probabilidad sobre un espacio de medidas de probabilidad, también se presenta la definición, caracterizaciones y algunas propiedades de los procesos de Dirichlet. Como por ejemplo aspectos de conjugación esquema de urna de Polya y la representación debida a Sethuraman.

En el capítulo 3, se define una nueva familia de medidas de probabilidad aleatorias cuyos elementos se construyen como una combinación convexa de un proceso de Dirichlet concentrado sobre los reales positivos y otro concentrado sobre los reales negativos y cuya combinación depende de un parámetro que controla la asimetría del proceso. A este nuevo proceso se le llamará Proceso Skew Dirichlet (SDP), se da su representación como mezclas de masa puntuales en el espíritu de la representación de Sethuraman, así como el esquema de urna en el espíritu del trabajo de Blackwell & Mac-Queen. Esta secuencia será llamada Skew Polya. Dentro del capítulo se incluyen diversas proposiciones

con sus respectivas demostraciones y se realiza una revisión de las distintas propiedades del SDP.

En el capítulo 4 se aplica el nuevo proceso para modelar un problema de regresión lineal, se incluyen en este capítulo proposiciones respecto a la distribución del error con sus respectivas demostraciones , también para una priori particular del parámetro de asimetría se obtiene el factor de Bayes que permite evaluar la posible simetría o asimetría del modelo. Finalmente se aplica el modelo a datos reales provenientes de una empresa vinera chilena.

En el capítulo 5, se considera el problema de estimación de una densidad donde se realiza la modelación de datos reales utilizando los SDP´s, observando como el modelo captura la multimodalidad y la asimetría de los datos.

En el capítulo final se discuten los principales resultados, conclusiones, posibles extensiones.

En el apéndice se entregan los cálculos de las distribuciones condicionales que se utilizan en el algoritmo de estimación, para la aplicación de los SDP en Modelos de Regresión.

1.2. Preliminares

En este capítulo se mostrarán los modelos probabilísticos asociados a variables aleatorias con dominio $\mathcal{X}$ finito, cuyo conocimiento básico nos permitirán comprender la construcción de medidas probabilidad en un espacio infinito dimensional. En este contexto partiremos por considerar un modelo bayesiano con una variable Bernoulli con probabilidad de éxito p y como priori para p una distribución Beta y a posterior en un caso más general consideramos un vector aleatorio con distribución multinomial a cuyos parámetros les daremos como distribución a priori una distribución de Dirichlet. Es de especial interés darse cuenta que ambas distribuciones a priori pueden ser entendidas como medidas de probabilidad definidas sobre un espacio finito dimensional.

1.3. Medidas de probabilidad sobre un espacio finito dimensional

1.3.1. Esquema para un modelo Beta-Binomial

Consideremos una variable aleatoria la cual puede tomar dos valores en el dominio $\mathcal{X} = \{1, 2\}$ con probabilidad p_1 y $p_2 = (1 - p_1)$ respectivamente. Lo anteriormente expuesto lo podemos escribir a través del modelo

$$X|p_1, p_2 \sim Bernoulli(p_1)$$

Observemos que para cada valor de (p_1, p_2) tal que $p_i \geq 0, i = 1, 2$ y $p_1 + p_2 = 1$ se define una medida de probabilidad sobre $\mathcal{X}$. Si además tenemos desconocimiento acerca de los parámetros p_1, p_2, podemos entonces poner una priori sobre p_1. Una distribución ampliamente utilizada para un parámetro p_1 , $0 \leq p_1 \leq 1$ es la distribución Beta cuya función de densidada está dada por:

$$\pi(p_1) = \frac{\Gamma(\alpha_1 + \alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} p_1^{\alpha_1 - 1}(1 - p_1)^{\alpha_2 - 1}$$

Con $\alpha_1, \alpha_2 > 0$ los parámetros de la distribución Beta.

Es interesante observar que cada α_i está asociado con cada elemento del dominio $\mathcal{X}$, entonces al considerar la siguiente notación $\alpha(x_i) = \alpha_i, i = 1, 2$ podemos entender a α como una medida sobre $\mathcal{X}$, donde $\alpha(\mathcal{X}) = \sum_{x_j \in \mathcal{X}} \alpha_j$

A partir de los parámetros de la distribución a priori Beta , podemos obtener el valor esperado de p_i como

$$E(p_i) = \frac{\alpha_i}{\alpha(\mathcal{X})}$$

También podemos obtener la distribución a posteriori de p_i dado las observaciones $X_1, X_2, ...X_n$ que está dada por una distribución beta. En efecto

$$\begin{aligned} \pi(p_1|X_1, X_2, ..., X_n) &\propto L(p_1|X_1, X_2, ..., X_n)\pi(p_1) \\ &\propto Beta(\alpha_1 + \sum_{i=1}^{n} \delta_{X_i}(1), \alpha_2 + \sum_{i=1}^{n} \delta_{X_i}(2)) \end{aligned}$$

Donde $\delta_{X_i}(x)$ es la función indicadora la cual es uno si $X_i = x$ y cero en otro caso.

La probabilidad conjunta de $X_1, X_2, ..., X_n$ la podemos obtener al marginalizar sobre p_1. Una expresión para esta probabilidad conjunta está dada por:

$$\begin{aligned} p(X_1, X_2, .., X_n) &= \frac{\Gamma(\alpha(\mathcal{X}))}{\Gamma(\alpha_1)\Gamma(\alpha_2)} \int p_1^{\alpha_1 + \sum_{i=1}^{n} \delta_{X_i}(1)} (1 - p_1)^{\alpha_2 + \sum_{i=1}^{n} \delta_{X_i}(2)} \\ &= \frac{\alpha_1^{\sum_{i=1}^{n} \delta_{X_i}(1)} \alpha_2^{\sum_{i=1}^{n} \delta_{X_i}(2)}}{(\alpha(\mathcal{X}))^{[n]}} \end{aligned}$$

donde $\alpha^{[n]}$ representa el factorial $\alpha(\alpha + 1).....(\alpha + n - 1)$

Finalmente podemos obtener la distribución condicional de una nueva observación X_{n+1} dado las observaciones previas $X_1, X_2, .., X_n$, como:

$$p(X_{n+1}|X_1, X_2, ..., X_n) = \frac{p(X_1, X_2, ..., X_n, X_{n+1}|\alpha_1, \alpha_2)}{p(X_1, ..., X_n|\alpha_1, \alpha_2)}$$
$$= \frac{\Gamma(\alpha(\mathcal{X}) + n)\Gamma(\alpha_1 + \sum_{i=1}^{n+1} \delta_{X_i}(1))\Gamma(\alpha_2 + \sum_{i=1}^{n+1} \delta_{X_i}(2))}{\Gamma(\alpha(\mathcal{X}) + n + 1)\Gamma(\alpha_1 + \sum_{i=1}^{n} \delta_{X_i}(1))\Gamma(\alpha_2 + \sum_{i=1}^{n} \delta_{X_i}(2))}$$

Esta última expresión la podemos escribir como:

$$p(X_{n+1}|X_1, X_2, ..., X_n) = \begin{cases} \dfrac{\alpha_1 + \sum_{i=1}^{n} \delta_{X_i}(1)}{\alpha_1 + \alpha_2 + n}, & X_{n+1} = 1; \\ \\ \dfrac{\alpha_2 + \sum_{i=1}^{n} \delta_{X_i}(2)}{\alpha_1 + \alpha_2 + n}, & X_{n+1} = 2. \end{cases}$$

1.3.2. Esquema para un modelo Dirichlet-Multinomial

La distribución de Dirichlet es una generalización multiparamétrica de la distribuición Beta, así como el modelo Multinomial es una generalización multiparamétrica del modelo Binomial.

Siguiendo el mismo esquema que en el caso Beta-Bernoulli, consideremos las variables aleatorias $X_1, X_2, ..., X_n$ las cuales pueden tomar k valores en el dominio $\mathcal{X} =$

$\{1, 2, 3, ..., k\}$ con probabilidades $p_i,\ i = 1, 2, ..., k$ para cada uno de los respectivos valores en $\mathcal{X}$.

Lo anteriormente expuesto lo podemos escribir a través del modelo Multinomial

$$(X_1, X_2, ..., X_n)|p_1, p_2, ..., p_k \sim Multinomial(p_1, p_2, ..., p_k)$$

donde $p_i > 0\ \ i = 1, 2, ..., k$ y $\sum_{j=1}^{k} p_j = 1$ (es decir el vector $(p_1, p_2, ..., p_k)$ pertenece a un Simplex de dimensión $(k-1)$)

Si asumimos desconocimiento sobre los parámetros $p_1, p_2, ..., p_k$ podemos dar como priori la distribución de Dirichlet como una extensión al caso Beta, la que está definida como:

$$\pi(p_1, p_2, .., p_k) = \frac{\Gamma(\sum_{j=1}^{k} \alpha_j)}{\prod_{j=1}^{k} \Gamma(\alpha_j)} \prod_{j=1}^{k} {p_j}^{\alpha_j - 1}$$

De forma similar a lo expuesto para la distribución Beta, podemos derivar distintas propiedades para el esquema Dirichlet-Multinomial, las que podemos resumir como sigue:

1) El valor esperado de p_i es $\frac{\alpha_i}{\sum_{j=1}^{k} \alpha_j}$

2) La distribución a posteriori de p_i dado una observación $(X_1, X_2, ..., X_n)$ de la distribución multinomial es también Dirichlet. En efecto

$$\pi(p_1, p_2, ..., p_k|X_1, X_2, ..., X_n) \propto \prod_{j=1}^{k} {p_j}^{\alpha_j + \sum_{i=1}^{n} \delta_{X_i}(j)}$$

Es decir

$$p_1, p_2, ..., p_k | X_1, X_2, ..., X_n \sim Dirichlet(\alpha_1 + \sum_{i=1}^{n} \delta_{X_i}(1), ..., \alpha_k + \sum_{i=1}^{n} \delta_{X_i}(k))$$

3) La distribución marginal de $X_1, X_2, ..., X_n$ puede ser expresada por

$$p(X_1, X_2, ..., X_n) = \frac{\Gamma(\alpha(\mathcal{X}))}{\prod_{j=1}^{k} \Gamma(\alpha_j)} \int \prod_{j=1}^{k} p_i^{\alpha_j + \sum_{i=1}^{n} \delta_{X_i}(j)} dp_1 .....dp_{k-1}$$

$$= \frac{\alpha_1^{\sum_{i=1}^{n} \delta_{X_i}(1)} ..... \alpha_k^{\sum_{i=1}^{n} \delta_{X_i}(k)}}{(\alpha(\mathcal{X}))^{[n]}}$$

4) La distribución condicional de una nueva observación X_{n+1} dadas las observaciones previas $X_1, X_2, ..., X_n$ está dada por la expresión:

$$p(X_{n+1} | X_1, X_2, ...X_n) = \frac{\Gamma(\alpha(\mathcal{X}) + n)}{\Gamma(\alpha(\mathcal{X}) + n + 1)} \prod_{j=1}^{k} \frac{\Gamma(\alpha_j + \sum_{i=1}^{n+1} \delta_{X_i}(j))}{\Gamma(\alpha_j + \sum_{i=1}^{n} \delta_{X_i}(j))}$$

Así podemos escribir la expresión:

$$p(X_{n+1} = j | X_1, X_2, ...X_n) = \frac{\alpha_j + \sum_{i=1}^{n} \delta_{X_i}(j)}{\alpha(\mathcal{X}) + n}$$

5) Es una propiedad conocida (Teorema de Consistencia de Kolmogorov) que si $B_1, B_2, ..., B_m$ es cualquier partición del dominio $\mathcal{X}$ entonces

$$(P(X \in B_1), ..., P(X \in B_m)) \sim Dirichlet(\alpha(B_1), ..., \alpha(B_m))$$

donde $\alpha(B_i) = \sum_{x_j \in B_i} \alpha_j$

Además de todas las propiedades de la distribución Dirichlet antes expuestas, tenemos que así como otras distribuciones en espacio finitos dimensionales, existe un modelo de urna asociado a dicha distribución, lo que será de gran importancia para la continuidad de este trabajo al extender los resultados a espacios infinito dimensionales. Esto es presentado brevemente en la siguiente sección.

1.4. Modelo de Urna asociado a una distribución de Dirichlet

Diferentes distribuciones de probabilidad pueden ser obtenidas a través de modelos de urna. Una revisión de algunas de ellas se encuentran en Johnson y Kotz (1977). Un caso particular en el que estamos interesados según el contexto del análisis a realizar está dado por el esquema de urna para la distribución Dirichlet, cuya representación es como sigue:

Consideremos una caja con α bolitas, de las cuales inicialmente hay αm_j del color j , $1 \leq j \leq n$. Asumiendo que $m_j\alpha$ es entero. Seleccionamos una bolita al azar desde la caja y en cada selección reemplazamos la bolita seleccionada por ella misma más una bolita del mismo color. Entonces si deseamos calcular la probabilidad de obtener una bolita de color j en la i-ésima selección, que podemos anotar como $P(X_i = j)$ lo podríamos determinar a partir de :

$$P(X_1 = j) = \frac{\alpha m_j}{\sum_{i=1}^{n} \alpha m_i}$$

$$P(X_{n+1} = j | X_1, X_2, ..., X_n) = \frac{\alpha m_j + \sum_{i=1}^{n} \delta_{X_i}(j)}{\alpha + n}$$

Podemos darnos cuenta que la expresión anterior coincide exactamente con la propiedad 4) de la distribución de Dirichlet, en el caso de un dominio $\mathcal{X}$ finito, es decir muestrear desde una distribución de Dirichlet es equivalente a generar una secuencia de variables aleatorias según el esquema de urna de Polya, especificado anteriormente. Blackwell y MacQueen (1973) muestran que este resultado se generaliza al caso en que el dominio $\mathcal{X}$ es infinito , en cuyo caso el muestreo es desde lo que llamaremos un proceso de Dirichlet.

Capítulo 2

Procesos de Dirichlet

2.1. Procesos de Dirichlet (DP)

Según lo observado en los párrafos correspondientes al capítulo precedente podemos entender la distribución Beta como una distribución sobre el parámetro p de la distribución Bernoulli, el cual define una medida de probabilidad sobre un espacio $\mathcal{X}$ finito dimensional. Asi mismo la distribución de Dirichlet como una extensión de la distribución Beta, la podemos entender como una distribución sobre los parámetros $(p_1, p_2, ..., p_k)$ de una distribución multimomial. En ambos casos lo que hemos hecho es asignar una medida de probabilidad a un espacio finito de medidas de probabilidad. La idea es extender esta forma de entender dichas distribuciones en el caso en que $\mathcal{X}$ es un conjunto infinito dimensional, típicamente un espacio métrico separable.

Es en la "realización" de esta idea donde la definición de los procesos de Dirichlet debida a Ferguson (1973), gana importancia como un caso particular de una definición más general como son medidas de probabilidad aleatorias.

2.1.1. Medidas de Probabilidad Aleatorias

Denotaremos por $\mathcal{M}(\mathcal{X})$ al conjunto de todas las medidas de probabilidad sobre $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$, con $\mathcal{B}(\mathcal{X})$ el correspondiente σ-álgebra de Borel sobre $\mathcal{X}$. Los casos analizados en el capítulo precedente consideraba a $\mathcal{X}$ como un conjunto finito. Será ahora de nuestro interés considerar $\mathcal{X} = \mathbb{R}$.

Desde el punto de vista Bayesiano requeriremos distribuciones a priori sobre $\mathcal{M}(\mathcal{X})$, es decir medidas de probabilidad sobre espacios de medidas de probabilidad.

La topología del espacio $\mathcal{M}(\mathcal{X})$ depende de la métrica definida sobre el espacio. Algunas de ellas son, la métrica de variación total o métrica L_1, la métrica de Hellinger, la métrica del supremo, la métrica de la convergencia débil etc..

En la revisión de la literatura las topologías más usadas provienen de las métricas de variación total y de la convergencia débil, y en este sentido miraremos como son los conjuntos abiertos del espacio $\mathcal{M}(\mathcal{X})$ con la idea de a posterior entender el concepto de consistencia de los estimadores bayesianos no paramétricos.

Métrica de Variación Total

Sean P y Q pertenecientes a $\mathcal{M}(\mathcal{X})$, se define la métrica de variación total entre P y Q como

$$||P - Q|| = 2sup_B|P(B) - Q(B)| \quad B \subset \mathcal{X}$$

Cuando la métrica de variación total se restringe al subconjunto, L_μ, de distribuciones que poseen densidad respecto de alguna medida σ-finita, μ. Llamemos p y q a las correspondientes densidades de P y Q, entonces

$$||P - Q|| = \int |p - q| d\mu$$

que corresponde a la distacia L_1 entre p y q. Es un resultado conocido que este subespacio, L_μ con la métrica L_1 es completo y separable.

Métrica de la convergencia Débil

En la literatura se establece que la topología más natural sobre $\mathcal{M}(\mathcal{X})$ es la que proviene de la métrica de la convergencia débil (weak topology).

Para cualquier $P_o \in \mathcal{M}(\mathcal{X})$ una base de vecindades en esta topología, consiste de los conjuntos de la forma

$$\bigcap_{1}^{k}\{|\int f dP_o - \int f_i dP| < \epsilon\}$$

donde $f_i, \ldots i = 1, 2, ..., k$ son funciones acotadas y continuas sobre $\mathcal{X}$.

Es importante destacar que bajo esta topología $\mathcal{M}(\mathcal{X})$ es un espacio métrico completo y separable por lo que se puede definir el σ-álgebra de borel $\mathcal{B}_{\mathcal{M}(\mathcal{X})}$ sobre $\mathcal{M}(\mathcal{X})$ como el menor σ-álgebra generado por todos los conjuntos abiertos (de esta topología). Otra forma es decir que $\mathcal{B}_{\mathcal{M}(\mathcal{X})}$ es el menor σ-álgebra que hace que las funciones $\{P \longrightarrow P(B) : B \in \mathcal{B}(\mathcal{X})\}$ sean medibles. Una discusión sobre los distintos σ-álgebras generados por las distintas topologías puede ser encontrado en Gaudard y Hadwin (1989).

Dentro del contexto del desarrollo de esta tesis es interesante el resultado obtenido en la siguiente proposición, cuya demostración puede ser encontrada en Ghosh y Ramamoorthi (2003)

Proposición 2.1. El conjunto de medidas de probabilidad discretas es un subconjunto medible de $\mathcal{M}(\mathcal{X})$.

Finalmente si consideramos μ una medida σ-finita sobre $\mathcal{X} = \mathbb{R}$, entonces L_μ es un subconjunto medible de $\mathcal{M}(\mathcal{X})$.

2.1.2. Medidas de probabilidad sobre $\mathcal{M}(\mathcal{X})$

Llamamos medida de probabilidad aleatoria $\mathbf{P}$ a la colección

$$\mathbf{P} = \{P(B, w), B \in \mathcal{B}(\mathcal{X}), w \in \Omega\}$$

definidas sobre algún espacio de probabilidad $(\Omega, \mathcal{B}(\mathcal{X}), P)$ tal que :

Para cada $w \in \Omega$ la función

$$B \rightarrow P(B, w)$$

es una medida de probabilidad sobre un espacio $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$, donde $\mathcal{B}(\mathcal{X})$ es un $\sigma-$álgebra asociado al espacio $\mathcal{X}$.

Para cada $B \in \mathcal{B}(\mathcal{X})$ la función

$$w \rightarrow P(B, w),$$

es una variable aleatoria que toma valores en $\mathcal{X}$.

Un ejemplo básico de una medida de probabilidad aleatoria es la función de distribución empírica. Sean $X_1, X_2, ..., X_n$ una muestra aleatoria desde una distribución desconocida F sobre $\mathcal{X}$. Sea $B \in \mathcal{B}(\mathcal{X})$. Se define la función

$$F_n(B) = \frac{1}{n} \sum_{i=1}^{n} I_B(X_i)$$

Observamos que F_n es una función medible desde el espacio $(\mathcal{X}^n, \mathcal{B}(\mathcal{X})^n)$ hasta el espacio de todas las medidas de probabilidad sobre $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$

2.1.3. Procesos de Dirichlet

Los procesos de Dirichlet (DP) fueron introducidos por ?, (en la cual estudiaba y analizaba formas de introducir medidas de probabilidad aleatorias que cumplieran con ciertos requisitos básicos), como una distribución a priori sobre el espacio de distribuciones de probabilidad sobre un espacio medible $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ dado.

Como una forma de profundizar en esta definición daremos una mirada a algunas de las caracterizaciones y representaciones del proceso que se han presentado a lo largo del tiempo.

Sea $\mathcal{X}$ un conjunto con un número infinito de elementos (por ejemplo $\mathbb{R}$, el intervalo $[0,1]$, $\mathbb{R}^k$ etc...) y sea $\mathcal{B}(\mathcal{X})$un $\sigma-$álgebra de subconjuntos de $\mathcal{X}$. Ferguson (1973) define un proceso de Dirichlet con parámetro α (α una medida finita , no nula) como una medida de probabilidad aleatoria P sobre $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$, tal que para toda partición $A_1, A_2, ..., A_k$ de conjuntos Borel medibles de $\mathcal{X}$ y , $k \in \mathbb{N}$ se tiene que la distribución del vector aleatorio $(P(A_1), P(A_2), ..., P(A_k))$ es la distribución de Dirichlet con parámetros $(\alpha(A_1), \alpha(A_2), ..., \alpha(A_k))$

Ferguson verifica las condiciones de consistencia de Kolmogorov para mostrar la existencia de este proceso.

En adelante y motivados por las distintas aplicaciones en las que se utiliza dichos procesos definiremos lo que es una distribución de probabilidad aleatoria , F, definida por un DP.

Una distribución de probabilidad aleatoria, F, es definida por un DP si para cualquier partición $A_1, A_2, ..., A_k$ del espacio muestral el vector probabilidades aleatorias

$$(F(A_1), F(A_2), ..., F(A_k)) \sim Dirichlet(MF_o(A_1), MF_o(A_2), ..., MF_o(A_k))$$

donde $M > 0$ es un parámetro de precisión, Walker y otros (1999) y F_o es la medida base, definida por el valor esperado $E(F(B)) = F_o(B)$

La notación a utilizar será $F \sim DP(M, F_o)$, para indicar que la distribución de probabilidad aleatoria F es definida por un proceso de Dirichlet con medida de centro F_o y parámetro de precisión M.

En adelante entenderemos indistintamente que la distribución de probabilidad alea-

toria F sigue un proceso de Dirichlet, cuando ella es definida por una medida de probabilidad aleatoria que sigue un DP.

Es importante para la interpretación de los parámetros del proceso recordar que la distribución de probabilidad donde se centra el proceso está definida por el valor esperado

$$E(F(A)) = F_o(A).$$

Y que el parámetro M interviene en la varianza del proceso según la igualdad

$$Var(F(A)) = \frac{F_o(A)(1 - F_o(A))}{M + 1}$$

Según esta expresión tenemos que si M es "grande" el proceso F está altamente concentrado alrededor de F_o justificando la terminología anterior, que define a M como un parámetro de precisión.

En particular, para cualquier $B \in \mathcal{B}(\mathcal{X})$, si $F_o(B) > 0$ entonces el proceso $F(B)$ sigue una distribución $Beta(MF_o(B), M(F_o(B^c))$ y con esto $F(B) > 0$ *c.s.*

Una propiedad a destacar de los procesos de Dirichlet es que la restricción del proceso de Dirichlet a un subconjunto A de $\mathcal{X}$ tal que $F_o(A) > 0$ es un proceso de Dirichlet F_A definido por $F_A(B) = \frac{F(A \cap B)}{F(A)}$ con parámetros $M_1 = MF_o(A)$ y $F_o|_A$. ($F_o|_A$ representa la restricción de F_o al conjunto A)

Es evidente las distintas propiedades que hace interesante la utilización de este proceso pero una crítica estándar a los procesos de Dirichlet es que son casi seguramente discretos, lo que los hace inadecuados para aplicar en distintos problemas. Para ver esta discretización resulta muy útil la construcción debida a Sethuraman (1994) que establece

que si $F \sim \mathcal{D}(M, F_o)$ entonces éste puede ser casi seguramente representado como:

$$F(.) = \sum_{h=1}^{\infty} w_h \delta_{\mu_h}(.);$$

donde $\mu_h \overset{iid}{\sim} F_o$ y $w_h = U_h \prod_{j<h}(1 - U_j)$ con $U_h \overset{iid}{\sim} Beta(1, M)$, donde $\delta_a(.)$denota una masa puntual en a.

Es decir, con probabilidad 1 desde los DP se obtienen distribuciones que pueden ser representados como mezclas infinitas de masas puntuales , donde las localizaciones de las masas puntuales, μ_h, son muestreados desde F_o y los pesos aleatorios w_h son generados por un proceso "stick-breaking", basado sobre variables aleatorias i.i.d. $Beta(1, M)$.

Un resultado destacable obtenido por Ferguson, especificamente para realizar inferencia a posterioric es que si $F \sim D(M, F_o)$ y $X_1, X_2, ...X_n$ es una muestra aleatoria de tamaño n desde F entonces la distribución condicional de F dado $X_1, X_2, ...X_n$ es un proceso de Dirichlet de parámetros $M + n$ y F_1 donde $F_1 \propto F_o + \sum_{i=1}^{n} \delta_{X_i}$ con

$$\delta_x(A) = \begin{cases} 1, & x \in A \\ 0, & x \notin A \end{cases}$$

2.1.4. Esquema de Urna De Polya

Blackwell y MacQueen (1973) extendieron la representación en esquema de urna de distribuciones sobre espacio finito dimensionales a las medidas de probabilidad aleatorias, específicamente a los procesos de Dirichlet, conectándolos con el esquema de Urna de Polya.

Partamos por definir una secuencia de Polya con parámetro F_o (F_o, distribución no nula definida sobre $\mathcal{X}$) como una secuencia de variables aleatorias $\{X_n\}_n$ con valores en $\mathcal{X}$ si $\forall B, B \subset \mathcal{X}$

1. $P(X_1 \in B) = F_o(B)$

2. $P(X_{n+1} \in B / X_1, ..., X_n) = F_n(B)/F_n(\mathcal{X})$

donde

$$F_n(.) = F_o(.) + \sum_1^n \delta_{X_i}(.)$$

En el caso particular en que $\mathcal{X}$ fuera un conjunto finito y F_o fuera una medida sobre $\mathcal{X}$ la secuencia $\{X_n\}_n$ puede ser entendida como una representación del resultado de sucesivas extracciones desde una urna donde inicialmente la urna contiene $F_o(x)$ bolitas de color x y después de cada extracción, la bolita seleccionada es reemplazada y otra bolita del mismo color es agregada a la urna.

Siguiendo esta idea y generalizando este resultado al caso en que el espacio de colores es un espacio continuo, Blackwell y MacQueen (1973) establecieron el siguiente resultado, el cual es crucial para la implementación computacional desde un proceso de Dirichlet.

Teorema 2.1. Si $\{X_n\}_n$ es una secuencia de Polya con valores en $\mathcal{X}$ con parámetro Q entonces se tiene que

1. $m_n = \frac{Q_n(.)}{Q_n(\mathcal{X})}$ converge con probabilidad 1 cuando n tiende a infinito a una medida de probabilidad aleatoria discreta Q^*

2. Q^* sigue un proceso de Dirichlet de parámetro Q

3. Dado Q^*, las variables $X_1, X_2, ...$ son independientes con distribución Q^*

Con este resultado podemos entender un proceso de Dirichlet , como una medida de probabilidad aleatoria límite de una secuencia de Polya.

2.2. Algunos procesos construidos a partir de los procesos de Dirichlet

A pesar de que los procesos de Dirichlet son los procesos más populares en la literatura no paramétrica, sabemos que son casi seguramente discretos y por tanto son inapropiados para modelar algunos fenómenos. Una extensión de estos procesos que buscan aliviar esta característica, particularmente en aplicaciones donde se requiere estimar una densidad, la podemos encontrar en dos trabajos, el primero debido a Ferguson (1983) y el más conocido debido a Lo (1984) , donde se realiza una convolución de un kernel de una distribución continua con una medida de probabilidad aleatoria que sigue un proceso de Dirichlet. Esta convolución se puede representar a través del siguiente modelo jerárquico.

$$\begin{aligned} Y_i|\theta_i & \sim f(.|\theta_i), i = 1, 2, ..., n \\ \theta_1, \theta_2, ..., \theta_n|F & \overset{iid}{\sim} F \\ F & \sim DP(M, F_o) \end{aligned}$$

En este modelo se mezclan densidades continuas dependientes de algún parámetro θ, cuya distribución F es desconocida y por tanto aleatoria a la que se da como priori un proceso de Dirichlet.

Estos procesos son conocidos por el nombre de *Dirichlet Mixtures, (MDP)*y son ampliamente utilizados. Algunas aplicaciones las podemos encontrar en Escobar (1994), West y otros (1994a), Escobar y West (1995a), Escobar y West (1995b),Müller y otros (1996) , Bush (1993) y MacEachern (1998).

En el caso particular del trabajo de West y otros (1994a), utilizan un kernel Normal, de la forma

$$f(x|\mu, S) \propto \exp\left[-(x-\mu)^t S^{-1}(x-\mu)/2\right]$$

y lo mezclan respecto de $\theta = (\mu, S)$.

La inferencia para estos modelos es basada en simulaciones MCMC. La eficiencia, rapidez y otras características de tales simulacions para modelos MDP es ampliamente discutida en los trabajos de Bush (1993), MacEachern (1998), West y otros (1994a) y Neal (2000).

Otra extensión de los procesos de Dirichlet es debida a Antoniak (1974) y son las mezclas de Procesos de Dirichlet,denotadas por

$$\int \mathcal{D}(M_\theta, H_\theta)\pi(d\theta)$$

donde cada proceso tiene medida de centro H_θ, que depende de un parámetro θ y de números positivos $M_\theta, \theta \in \Theta$ y la mezcla se hace sobre la distribución a priori π de θ.

Una forma de entender estos procesos es pensar que en una primera etapa se selecciona θ desde la distribución π y en una segunda etapa se condiciona en θ y se asigna a F una priori Dirichlet

$$\mathcal{D}(M_\theta, H_\theta)$$

En este contexto un modelo ampliamente utilizado es aquel en donde $\{H_\theta\}$ pertenece a la familia de distribuciones Normal de media μ y varianza σ^2 y π es una priori sobre $\theta = (\mu, \sigma^2)$, donde se consideran μ_i i.i.d. F con

$$F \sim \mathcal{D}(M_\theta, H_\theta)$$

Este modelo es de gran interés en publicaciones especialmente médicas que realizan Meta-análisis , en cuyo caso el objeto de interés en el modelo es μ. Es importante destacar que en el contexto de estas mezclas este parámetro no tienen una interpretación clara, como por ejemplo representar la media o la mediana de F. Esta limitación es abordada por Doss (1985) quien utiliza en este tipo de modelos una nueva clase de

procesos en los cuales, μ tiene una clara interpretación, y corresponde a la mediana del proceso con probabilidad 1.

2.2.1. Procesos de Dirichlet Invariantes

Los procesos de Dirichlet invariantes fueron introducidos por Dalal (1979) el objetivo de su publicación era hallar una clase de distribuciones prioris sobre un espacio de medidas de probabilidad menos amplio que el que interesaba a Ferguson, ya que su estimación no paramétrica tenía supuestos de simetría o permutabilidad lo cual requería un espacio de distribuciones más restringido y específico, de hecho las distribuciones deben ser invariantes bajo cierto grupo de transformaciones. Para resolver esto define un proceso un poco más complejo que un Proceso de Dirichlet, al que dió el nombre de *Proceso de Dirichlet Invariante* en los que desarrolla propiedades análogas a las desarrolladas por Ferguson.

Además del trabajo de Dalal existe otro trabajo donde se presenta un método alternativo para obtener una medida de probablidad aleatoria simétrica, ésta es sugerida por Hannum y Hollander (1983) y consiste en elegir una medida de probabilidad que siga un proceso de Dirichlet y luego simetrizarlo, más específicamente si P sigue un proceso de Dirichlet sobre $(\mathbb{R}^+, \mathcal{B}(\mathbb{R}^+))$ tal que $F(x) = P((-\infty, x])$ y $F(x^-) = P((-\infty, x))$, entonces

$$F^*(x) = \frac{1}{2}F(x) + \frac{1}{2}(1 - F(-x^-)$$

es una medida de probabilidad aleatoria simétrica alrededor de cero. Los autores prueban que F^* sigue el proceso de Dirichlet Invariante, definido por Dalal (1979).

Freedman y Diaconis (1982) realizan inferencia y muestran problemas de inconsistencia cuando se utilizan ciertos tipos de medidas basales del proceso.

Como hemos visto en la construcción anterior, se considera una medida de probabilidad , F, concentrada sobre $[0, +\infty)$ y se obtiene su versión simetrizada dada por

$$F_s(t) = \frac{1}{2}F(t) + \frac{1}{2}(1 - F(-t^-))$$

Esto nos lleva a afirmar que la construcción de funciones de distribuciones aleatorias sobre $\mathbb{R}$ simétricas alrededor de cero es equivalente a la construcción de funciones de distribución aleatorias sobre $[0, +\infty)$.

Es importante destacar que las propiedades de los Procesos de Dirichlet Invariantes estudiadas por Dalal, son análogas a los de procesos de Dirichlet. En particular se tiene que si F es un proceso de Dirichlet Invariante de parámetro α sobre $\mathcal{X}$ y $X_1, X_2, ..., X_n$ es una muestra de tamaño n desde F , entonces la distribución condicional de F dado $X_1, X_2, ...X_n$ es un procesos de Dirichlet Invariante con medida de centro $\alpha + \sum_{j=1}^{n} \delta^g_{X_j}$ donde

$$\delta^g_{X_j}(.) = \frac{1}{k}\sum_{i=1}^{k} \delta_{g_i X_j(.)}$$

$\{g_i\}_{i=1}^n$ es un grupo de transformaciones invariantes sobre $\mathcal{X}$.

En el caso particular en que k=2, y se considera el grupo de transformaciones $\{g_1(x) = x, g_2(x) = -x\}$ se obtiene una representación para un proceso de Dirichlet simetrizado.

2.2.2. Otras medidas de probabilidad aleatorias de interés

Doss (1984) construye una clase de prioris que dan probabilidad 1 al espacio de las funciones de distribución simétricas. Esta clase contiene los procesos de Dirichlet Simétricos usados por Dalal(1979) y Diaconis y Freedman(1982).

La construcción de la priori definida sobre el espacio de las funciones de distribución simétricas alrededor de cero es basada en el trabajo de Doksum (1974) sobre prioris

neutrales a la derecha e incluyen la priori definida por Dalal(1979). En otro trabajo, Doss(1985) también construye ciertos procesos que son asimétricos con probabilidad 1 y que concentran su masa en distribuciones con mediana igual a cero. Éstos contienen como caso particular a los procesos de Dirichlet simétricos definidos por Dalal(1979). La construcción de tales procesos es la siguiente:

Sea $\alpha = M\alpha_o$ una medida finita no nula sobre $\mathbb{R}$ con mediana igual a cero. Sean α_- y α_+ las restricciones de α a $(-\infty, 0)$ y $(0, +\infty)$ respectivamente.

Eligiendo F_- y F_+ independientemente desde los procesos de Dirichlet con medida basal α_- y α_+ respectivamente se obtienen la medida de probabilidad aleatoria

$$F(t) = \frac{1}{2}F_-(t) + \frac{1}{2}F_+(t)$$

que tiene mediana igual a cero pero con probabilidad 1 no es simétrico. Doss hace notar que si el soporte de α es suficientemente grande, así lo es el soporte del proceso, también se rescata la interpretación de los parámetros, donde M indica el grado de concentración del proceso alrededor de α_o.

Este nuevo proceso puede ser visto como la distribución condicional de un proceso de Dirichlet corriente con parámetro α, dado que la mediana de F es igual a cero.

Capítulo 3

Procesos Skew Dirichlet

En ciertas aplicaciones es de interés definir distribuciones con soporte contenido en algún subconjunto del espacio de distribuciones continuas. Como por ejemplo el espacio de distribuciones simétricas, si consideramos el caso de los errores en un modelo de regresión que tienen distribución simétrica, los procesos de Dirichlet no serían adecuados para la modelación ya que ellos son asimétricos con probabilidad 1. Dalal(1979) introduce una clase de medidas de probabilidad concentrada sobre las distribuciones simétricas a las que llamó procesos de Dirichlet Invariantes. En el caso especial en el que el espacio de soporte es los reales y considerando particiones de los reales que estén constituidas por conjuntos simétricos, tal proceso concentra su masa sobre las distribuciones simétricas. Esta construcción ha sido insuficiente ya que en general se desea obtener las distribuciones simétricas como un caso especial dentro de un grupo más general. En este aspecto ni los procesos de Dirichlet ni los procesos de Dirichlet Invariantes son flexibles para acomodar asimetría y simetría como un caso especial.

En este trabajo se introduce una nueva familia de medidas de probabilidad aleatorias cuyos elementos se construyen como una mezcla convexa de un proceso de Dirichlet concentrado sobre los reales positivos y otro concentrado sobre los reales negativos y cuya mezcla depende sólo de un parámetro de asimetría que designamos por θ y

tal que la simetría se obtiene como un caso particular cuando $\theta = 1/2$ y cuando en la definición de las distribuciones finito dimensionales la partición se concentra sobre conjuntos simétricos de los reales.

La idea básica parte desde la representación de cualquier variable aleatoria X con distribución simétrica como $U \cdot V$ donde V tiene la distribución de $|X|$ (concentrada sobre $\mathbb{R}^+$) y U es uniforme sobre $\{-1, 1\}$. Con esta idea en mente tenemos que si U es distribuida tal que $P(U = 1) = \theta = 1 - P(U = -1)$ con $0 < \theta < 1$ entonces la distribución de $U \cdot V$ es asimétrica y es simétrica si y sólo si $\theta = \frac{1}{2}$.

3.1. Definiciones y algunas propiedades

Se define una nueva familia de medidas de probabilidad aleatorias que se deriva a partir de la definición de procesos de Dirichlet y que incluyen como caso especial a los procesos de Dirichlet invariantes.

Definición 3.1. Una medida de probabilidad aleatoria F se llamará *Proceso Skew Dirichlet* sobre $\mathbb{R}$ con parámetros M, F_0, θ si para cualquier partición medible $A_1, A_2, ..., A_n$ de $\mathbb{R}$ se tiene que

$$(F(A_1), \ldots, F(A_k)) \sim \theta Dir(MF_0(A_1), \ldots, MF_0(A_k)) + (1-\theta) Dir(MF_0(-A_1), \ldots, MF_0(-A_k)), \quad (3.1)$$

donde F_0 es una *función de distribución* concentrada sobre $\mathbb{R}^+$, $0 < \theta < 1$ es el *parámetro de asimetría*, $M > 0$ es el parámetro de *masa total* o *peso a priori* y $Dir(\alpha_1, ..., \alpha_k)$ representa la Distribución de Dirichlet con parámetros $(\alpha_1, \ldots, \alpha_k)$.
En adelante denotaremos a este proceso por $F \sim SDP(M, F_0, \theta)$.

Comparando (3.1) con la definición de un DP dada por Ferguson (1973), los SDP pueden ser entendidos como una reflexión alrededor de 0 de un DP regular concentrado sobre $\mathbb{R}^+$ pero donde el signo positivo es asignado con probabilidad θ.

Podemos entender al parámetro θ como un indicador, según su valor se encuentre más cercano a $1/2$, a que el soporte del proceso F este constituido por distribuciones simétricas o asimétricas.

3.1.1. Algunas propiedades de los procesos Skew Dirichlet

Es interesante destacar que los SDP además de la definición a partir de las trayectorias finito dimensionales pueden ser construidos de la siguiente forma:

Sea $M > 0$ y sea F_{0+} una distribución concentrada en $\mathbb{R}^+$, y consideremos el proceso de Dirichlet, F_+, con distribución a priori $DP(M, F_{0+})$. A partir de este proceso construyamos un nuevo proceso F, $\forall t \in \mathbb{R}$ tal que :

$$F(t) = \theta F_+(t) + (1-\theta)(1 - F_+(-t^-))$$

Es claro que estos procesos son *Procesos Skew Dirichlet*. Esto será demostrado más adelante en el *Teorema* 3.5.

Esta representación tiene la ventaja de permitirnos identificar e interpretar más fácilmente a los parámetros del proceso. En efecto desde esta identidad podemos observar sin mayor dificultad que el percentil $(1-\theta)$ de F es igual a cero con probabilidad 1 y que

$$E(F(t)) = \theta F_{0+}(t) + (1-\theta)(1 - F_{0+}(-t^-)) \tag{3.2}$$

es decir el valor esperado de un SDP es una distribución asimétrica, salvo el caso en

que $\theta = 1/2$.

También se rescata la interpretación del parámetro M como indicador del grado de concentración del proceso Skew Dirichlet alrededor de F_0. En efecto

- Si $F \sim SDP(M, F_0, \theta)$ y A es cualquier evento en $\mathcal{B}(\mathbb{R})$ entonces

$$\mathbb{V}(F(A)) = \frac{1}{M+1}\{\theta^2\mathbb{V}(F_+(A)) + (1-\theta)^2\mathbb{V}(1 - F_+(-A))\}. \tag{3.3}$$

  Ya que en el proceso de Dirichlet regular F_+ se tiene que

$$\mathbb{V}(F_+(A)) = \frac{F_0^+(A)(1 - F_0^+(-A))}{M+1}$$

- En el caso particular donde $\theta = \frac{1}{2}$ y los eventos A están restringidos a satisfacer $A = -A$, entonces los SDP llegan a ser los procesos de Dirichlet Invariantes definidos por Dalal (1979).

- Cualquier SDP sobre $\mathbb{R}$ es una combinación convexa de dos DPs, uno concentrado sobre $\mathbb{R}^+$ y el otro sobre $\mathbb{R}^-$.

Observación 3.1. En general los SDP no son conjugados bajo muetreo independiente. A pesar de ello esta condición no es restrictiva ya que las simulaciones del proceso en las distintas aplicaciones se pueden obtener utilizando muestreo de Gibbs en MCMC.

3.1.2. Otra representación de los procesos Skew Dirichlet

El siguiente resultado extiende la representación dada por Sethuraman (1994) a los procesos de Dirichlet para los procesos Skew Dirichlet , como sigue:

Teorema 3.1. Cualquier $F \sim SDP(M, F_0, \theta)$ puede ser representado como

$$F(.) = \theta \sum_{i=1}^{\infty} \omega_h \delta_{V_h}(.) + (1-\theta) \sum_{i=1}^{\infty} \omega_h \delta_{-V_h}(.)$$

donde $V_1, V_2, ..., . \overset{iid}{\sim} F_0$ y $\omega_1, \omega_2, ...$ son pesos estocásticamente ordenados que siguen un proceso " stick breaking " como sigue

$$\omega_1 = U_1$$

y

$$\omega_h = \prod_{i=1}^{h-1}(1-U_i)U_h \quad \forall h > 2,$$

con $U_1, U_2, ... \overset{iid}{\sim} Beta(1, M)$

Esta representación muestra que cualquier $F \sim SDP(M, F_0, \theta)$ puede ser mirado como una suma ponderada de dos mezclas infinitas de masas puntuales con distribución marginales F_0 y $1 - F_0 o \varphi$ respectivamente (entenderemos en adelante que $\varphi(x) = -x$). Este resultado también sirve para verificar la afirmación de que la simetría se obtiene sólo si $\theta = 1/2$.

Además podemos notar que cuando F_0 tiene densidad f_0, F tienen densidad

$$\theta f_0(x) + (1-\theta) f_0(-x) \forall x \in \mathcal{X}. \tag{3.4}$$

Observemos que esta densidad es discontinua en $x = 0$ a menos que $f_0(0) = 0$.

Una simple solución a este problema es considerar para $0 < \theta < 1$

$$F(\cdot) = \sum_{h=1}^{\infty} w_h \delta_{\{\theta V_h\}}(\cdot) + \sum_{h=1}^{\infty} w_h \delta_{-\{(1-\theta)*V_h\}}(\cdot), \tag{3.5}$$

con ponderaciones $\{w_h\}$ y variables aleatorias $\{V_h\}$ definidas como antes.

Bajo (3.5) la densidad marginal de cada componente mezcla es

$$f_0\left(\frac{x}{\theta}\right) I\{x \geq 0\} + f_0\left(-\frac{x}{1-\theta}\right) I\{x < 0\},$$

(ver Arellano-Valle y otros (2005)) la cual *es continua.*

Bajo esta construcción alternativa,la teoría desarrollada es válida, con los cambios correspondientes. Notemos que la simetría es obtenida, si y sólo si $\theta = \frac{1}{2}$.

3.1.3. Representación de SDP en esquema de urna

Aplicaciones prácticas de SDP requieren algunos resultados extras. Específicamente y en analogía con los procesos de Dirichlet, se discutirá la representación en esquema de urna de la distribución marginal de una muestra

$$X_1, \ldots, X_n | F \overset{iid}{\sim} F$$

$$F \sim SDP(M, F_0, \theta)$$

Tal representación es particularmente útil para el esquema de simulación a posteriori.

La descripción siguiente fue desarrollada en el espíritu de la aplicación de los resultados debidos a Blackwell y MacQueen (1973).

Partamos notando que si $X_1, X_2, \ldots, X_n | F \overset{iid}{\sim} F$ and $F \sim SDP(M, F_0, \theta)$ entonces para todo $B \in \mathcal{B}(\mathbb{R})$

$$P(X_1 \in B) = \mu(B) \tag{3.6}$$

$$P(X_{l+1} \in B | X_1, X_2, \ldots, X_l) = \frac{\mu_l(B)}{\mu_l(\mathbb{R})}, \tag{3.7}$$

para todo $l = 1, \ldots, n-1$, donde $\varphi(B) = -B$, $\mu(B) = \mathbb{E}(P(B))$ dado en (3.2), y para $l \geq 1$

$$\mu_l(B) = M\mu(B) + \theta \sum_{i=1}^{l} \delta_{|X_i|}(B) + (1 - \theta) \sum_{i=1}^{l} \delta_{|X_i|}(\varphi(B)) \quad (3.8)$$

así que $\mu_l(\mathbb{R}) = 2(M + l)$. Aquí $\delta_z(\cdot)$ representa una distribución degenerada concentrada z.

Cualquier secuencia de variables aleatorias satisfaciendo (3.6) y (3.7) será llamada *Secuencia Skew Polya.*

Observación 3.2. Las variables que conforman una secuencia de Polya son construídas asignando signo positivo o negativo a los valores absolutos de una secuencia de Polya usual, donde la probabilidad de asignar el signo positivo es θ.

A continuación se presentam una serie de proposiciones que permiten establecer relaciones entre una secuencia Skew Polya con una secuencia de Polya usual.

En primer lugar el siguiente teorema establece, entre otras cosas, la independencia condicional del módulo de una variable futura de la secuencia de Skew Polya con la muestra $X_1, X_2, ...X_n$.

Teorema 3.2. Sea $\{X_n, n \geq 1\}$ una secuencia Skew Pólya con parámetros M , F_0 y θ. Entonces

a) $|X_{n+1}| \amalg (X_1, X_2, ...X_n)/|X_1|, |X_2|, ..., |X_n|$

b) La secuencia de variables aleatorias $\{Y_n = sign(X_n), n \geq 1\}$ es i.i.d.

Demostración 3.1. Según vemos en la ecuación (3.8) la probabilidad predictiva sólo depende de los valores previos de la secuencia, $X_1, X_2, ..., X_n$, a través de sus valores absolutos $|X_1|, |X_2|, ..., |X_n|$ con esto se muestra que

$$P(X_{n+1} \in B \mid X_1, \ldots, X_n) = P(X_{n+1} \in B \mid |X_1|, \ldots, |X_n|), \tag{3.9}$$

desde aquí (a) se sigue fácilmente.

Además

$$P(Y_{n+1} = 1|Y_1, ..., Y_n) \quad = \quad P(X_{n+1} > 0|Y_1, ..., Y_n)$$

Pero

$$E(I_{X_{n+1}}|Y_1, ..., Y_n) = E(E(I_{X_{n+1}}|Y_1, .., Y_n, |X_1|, .., |X_n|)|Y_1, .., Y_n)$$

Y

$$E(E(I_{X_{n+1}}|X_1, ..., X_n)|Y_1, .., Y_n) = E(P(X_{n+1}) > 0/X_1, .., X_n)|Y_1, ..., Y_n)$$

Además por (3.7), $P(X_{n+1} > 0 \mid X_1, \ldots, X_n) = \theta$, asi que

$$P(Y_{n+1} = 1 \mid Y_1, \ldots, Y_n) = \theta,$$

desde lo cual se sigue (b). ■

Es interesante establecer relaciones entre la secuencia Skew Polya con una secuencia de Polya usual, la proposición que se presenta a continuación con su respectivo corolario muestra como a partir de una secuencia Skew Polya sobre $\mathbb{R}$ se puede obtener una secuencia de Polya sobre $\mathbb{R}^+$ y consecuentemente otra secuencia de Polya sobre $\mathbb{R}^-$

Proposición 3.1. Si $\{X_n, n \geq 1\}$ es una Secuencia Skew Polya con medida de centro F_0 y parámetro de concentración alrededor de F_0 a priori M y parámetro de asimetría θ, entonces $\{|X_n|, n \geq 1\}$ es una secuencia de Polya con medida de centro F_0 y parámetro de precisión M.

Demostración 3.2. Ya que $\{X_n, n \geq 1\}$ es una secuencia Skew Pólya , entonces la probabilidad predictiva sólo depende de las $X_1, X_2, ..., X_n$ a través de los valores absolutos $|X_1|, |X_2|, ..., |X_n|$ con esto se muestra que

$$P(X_{n+1} \in B/X_1, X_2, ..., X_n) = P(X_{n+1} \in B/|X_1|, |X_2|, ..., |X_n|) \tag{3.10}$$

Se sigue desde (3.10) que para todo $A \in \mathcal{B}(\mathbb{R}^+)$

$$\begin{aligned} P(|X_{n+1}| \in A/|X_1|, ..., |X_n|) =& P(X_{n+1} \in A/|X_1|, ..., |X_n|) + P(X_{n+1} \in -A/|X_1|, ..., |X_n|) \\ =& P(X_{n+1} \in A/X_1, ..., X_n) + P(X_{n+1} \in -A/X_1, ..., X_n) \\ =& \theta \frac{MF_0(A) + \sum_{i=1}^n \delta_{|X_i|}(A)}{\mu(\mathbb{R}^+) + n} + (1-\theta) \frac{F_0\varphi(A) + \sum_{i=1}^n \delta_{|X_i|}(\varphi(A))}{F_0(\mathbb{R}^+) + n} \end{aligned}$$

Consecuentemente,

$$P(|X_{n+1}| \in A/|X_1|, |X_2|, ..., |X_n|) = \frac{F_0(A) + \sum_{i=1}^n \delta_{|X_i|}(A)}{F_0(\mathbb{R}^+) + n}$$

y la probabilidad

$$P(|X_1| \in A) = \frac{MF_0(A)}{F_0(\mathbb{R}^+)} = F_0(A)$$

■

Corolario 3.3. Sea $\{X_n, n \geq 1\}$ una Secuencia Skewed Pólya con parámetros M, F_0, θ entonces $\{-|X_n|, n \geq 1\}$ es una Secuencia de Pólya con parámetros $M, F_0\varphi$.

Como hemos visto en las proposiciones precedentes se obtiene de forma natural secuencias de Polya sobre $\mathbb{R}^+$ y $\mathbb{R}^-$ a partir de secuencias Skew Polya sobre $\mathbb{R}$. El siguiente Teorema establece el recíproco, es decir la construcción de una Secuencia Skew Polya con parámetros M, F_o, θ a partir de una secuencia de Polya con parámetros M, F_o donde $F_o(0) = 0$.

Teorema 3.4. Sea $\{|X_n|, n \geq 1\}$ una Secuencia de Pólya con parámetros M y F_0 entonces $Y_i = U_i|X_i|$ con U_i iid $P(U_i = 1) = \theta = 1 - P(U_i = -1)$ y U_i independiente de $|X_i|$ entonces $\{Y_i\}$ es una Secuencia Skew Pólya con parámetros M, F_0, θ

Demostración 3.3. Ya que $Y_i = \begin{cases} |X_i| & U_i = 1 \\ -|X_i| & U_i = -1 \end{cases}$

Tenemos que en general

$$|X_{n+1}| \coprod U_1, U_2, ..., U_n / |X_1|, |X_2|, ..., |X_n|$$

y esto dice que

$$P(Y_{n+1} \in b|Y_1, .., Y_n) = \begin{cases} P(|X_{n+1}| \in B||X_1|, ..., |X_n|), & \text{si } U_{n+1} = 1 \\ P(-|X_{n+1}| \in B| - |X_1|, ..., -|X_n|), & \text{si } U_{n+1} = -1 \end{cases}$$

Entonces

$$P(Y_{n+1} \in B|Y_1, .., Y_n) = \theta P(|X_{n+1}| \in B||X_1|, ..., |X_n|) + (1-\theta) P(-|X_{n+1}| \in B| - |X_1|, ..., -|X_n|)$$

Pero como

$$Y_i = U_i|X_i|$$

entonces

$$|Y_i| = |X_i|$$

Con esto claramente

$$P(Y_{n+1} \in b|Y_1,..,Y_n) = \theta(F_0(B) + \sum_{i=1}^{n} \delta_{|Y_i|}(B)) + (1-\theta)(F_0(B) + \sum_{i=1}^{n} \delta_{-|Y_i|}(B))$$

Entonces $\{Y_n\}_{n\geq 1}$es una Secuencia Skew Pólya. ■

El siguiente Teorema establece la convergencia de la mezcla convexa de los procesos de Dirichlet con soportes ortogonales a un proceso Skew Dirichlet.

Teorema 3.5. La distribucion condicional (3.7) converge con probabilidad 1 cuando $n \to \infty^+$ a la medida de probabilidad aleatoria

$$F^* = \theta F_0^* + (1-\theta)F_0^*\varphi$$

donde $F_0^* \sim DP(M, F_0)$.

Demostración 3.4. La demostración se basa en el trabajo de Blackwell y MacQueen (1973) quienes demuestran que un DP es una medida de probabilidad aleatoria límite de una Secuencia de Polya. En efecto sea

$$m_n(.) = \theta\frac{\mu_n(.)}{\mu_n(\mathbb{R}^+)} + (1-\theta)\frac{\mu_n(\varphi(.))}{(\mu_o\varphi)_n(\mathbb{R}^-)}$$

como por el trabajo de Blackwell y MacQueen (1973)

$$\frac{\mu_n(.)}{\mu(\mathbb{R}^+)}$$

converge con probabilidad 1 cuando $n \longrightarrow \infty$ a μ^*

donde μ^* sigue una distribución de Ferguson sobre $\mathbb{R}^+$ con parámetro μ y

$$\frac{\mu_o\varphi_n(.)}{(\mu_o\varphi)_n(\mathbb{R}^-)}$$

converge con probabilidad 1 cuando $n \longrightarrow \infty$ a $(\mu_o\varphi)^*$

donde $(\mu_o\varphi)^*$ sigue una distribución de Ferguson concentrada sobre $\mathbb{R}^-$ con parámetro $\mu_o\varphi$.

Se obtiene entonces que consecuentemente $m_n(.)$ converge a $\theta\mu^* + (1-\theta)(\mu_o\varphi)^*$. ■

Este último resultado nos lleva a obtener el siguiente Corolario.

Corolario 3.6. Sea F un DP con parámetros M, F_0 tal que $F_0(0) = 0$ entonces

$$G = \theta F + (1-\theta)F\varphi$$

es un SDP con parámetros M, F_0, θ.

Observación 3.3. Como vimos en el conjunto de resultados anteriores se puede obtener una secuencia Skew Polya a partir de una secuencia de Polya y vice versa. Además el resultado establece la convergencia de una combinación convexa particular de secuencias de Polya una sobre $\mathbb{R}^+$y la otra sobre $\mathbb{R}^-$ a un proceso Skew Polya.

Capítulo 4

Regresión Bayesiana No paramétrica con SDP

4.1. Introducción

La idea de considerar distribuciones que van más allá de la normal para los términos de error en un modelo de regresión dentro de la perspectiva bayesiana, no es reciente, inicialmente se consideraron distribuciones que incluían a la familia normal, pero que preservaban la simetría, como por ejemplo los trabajos pioneros de Zellner (1975) y Box y Tiao (1975), el primero trata con la distribución t-multivariada que preserva la esfericidad de la normal, pero viola la independencia, el segundo trabaja con distribuciones marginales simétricas, en la familia exponencial potencia, preservando el supuesto de independencia. A partir de entonces las extensiones van desde la considereación de mezcla en parámetros de escala de miembros de familias paramétricas multivariadas o univariadas de distribuciones simétricas, como la familia normal, la familia de distribuciones uniformes simétricas en torno al cero y la familia de distribuciones de Laplace.

La consideración de familias de distribuciones construídas a partir de asimetrización de una familia simétrica conteniendo a esta última, no tardó en llegar, uno de los primeros trabajos que considera términos de error independientes con distribución

en una familia de distribuciones asimétricas con moda cero es el de Fernández y Steel (1998) focalizado en el hallazgo de distribuciones a priori no informativas y condiciones que conducen a posterioris propias, Yu y Moyeed (2001) consideran mezclas de distribuciones asimétricas de Laplace y utilizan prioris impropias.

Posteriormente, Sahu y otros (2003) consideran asimetrizaciones de distribuciones simétricas multivariadas como familia de distribuciones para el vector de términos de error cuya representación facilita el uso de Gibbs Sampling , para ver aplicaciones Branco y Dey (2002) , Arellano-Valle y otros (2003), Liseo (2004) y Ferreira y Steel (2004). Una discusión general de distribuciones asimétricas univariadas pueden ser encontradas en Ferreira y Steel (2005).

Uno de los potenciales problemas que se tiene con estas familias es que no necesariamente tienen media o mediana cero, lo cual dificulta la interpretación de los coeficientes de regresión y por tanto la especificación de genuinas distribuciones a priori sobre dichos parámetros. Una posible solución es utilizar prioris no informativas, en especial, si el objetivo es la inferencia de una nueva observación o bien realizar una regresión sobre la mediana donde se restringen a subfamilias de distribuciones simétricas con mediana cero si los coeficientes de regresión son de especial interés o bien en el caso más general realizar una regresión cuantil como ocurre con frecuencia en aplicaciones en finanzas. Algunos trabajos en esta dirección adoptando un enfoque bayesiano paramétrico son los de Yu y Moyeed (2001), que consideran distribuciones de Laplace asimétricas para regresión cuantil fijando el p-ésimo cuantil en cero. Discusiones, aplicaciones y referencias pueden ser encontradas en Yu y otros (2005) y una aproximación a la verosimilitud puede ser encontrada en Dunson y Taylor (2005) basados en el trabajo de Lavin (1995).

Gelfand (1999) provee de una discusión general de aproximaciones semiparamétricas en regresión lineal. West y otros (1994b) consideran regresión lineal con mezclas sobre la medidad basal de procesos de Dirichlet sobre los coeficientes de regresión y las varianzas.

Brunner y Lo (1989), Brunner (1995b) consideran regresión con errores provenientes de una densidad unimodal asimétrica.

Regresión mediana y regresión cuantil también han sido estudiadas desde el punto de vista bayesiano semiparamétrico como en Kottas y Gelfand (2001) y Gelfand y Kottas (2003) donde estudian un modelo de regresión mediana considerando una familia de distribuciones asimétricas de mediana cero usando una mezcla de procesos de Dirichlet con soporte sobre distribuciones con mediana cero y rango intercuartil fijo. Ellos consideran la familia de distribuciones definidas como una media aritmética de dos funciones de densidad no decrecientes, una sobre el eje real positivo y la otra sobre el eje real negativo, las que a su vez admiten representación como mezcla de distribuciones uniformes Feller (1971) y se escoge como medidas en cada mezcla, dos procesos de Dirichlet independientes. Por construcción cada elemento de la familia tienen mediana cero, con mejor desempeño en el enfoque completamente no paramétrico. Los resultados obtenidos son bastante inestables, como lo mencionan los propios autores, de hecho los mismos autores consideran un enfoque semi paramétrico considerando MDP a partir de mezclas en parámetros de escala de una particular familia de distribuciones asimétricas con mediana cero, con mejor desempeño que el enfoque completamente no paramétrico.

Esta propuesta difiere de las anteriores, pues considera Procesos Skew Dirichlet (SDP), con mezcla, no en el parámetro de escala, si no, en parámetros de localización, de modelos de regresión normal. Específicamente se considera que la asimetría puede ser introducida por la presencia de outliers o grupos definidos como saltos en la media, como consecuencia la distribución de la variable respuesta después de integrar el SDP es parametrizada por los coeficientes de regresión y el parámetro de escala tiene distribución continua que es simétrica solo sí el parámetro de asimetría es $1/2$.

Puesto que los Modelos no paramétricos que utilizan como priori un SDP son esencialmente jerárquicos, los parámetros tienen la interpretación que corresponde a un

modelo de regresión normal con diferentes interceptos en la primera etapa. Así la especificación de distribución a priori es hecha condicionalmente a los interceptos. La elección de un SDP para los interceptos no sólo se justifica por su flexibilidad si no también por sus implicaciones sobre la simetría de la distribución marginal de la componente no sistemática en el modelo de regresión.

4.1.1. Análisis de Regresión utilizando mezclas de SDP

Consideraremos un modelo de regresión lineal en el cual utilizaremos mezclas de SDP sobre el parámetro de localización. Especificamente , sean $(y_i, \mathbf{x_i})$ el vector de respuestas y covariables p dimensionales asociadas al iésimo sujeto , $i = 1, 2, ..., n$. Consideremos el siguiente modelo jerárquico

$$\begin{aligned} y_i|\mu_i, \beta, \sigma^2 &\sim N(\mu_i + \beta^T x_i, \sigma^2) \\ \mu_1, \ldots, \mu_n|F &\sim F \\ F|\theta, M, \tau &\sim SDP(\theta, M, F_0(\tau)) \\ \sigma^{-2} &\sim \Gamma(\nu_0, \nu_1) \\ \tau^{-1} &\sim \Gamma(\lambda_0, \lambda_1) \\ \beta &\sim N(0, S) \end{aligned} \tag{4.1}$$

donde $F_0(\tau)$ es la distribución de $|Z|$ con $Z \sim N(0, \tau)$ y hiperparámetros conocidos $(a_1, b_1, \nu_0, \nu_1, \lambda_0, \lambda_0, S)$.

Esta estructura jerárquica es motivada a partir de la consideración que la asimetría es introducida por la presencia de outliers o grupos de puntos con desviaciones similares. Debido a la flexibilidad de los SDP, éstos pueden capturar las anormalidades descritas a través del intercepto μ_i como en el método discutido por Quintana y Iglesias (2003) en el contexto de modelos partición producto.

Más adelante se verificará que la distribución marginal de las respuestas, después de integrar sobre el SDP es caracterizada por los coeficientes de regresión y el parámetro de asimetría θ, y la simetría se obtienen si y sólo si $\theta = 1/2$

Finalmente este modelo jerárquico tiene la ventaja de la interpretabilidad, ya que al modelar no paramétricamente los interceptos μ_i y dados los μ_i y la medida de probabilidad aleatoria F, los coeficientes de regresión y la varianza tienen la interpretación usual.

Observemos que la verosimilitud en el modelo jerárquico especificado puede ser obtenida a partir de las representaciones en términos de densidades del modelo

$$y_i = \beta^T \mathbf{x_i} + \epsilon_i \quad i = 1, 2, ..., n \tag{4.2}$$

donde los ϵ_i son i.i.d. con distribución común $N(0, \sigma^2)$ y son independientes de todos los otros términos del modelo.

La expresión (4.2) puede ser ser reescrita como

$$y_i = \beta^T \mathbf{x_i} + r_i \quad i = 1, 2, ..., n \tag{4.3}$$

donde $r_i = \mu_i + \epsilon_i$ representa el error en esta nueva especificación del modelo.

Adelante se mostrará formalmente que r_i tienen distribución marginal simétrica si y sólo si $\theta = 1/2$.

Para obtener este resultado, en primer lugar se demostrarán un par de proposiciones sobre las que se basará la demostración del teorema principal.

Proposición 4.1. En el modelo (4.1) la distribución condicional de $\mu_i|\sigma^2, \beta$ es simétrica respecto de cero si y sólo si la distribucón condicional de la componente aleatoria $r_i|\sigma^2, \beta$

es simétrica respecto de cero, donde $r_i = \mu_i + \epsilon_i$.

Demostración 4.1. Como $r = \mu + \epsilon$ es la componente aleatoria en el modelo de regresión lineal (4.3), entonces podemos utilizar esta identidad estocástica

$$r|\epsilon, F \stackrel{d}{=} \mu + \epsilon|\epsilon, F$$

y por tanto obtener que

$$f(r|\epsilon) = \int f(r|\epsilon, F)d\mu(F)$$

Entonces

$$f_R(r|\epsilon, F) = f_\mu(r - \epsilon|\epsilon, F),$$

con esto

$$f(r|\epsilon) = \int f_R(-r|\epsilon, F)d\mu(F) = f_R(-r|\epsilon)$$

y

$$f_R(r) = \int f(r|\epsilon)N(0, \sigma^2)d\epsilon = f_R(-r).$$

En el otro sentido, sólo es necesario utilizar la identidad estocástica en sentido inverso con lo que obtenemos

$$\mu|\epsilon, F \stackrel{d}{=} r - \epsilon|\epsilon, F$$

Con esto concluye la prueba ■

Observación 4.1. Observemos que en la proposición anterior se establece simetría de la componente aleatoria en relación directa y biunívoca con la simetría de los interceptos del modelo.

En la siguiente proposición se establece también una relación biunívoca en el sentido que la simetría de la distribución de los interceptos del modelo se obtiene si y sólo si el valor del parámetro θ es $1/2$.

Proposición 4.2. En el modelo de regresión lineal (4.1) $\mu_i|\sigma^2, \beta$ tiene distribución simétrica respecto de cero si y sólo si $\theta = 1/2$.

Demostración 4.2. Es suficiente probar que la distribución de μ_1 es simétrica respecto de cero si y sólo si $\theta = 1/2$, ya que todos los μ_i´s , tienen la misma distribución marginal (permutabilidad). En efecto y de acuerdo a la representación en (3.4) el esquema de urna μ tiene densidad

$$\theta f_o(x) + (1-\theta) f_o(-x)$$

La cual es simétrica si y sólo si $\theta = 1/2$, por lo tanto la distribución marginal de μ_1 es simétrica si y sólo si $\theta = 1/2$. ■

Teorema 4.1. Bajo el modelo (4.1), la distribución de los errores tienen densidad dada por $f(x) = \int N(x;\, \mu, \sigma^2)\, dF(\mu)$, y ella es simétrica si y sólo si $\theta = \frac{1}{2}$.

Demostración 4.3. La componente aleatoria del modelo (4.1) es $r_i = \mu_i + \epsilon_i$ donde μ_i, ϵ_i son condicionalmente independientes dado σ^2, F y $\epsilon_i|\sigma^2 \overset{iid}{\sim} N(0, \sigma^2)$ entonces la densidad de r_i es la convolución de μ_i con ϵ_i y está dada por $f(r|\sigma^2, F) = \int N(r; \alpha, \sigma^2)\, dF(\alpha)$ Además $\mu_i|F \overset{iid}{\sim} F$,entonces $\mu_i's$ dado F son permutables y al aplicar las dos proposiciones anteriores obtenemos que $f_R(r) = f_R(-r)$. ■

Observación 4.2. En el modelo jerárquico es necesario especificar una priori para θ. Para realizar esto es interesante considerar la simetría como un caso especial. Si se considera la siguiente distribución a priori:

$$\theta \sim r_0 \delta_{1/2}(\theta) + (1 - r_0) Beta(\theta; a_0, b_0) \tag{4.4}$$

donde $0 < r_0 < 1$, a_0, y b_0 son números conocidos. Podemos observar que esta mezcla asigna una probabilidad positiva al modelo simétrico y también facilita el cálculo del factor de Bayes que está dado por:

$$\frac{P(\theta = \frac{1}{2}|\text{data})(1 - r_0)}{P(\theta \neq \frac{1}{2}|\text{data})r_0}, \tag{4.5}$$

el cual puede ser evaluado desde cualquier apropiado esquema de simulación.

4.1.2. Aplicación : Retorno de acciones del mercado chileno

Consideremos los datos mensuales de las series de retornos de acciones desde la productora de vino chileno, Concha ỹ Toro.

Es interesante estudiar la relación entre estos retornos de acciones y una variable que describa el comportamiento del mercado. Serán utilizados los valores correspondientes del IPSA.

Consideremos el modelo de regresión $y_i \sim N(\mu_i + \beta^T x_i)$, donde y_i representa el exceso del retorno del activo de la tasa libre de riesgo durante el i-ésimo período , y x_i es el exceso de retorno del activo en el portafolio del mercado en el i-ésimo período. Bajo ciertas condiciones la pendiente en el modelo de regresión, que se entiende como la variación relativa del exceso de retorno en el mercado, su estimación, en el ámbito de las finanzas representa la estimación del *riesgo sistemático.*

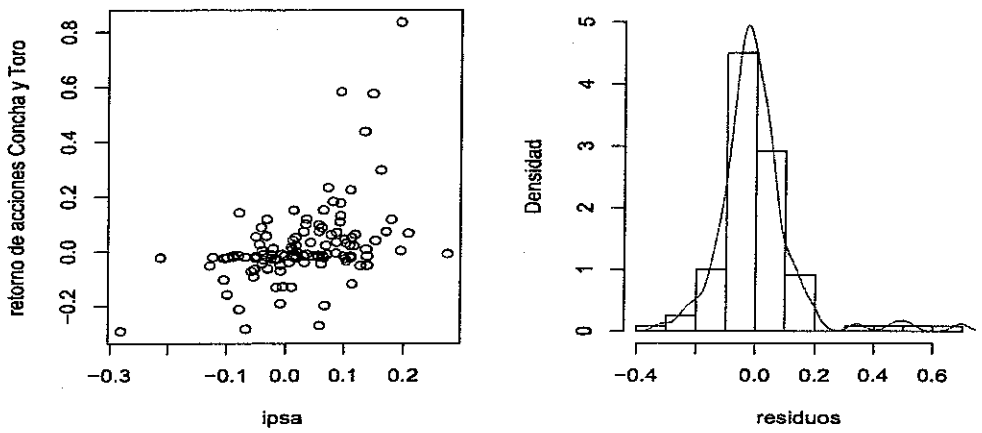

Figura 4.1: *Gráfico de dispersión retorno de acciones de Concha y Toro versus el IPSA.*

La figura (4.1) muestra el gráfico de dispersión de los datos y el histograma de los residuos del modelo de regresión después de haber ajustado la recta de regresión por mínimos cuadrados. El histograma contiene un estimador de kernel de la densidad que provee la función density soportada en el sistema R.

El gráfico sugiere la necesidad de usar distribuciones asimétricas, de hecho el estimador de kernel sugiere la presencia de varias modas menores en la distribución de los errores.

Es conocido que observaciones atípicas pueden tener un severo impacto en la estimación de los coeficientes de regresión. Para reparar tales problemas, Quintana e Iglesias (2003) proponen utilizar modelos partición producto con efectos aleatorios en la forma de interceptos para sujetos específicos. Ahora se reanalizan los datos utilizando el modelo especificado donde la variable exceso de retorno es la respuesta y los valores del IPSA es la variable predictora.

4.1.3. Algunos resultados obtenidos en la modelación

Se ajusta un modelo semiparamétrico, utilizando $M = 1$ (en el esquema MCMC se puede incorporar conocimiento inicial a cerca de M, ver Escobar and West (1995)) y prioris vagas para los otros parámetros. En primer lugar, en un análisis exploratorio se utiliza como priori $p(\theta) = Beta(\theta; a_0, b_0)$, donde se obtienen las distribuciones a posteriori mostradas en la figura (4.2).

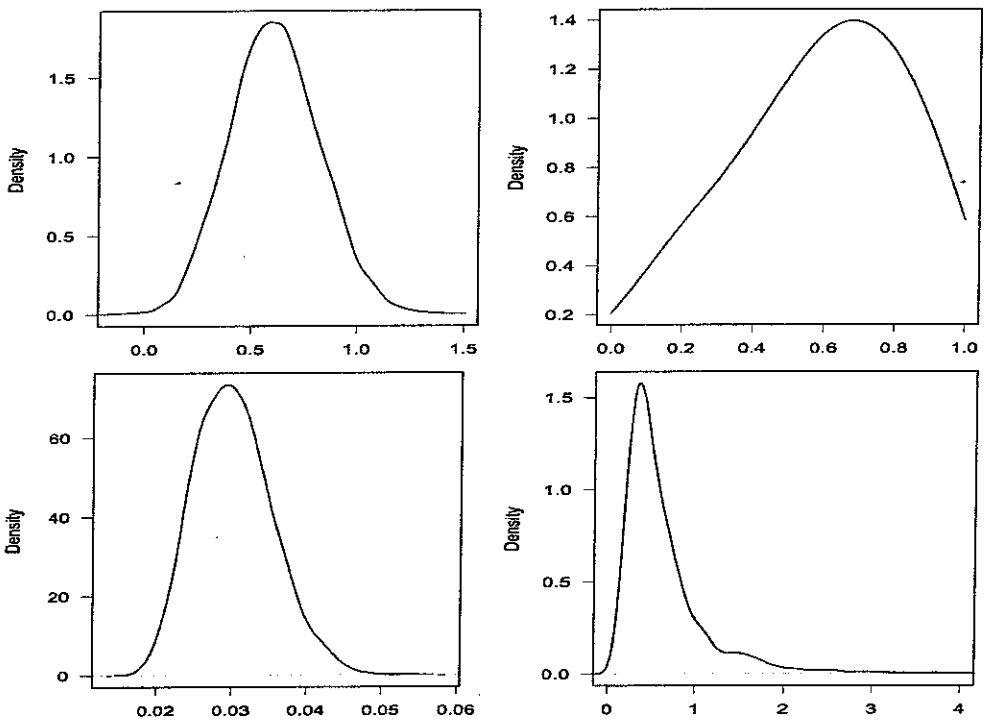

Figura 4.2: *Distribuciones a posteriori de los parámetros del modelo*

En la figura (4.2) el gráfico superior izquierdo muestra la densidad a posteriori marginal de β. El gráfico superior derecho muestra la posteriori marginal condicional $p(\theta|\theta \neq 1/2, y)$.Los gráficos inferiores de izquierda a derecha muestran las densidades posterioris marginales de σ^2 y τ.

La posteriori $p(\theta|\text{data})$ sugiere alta asimetría en el modelo. Se repite el análisis utilizando esta vez (4.4) con $r_0 = 0{,}5$ dando el factor de Bayes calculado en (4.5) tuvo un valor de 0,033, confirmando la afirmación anterior. Para entender como el modelo ajusta los puntos outliers consideramos las regiones de confiabilidad a posteriori del 95 % para los residuos de la regresión, los que se muestran en la figura (4.3).

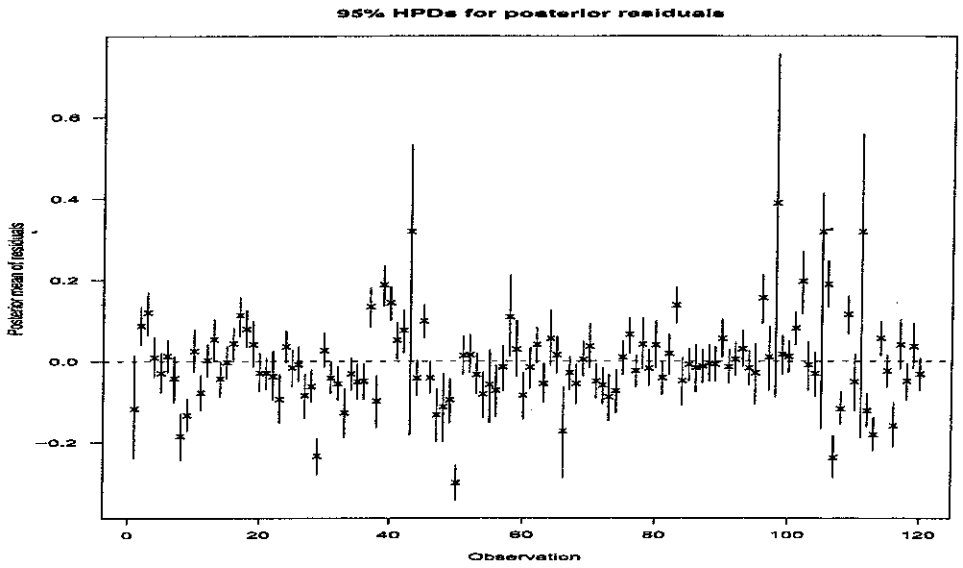

Figura 4.3: *HPD's de 95 % para los residuos de la regresión.*

La mayoría de los intervalos de credibilidad incluyen el 0, pero algunos de ellos no. Esto puede ser interpretado como un indicador de observaciones atípicas. Podemos ordenar los valores absolutos de los residuos y considerar los tres más pequeños, los tres centrales y los tres mayores. Sus correspondientes densidades a posteriori se muestran en la figura (4.4).

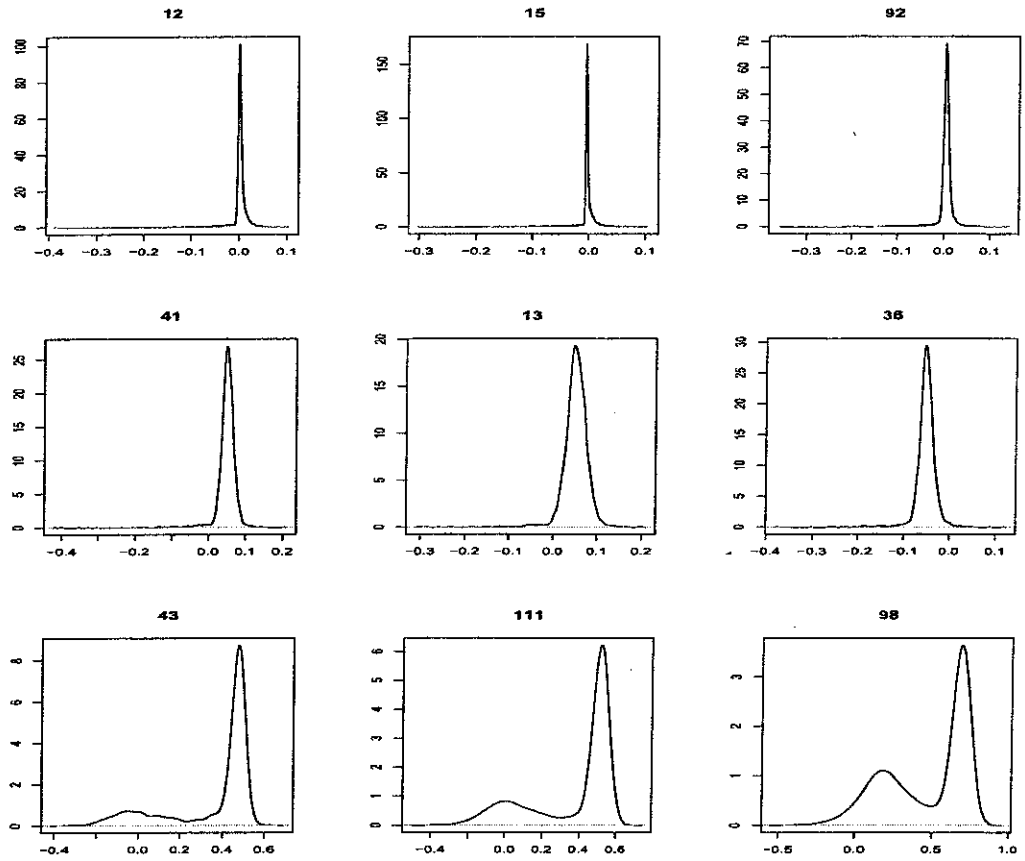

Figura 4.4: *Distribución a posteriori para algunos residuos en el modelo.*

No sorprende ver en la figura (4.4) el hecho que las posterioris en la primera fila están altamente concentradas, menos concentradas en la fila central y la última fila presenta bimodalidad, revelando que un modelo paramétrico podría ser inapropiado para ajustar la distribución del error.

Tabla 4.1: Algunos resúmenes de las posterioris para el ejemplo de regresión.

Parameter	Posterior Mean	95 % HPD
M	0,929 (0,0268)	(0,209, 1,830)
β	0,611 (0,0056)	(0,174, 0,990)
σ^2	0,030 (0,0002)	(0,020, 0,040)
τ	0,652 (0,0222)	(0,102, 1,654)
θ	0,574 (0,0162)	(0,164, 0,999)

En la tabla (4.1) los errores estándar de Monte Carlo son puestos entre paréntesis.

Capítulo 5

Estimación de una densidad

5.1. Introducción

Combinar un kernel Gaussiano con un DP resulta uno de los procedimientos más populares de estimación de la densidad desde la perspectiva Bayesiana no paramétrica. Tales modelos fueron introducidos por Lo (1984). Otros estudios en este mismo enfoque se pueden encontrar en los trabajos de Ferguson (1983), Escobar y West (1995b), Gasparini (1996). Una discusión del tema se puede ver en el trabajo de Roeder y Wasserman (1997).

Debido a la naturaleza discreta de los Procesos de Dirichlet en el problema de estimación de una densidad es natural pensar en los modelos de mezcla, donde los Procesos de Dirichlet juegan un rol fundamental. De hecho se realiza una convolución de un kernel de una distribución continua con una medida de probabilidad aleatoria que sigue un proceso de Dirichlet. Esta convolución se puede representar a través del modelo.

$$\begin{aligned} Y_i|\nu_i &\sim f(.|\nu_i), i = 1, 2, ..., n \\ \nu_1, \nu_2, ..., \nu_n|F &\overset{iid}{\sim} F \\ F &\sim DP(M, F_o) \end{aligned}$$

En este modelo se mezclan densidades continuas dependientes de algún parámetro ν, cuya distribución F es desconocida y por tanto aleatoria a la que se da como priori un proceso de Dirichlet. En particular se puede considerar un Kernel Normal y en este caso el parámetro ν corresponde al vector de media y varianza de la Normal. Estos procesos son conocidos por el nombre de *Dirichlet Mixtures, (MDP).* Algunas aplicaciones las podemos encontrar en Escobar (1994), West y otros (1994a), Escobar y West (1995a), Escobar y West (1995b), Müller y otros (1996), Bush (1993) y MacEachern (1998). Es importante destacar que la clausura (topología débil) de estas densidades es $\mathfrak{M}(\mathbb{R})$.

En el problema de estimación de una densidad lo natural es preguntarse por la consistencia de la posteriori, es aquí donde Barron y otros (1999) provee de condiciones suficientes para dicha consistencia y en el capítulo 4 del libro de Ghosh y Ramamoorthi (2003) se muestra el *teorema de Schwartz* como una herramienta poderosa para establecer consistencia de la posteriori.

Lijoi y otros (2005) muestran que una MDP es fuertemente consistente si la medida de base del proceso de Dirichlet tiene media finita y la priori de la desviación estándar tiene un decaimiento exponencial en una vecindad del cero.

Una demostración de la consistencia fuerte de la posteriori en el modelo MDP, con kernel normal con una priori gama invertida para la varianza puede ser encontrado en Ghosh y Ramamoorthi (2003).

5.1.1. Modelo de mezclas de normales con DP

Los MDP pueden ser representados a través de una medida de probabilidad aleatoria que es obtenida por una convolución de un kernel gaussiano con una medida de probabilidad aleatoria proveniente de un DP.

El modelo es de la forma

$$\begin{aligned} F(x) &= \int f(x|\theta)G(d\theta) \\ G &\sim DP(M, G_o) \end{aligned} \tag{5.1}$$

donde el kernel es $f(x|\theta) = \phi_{\mu,S} \propto exp[-(x-\mu)^T S^{-1}(x-\mu)/2]$, y se mezcla respecto al parámetro $\theta = (\mu, S)$.

La inferencia a posteriori en el modelo (5.1) es basado en simulaciones MCMC. Algunas Simulaciones MCMC eficientes para este modelo son discutidas entre otros en Bush (1993), West y otros (1994a), Müller y MacEachern (1998) y Neal (2000).

El modelo de mezcla de normales básico que describe Ferguson (1983) es como sigue:

$$\begin{aligned} Y_1, Y_2, ..., Y_n|\mu_i, V_i, G &\overset{ind}{\sim} N(\mu_i, V_i) \\ (\mu_i, V_i)|G &\overset{iid}{\sim} G \\ G &\sim DP(M, G_o) \end{aligned} \tag{5.2}$$

En este modelo (μ_i, V_i) son los parámetros correspondiente a la media y varianza de distribuciones normales que determinan los parámetros bidimensionales $\pi_i = (\mu_i, V_i)$ que provienen de una distribución G sobre $\mathbb{R}x\mathbb{R}^+$, donde esta distribución es desconocida y por tanto aleatoria, la que es modelada como un proceso de Dirichlet, definida por el parámetro de precisión M y la medida basal G_o que es la esperanza a priori de G.

El objetivo ahora es plantear un modelo similar a (5.2), pero utilizando como priori un proceso SDP, con lo que se pretende resolver el problema de estimación de una densidad utilizando un mezcla de kernel tradicional en el espíritu descrito en Escobar y West (1995b).

5.1.2. Estimación de una densidad utilizando SDP

Como una ilustración de la capacidad de modelación basada en SDP, se adaptará el modelo en Escobar y West (1995b) y presentado en (5.2). Haciendo esto tenemos la flexibilidad que provee un modelo Semiparamétrico pero con la ventaja extra de un procedimiento explícito para controlar la simetría. Específicamente consideremos el siguiente modelo jerárquico:

$$y_i \mid \mu_i, V_i \sim N(\mu_i, V_i) \tag{5.3}$$

$$(\mu_1, V_1), \ldots, (\mu_n, V_n) \mid F \sim F \tag{5.4}$$

con

$$F(\cdot) = \sum_{h=1}^{\infty} w_h \delta_{(m+\theta A_h, B_h)} + \sum_{h=1}^{\infty} w_h \delta_{(m-(1-\theta)A_h, B_h)}, \tag{5.5}$$

donde los ponderadores $\{w_h\}$ provienen desde un proceso "stick-breaking " con parámetro M, y donde $\{(A_h, B_h)\}$ son muestreados desde $F_0(A, B)$ tal que

$$A \mid B \sim |N(0, \tau B)|, \qquad B^{-1} \sim \Gamma(a_1, b_1),$$

i.e., $A \mid B$ tinen la misma distribución que el valor absoluto de la variable $\mathcal{A} \sim N(0, \tau B)$. Asumamos también que

$$\tau^{-1} \sim \Gamma(\lambda_0, \lambda_1), \quad m \sim N(m_0, S), \quad M \sim \Gamma(a, b),$$

con hiperparámetros conocidos $(a_1, b_1, a, b, \lambda_0, \lambda_1, m_0, S)$.

Para el parámetro de asimetría se asume la siguiente a priori:

$$\theta \sim (1-\pi)B(a_0, b_0) + \pi\delta_{1/2}(\theta), \tag{5.6}$$

para valores conocidos de (π, a_0, b_0). Esta distribución corresponden a una mezcla entre una masa puntual en $1/2$ y una distribución Beta, con ponderadores definidos por π. Esta priori permite tener una evaluación explícita de la simetría, ya que la correspondiente posteriori $p(\theta \mid y_1, \ldots, y_n)$ será otra vez una mezcla entre una distribución continua y una masa puntual en $1/2$.

El modelo anterior pone una medida de probabilidad aleatoria a priori sobre una distribución desconocida F sobre $\mathbb{R} \times \mathbb{R}^+$ que al proyectarla en la coordenada de las localizaciones mantiene la estructura de un SDP sobre $\mathbb{R}$, este proceso consta de tres partes: un centro m común, localizaciones individuales para todos los componentes de la mezcla, provenientes de la representación SDP (3.5)y expresada como $(m+\theta A_h, m-(1-\theta)A_h)$, y las varianzas correspondientes B_h, $h \geq 1$. El modelo marginal resultante se parece al procedimiento de estimación de una densidad utilizando un kernel tradicional en el espíritu descrito en Escobar y West (1995b), pero basados en los SDPs.

Se ajusta el modelo especificado anteriormente al conjunto de datos "galaxia " descrito en Roeder (1992) y que han sido analizados en varios trabajos, incluyendo los trabajos de Escobar y West (1995b), Richardson y Green (1997), Stephens (2000), and Ishwaran y James (2003). Los datos consisten de $n = 82$ mediciones de velocidades relativas de otras galaxias a nuestra galaxia (x 10^3 km/s) desde seis secciones cónicas del espacio. Siguiendo la idea de Escobar y West (1995b), los valores de los hiperparámetros fueron tomados como $m_0 = 20$, $S = 10^5$, $\lambda_0 = 1/2$, $\lambda_1 = 50$, $a = 2$, $b = 4$, $a_1 = 2$, $b_1 = 1$. Además se elige $a_0 = b_0 = 1$ y $\pi = 0{,}2$. La figura (5.1) muestra la distribución predictiva a posteriori $p(y_{n+1} \mid y_1, \ldots, y_n)$, i.e., la estimación Bayesiana no paramétrica de la densidad, además el histograma de los datos. Es claro que la estimación captura

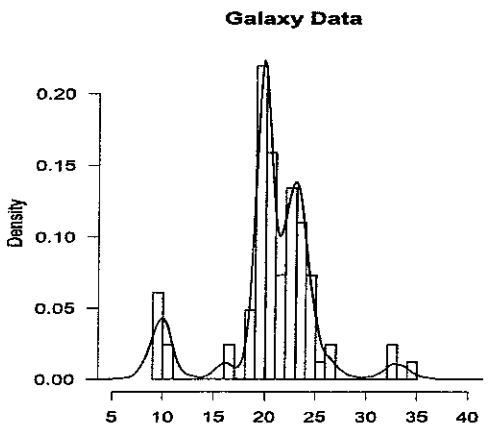

Figura 5.1: *Histograma y Densidad Estimada* .

bastante bien las características principales de los datos, poniendo componentes mezclas en todos los lugares pertinentes.

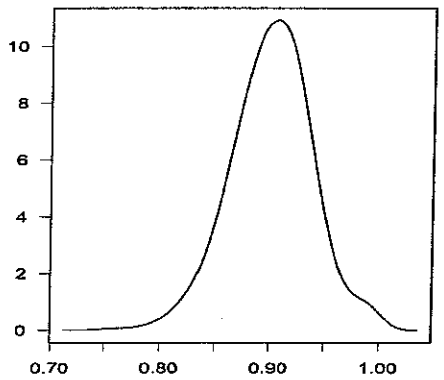

Figura 5.2: *densidad a posteriori condicional* $p(\theta \mid y_1, \ldots, y_n, \theta \neq 1/2)$.

La figura (5.2) muestra la densidad a posteriori condicional $p(\theta \mid y_1, \ldots, y_n, \theta \neq 1/2)$, i.e., la parte continua de la posteriori marginal completa $p(\theta \mid y_1, \ldots, y_n)$.

Es interesante indicar que el valor estimado de $p(\theta = 1/2 \mid y_1, \ldots, y_n)$ fue esencialmente 0, el cual fue el resultado al aumentar π a 0.5. Esto significa realmente un fuerte soporte para comparar asimetría con simetría.

La media posteriori marginal de m fue 10,91 (error estándar de Monte Carlo 0,052) y 95 % HPD (9,58, 12,00), la posteriori coloca esencialmente toda la masa a la derecha de 10,91, lo cual explica como $p(\theta \mid y_1, \ldots, y_n)$ está concentrado cerca de 1.

Observación 5.1. Es importante destacar que en estimación de densidades el modelo utilizado ajusta bastante bien las características principales de los datos, sería interesante poder comparar estos ajustes con los realizados por Escobar y West (1995b). En este aspecto se espera que no existan mayores diferencias, teniendo siempre en cuenta que el aporte del SDP en este aspecto es poder controlar asimetría con un parámetro, a diferencia de los DP que son asimetricos por construcción.

Capítulo 6

Notas Finales

6.1. Conclusiones

El objetivo básico de esta tesis fue construir y definir a partir de un proceso de Dirichlet usual sobre $\mathbb{R}^+$ una familia de medidas de probabilidad aleatoria que incorporan un parámetro de asimetría de tal forma de obtener medidas de probabilidad aleatorias simetrícas como un caso particular. De tal manera que estas nueva familia de medidas de probabilidad aleatoria actúen como prioris no paramétricas para la inferencia bayesiana en modelos regresión, estimación de la densidad y otros.

El parámetro de asimetría es tal que con probabilidad 1 el soporte del proceso son las distribuciones simétricas si y sólo si su valor es 1/2. En la incorporación de este parámetro es interesante obtener la simetría como un caso particular ya que en algunos modelos comparar ambos escenarios tiene importancia.

Como sabemos los DP y procesos de Dirichlet invariantes presentan construcciones que contienen clases de distribuciones simétricas o asimétricas, excluyentemente. Los SDP cumplen propiedades como poseer la flexibilidad de la modelación no paramétrica así como también tienen la posibilidad de incorporar la simetría como un caso particular.

En el desarrollo de esta tesis entre otras cosas se derivaron algunas útiles represen-

taciones de los SDP como un esquema de Urna y una representación tipo Sethuraman, importantes ambas, para observar la discretización del proceso, la concentración de él alrededor de distribuciones simétricas o asimétricas (según sea el valor de θ) y en la obtención de las posteriores condicionales para el muestreo en la simulación de estos procesos.

Este trabajo abre la posibilidad de aplicar los SDP en otros modelos estadísticos teniendo la particularidad que lo diferencia de un proceso de Dirichlet y un proceso de Dirichlet invariante, ya que en su soporte se encuentran tanto distribuciones simétricas como asimétricas, unimodales y multimodales, determinadas por una distribución basal , un parámetro de precisión y un parámetro de asimetría.

Es importante destacar que a pesar de que estos procesos no son conjugados, está condición no es limitante pues la obtención de la posteriori se realiza a través del muestreo de Gibbs (MCMC). Además el manejo computacional del proceso es heredado de los procesos de Dirichlet, en el sentido de que ellos pueden ser adecuados agregando un paso para el parámetro de asimetría, dentro de la implementación computacional.

Se aplicó el proceso SDP para modelar el problema de estimación del riesgo sistemático del exceso de retorno desde una serie de indices de valores de una empresa vinera chilena. El contexto específico fue un modelo de regresión lineal con presencia de asimetría y outliers. La discreticidad del proceso fue controlada considerando mezclas con un kernel continuo en el modelo de Regresión. Un factor de Bayes para evaluar simetría pudo ser obtenido desde el output de la simulación a posteriori. El modelo ajustado captura ambos problemas, presencia de asimetría y outliers, en forma simultánea.

La otra aplicación consistió en aplicar el proceso en el problema de estimación de una densidad. La estimación captura bastante bien las características principales de los datos, poniendo componentes mezclas en todos los lugares pertinentes, con resultados satisfactorios.

6.2. Trabajo Futuro

Queda abierta la posibilidad de estudiar la posible aplicación de los SDP a problemas como modelos con errores en las variables y modelos con efectos aleatorios. En este último modelo es interesante el trabajo realizado por Burr y Doss (2005) en Meta análisis donde se utiliza el modelo

$$\begin{array}{lll} \text{Condicional en } \psi_i, & D_i \overset{iid}{\sim} N(\psi_i, \sigma_i^2) & i = 1, 2, .., m \\ \text{Condicional en } F, & \psi_i \overset{iid}{\sim} F & i = 1, 2, .., m \\ \text{Condicional en } \mu, \tau & F \sim D(M, N(\mu, \tau^2)) & \\ \text{Condicional en } \tau & \mu \sim N(c, d\tau^2) & \\ & 1/\tau^2 \sim Gama(a, b) & \end{array}$$

donde D_i representa los datos , ψ_i representa el efecto del i-ésimo estudio,(en estudios médicos puede ser la diferencia entre dos probabilidades binomiales) F es una medida de probabilidad aleatoria construida a partir de una mezcla de procesos Dirichlet condicionales, que se caracterizan por ser asimétricos con probabilidad 1 y cuya mediana es μ.

Este modelo es justificado por los autores basándose en el rol que juega μ que es bien definido y el cual es de fácil estimación pues se obtiene del output del muestreo de Gibbs y pues su distribución a posteriori no es influenciada por el hecho de tener pocas observaciones.

Para el proceso SDP, y en el espíritu del trabajo de Burr y Doss (2005) , el parámetro μ representaría el percentil $(1 - \theta)$ de la distribución F, para $\theta = 1/2$ correspondería a la mediana de F, lo cual ampliaría , en cierto sentido, la familia de procesos y las distribuciones desde las cuales se obtienen los ψ_i además de que μ mantiene un rol definido. ■

Apéndice A

Modelo de Regresión

A.1. Algoritmo de estimación en el Modelo de Regresión SDP

El modelo de regresión especificado en el capítulo 5 era el siguiente

$$\begin{aligned} y_i|\mu_i,\beta,\sigma^2 &\sim N(\mu_i + x_i\beta, \sigma^2) \\ \mu_1,.....,\mu_n|F &\sim F \\ F|\theta,M,\tau &\sim SDP(\theta,M,F_o(\tau)) \\ F_o(\tau) &= |Z|conZ \sim N(0,\tau) \\ \beta|\sigma^2 &\sim N(\beta_o,\sigma^2|\kappa_o) \\ \sigma^2 &\sim GI(\upsilon_o,\upsilon_1) \\ \theta &\sim Beta(a_o,b_o) \\ \tau &\sim GI(\lambda_o,\lambda_1) \end{aligned} \tag{A.1}$$

con hiperparámetros conocidos $(a_o, b_o, M, \lambda_o, \lambda_1, \beta_o, \kappa_o, \upsilon_o, \upsilon_1)$

Para desarrollar el algoritmo de estimación a través de Muestreo de Gibbs es necesario determinar todas las distribuciones condicionales de los parámetros del modelo , además de considerar la útil construcción del SDP a través del esquema de urna Skew Polya que especifica lo siguiente:

- $P(\mu_1 \in B) = v(B)$
- $P(\mu_{k+1} \in B | \mu_1, ..., \mu_k) = \frac{v_k(B)}{v_k(\mathbb{R})}$

donde

- $v(B) = F_o(\frac{B}{\theta}) + F_o(\frac{-B}{(1-\theta)})$
- $v_k(B) = Mv(B) + \sum_{i=1}^{k} \delta_{\theta|\mu_i|}(B) + \sum_{i=1}^{k} \delta_{(1-\theta)|\mu_i|}(-B)$
- $v_k(\mathbb{R}) = 2(M+k)$

Desde la especificación del esquema de urna anterior se obtiene que la distribución marginal de los parámetros $\mu_1, \mu_2, ..., \mu_n$ está dada por la siguiente expresión:

$$f_o(\frac{\mu_i}{\theta}) I_{\{\mu_i \geq 0\}} + f_o(\frac{-\mu_i}{(1-\theta)}) I_{\{\mu_i < 0\}} \tag{A.2}$$

Además de las expresiones anteriores es necesario determinar la distribución condicional de cada uno de los parámetros que intervienen en el modelo a modo de aplicar Muestreo de Gibbs para la estimación de la posteriori.

A.1.1. Distribuciones Condicionales de los parámetros

$$
\begin{aligned}
p(\mu_i|\beta,\sigma^2,y_i,\theta,\tau) &\propto p(y_i|\beta,\sigma^2,\mu_i)p(\mu_i|\theta,\tau) \\
&\propto exp\{-\frac{1}{2\sigma^2}(y_i-\mu_i-\beta x_i)^2\}exp\{\frac{1}{2\tau}(\frac{\mu_i}{\theta})^2\}I\{\mu_i\geq 0\} \\
&+ exp\{\frac{1}{2\tau}(\frac{\mu_i}{1-\theta})^2\}I\{\mu_i<0\} \\
&= exp\{-\frac{1}{2}(\frac{\sigma^2+\tau\theta^2}{\sigma^2\tau\theta^2})(\mu_i-(\frac{\tau\theta^2}{\sigma^2+\tau\theta^2})m_i)^2\} \\
&exp\{-\frac{{m_i}^2}{2(\sigma^2+\tau\theta^2)}\}I\{\mu_i\geq 0\} \\
&+ exp\{-\frac{1}{2}(\frac{\sigma^2+\tau(1-\theta)^2}{\sigma^2\tau(1-\theta)^2})(\mu_i-(\frac{\tau(1-\theta)^2}{\sigma^2+\tau(1-\theta)^2})m_i)^2\} \\
&exp\{-\frac{{m_i}^2}{2(\sigma^2+\tau(1-\theta)^2)}\}I\{\mu_i<0\}
\end{aligned}
$$

con $m_i=\beta x_i-y_i$

La expresión anterior muestra que la distribución condicional de los μ_i dado los parámetros$(\beta,\sigma^2,y_i,\theta,\tau)$ corresponde a una mezcla de normales truncadas en a la izquierda y derecha de cero.

Según la expresión A.2 la distribución condicinal de μ_i dado μ_{-i} con la notación convencional $\mu_{-i}=(\mu_1,\mu_2,...,\mu_{i-1},\mu_{i+1},...,\mu_n)$ es de la forma

$$p(\mu_i|\mu_{-i})=\frac{1}{2(M+n)}\{M[f_o(\frac{\mu_i}{\theta})+f_o(\frac{-\mu_i}{1-\theta})]+\sum_{j\neq i}\delta_{\theta|\mu_j|}(\mu_i)+\sum_{j\neq i}\delta_{(1-\theta)|\mu_j|}(-\mu_i)\} \quad \text{(A.3)}$$

donde

$$f_o(\frac{\mu_i}{\theta})=\frac{1}{\sqrt{2\pi\tau}}\exp\{-\frac{1}{2\tau}\frac{{\mu_i}^2}{\theta^2}\}I_{\{\mu_i\geq 0\}}$$

$$f_o(\frac{-\mu_i}{1-\theta})=\frac{1}{\sqrt{2\pi\tau}}\exp\{-\frac{1}{2\tau}\frac{{\mu_i}^2}{(1-\theta)^2}\}I_{\{\mu_i<0\}}$$

Considerando la expresión A.3 obtenemos la siguiente distribución condicional que se entiende como que el muestreo se realiza desde normales truncadas con probabilidad proporcional a los coeficientes calculados o bien se remuestrea utilizando los valores actuales.

$$
\begin{aligned}
p(\mu_i|\mu_{-i},\beta,\sigma^2,y_i,\theta,\tau) &\propto p(y_i|\mu_i,\beta,\sigma^2)p(\mu_i|\mu_{-i}) \\
&\propto exp\{-\frac{1}{2\sigma^2}(y_i-\mu_i-\beta x_i)^2\}\{\frac{M}{\sqrt{2\pi\tau}}[exp\{\frac{1}{2\tau}\frac{{\mu_i}^2}{\theta^2}\}I\{\mu_i\geq 0\} \\
&+ exp\{\frac{1}{2\tau}\frac{{\mu_i}^2}{(1-\theta)^2}\}I\{\mu_i<0\}]\} \\
&+\sum_{j\neq i}\delta_{\theta|\mu_j|}(\mu_i)+\sum_{j\neq i}\delta_{\theta|\mu_j|}(-\mu_i) \\
&=\frac{M}{\sqrt{2\pi\tau}}exp\{-\frac{1}{2}(\frac{\sigma^2+\tau\theta^2}{\sigma^2\tau\theta^2})(\mu_i-(\frac{\tau\theta^2}{\sigma^2+\tau\theta^2})m_i)^2\} \\
&exp\{-\frac{{m_i}^2}{2(\sigma^2+\tau\theta^2)}\}I\{\mu_i\geq 0\} \\
&+exp\{-\frac{1}{2}(\frac{\sigma^2+\tau(1-\theta)^2}{\sigma^2\tau(1-\theta)^2})(\mu_i-(\frac{\tau(1-\theta)^2}{\sigma^2+\tau(1-\theta)^2})m_i)^2\} \\
&exp\{-\frac{{m_i}^2}{2(\sigma^2+\tau(1-\theta)^2)}\}I\{\mu_i<0\} \\
&+exp\{-\frac{1}{2\sigma^2}(\mu_i-m_i)^2\}\sum_{i\neq j}^{l}\delta_{\theta|\mu_j|}(\mu_i)I\{\mu_i\geq 0\} \\
&+exp\{-\frac{1}{2\sigma^2}(\mu_i-m_i)^2\}\sum_{j\neq i}\delta_{(1-\theta)|\mu_j|}(-\mu_i)I\{\mu_i<0\}
\end{aligned}
$$

Según las prioris especificadas en A.1 obtenemos lo siguiente

$$p(\beta,\sigma^2|\mu,y_1,..,y_n) \propto p(y_1,...,y_n|\beta,\sigma^2,\mu_1,...,\mu_n)p(\beta|\sigma^2)p(\sigma^2)$$

$$\propto(\sigma^2)^{-n/2}exp\{-\frac{1}{2\sigma^2}(\sum_{i=1}^{n}(y_i-\mu_i)-\beta x_i)^2\}$$

$$(\sigma^2)^{-1/2}exp\{-\frac{\kappa_o}{2\sigma^2}(\beta-\beta_o)^2\}exp\{-\upsilon_o/\sigma^2\}(\sigma^2)^{-\upsilon_1-1}$$

$$\propto(\sigma^2)^{-n/2-1/2+\upsilon_1+1}exp\{-\frac{1}{2\sigma^2}(\sum_{i=1}^{n}x_i^2+\kappa_o)[\beta-\frac{\sum_{i=1}^{n}x_i(y_i-\mu_i)+\kappa_o\beta_o}{\sum_{i=1}^{n}x_i^2+\kappa_o}]^2\}$$

$$exp\{-\frac{1}{2\sigma^2}[\frac{(\sum_{i=1}^{n}x_i(y_i-\mu_i)+\kappa_o\beta_o)^2}{\sum_{i=1}^{n}x_i^2+\kappa_o}-\sum_{i=1}^{n}(y_i-\mu_i)^2-\kappa_o\beta_o^2-2\upsilon_o]\}$$

Integrando respecto de β , determinamos la distribución condicional de σ^2 dado $\mu = (\mu_1,...,\mu_n)$ e $\mathbf{y} = (y_i,...,y_n)$ como sigue

$$p(\sigma^2|\mu,\mathbf{y}) = \int p(\beta,\sigma^2|\mu_1,...,\mu_n,y_1,..,y_n)d\beta$$

$$\propto(\sigma^2)^{n/2+\upsilon_1+1}exp\{-\frac{1}{2\sigma^2}[\frac{(\sum_{i=1}^{n}x_i(y_i-\mu_i)+\kappa_o\beta_o)^2}{\sum_{i=1}^{n}x_i^2+\kappa_o}$$

$$-\sum_{i=1}^{n}(y_i-\mu_i)^2-\kappa_o\beta_o^2-2\upsilon_o]\}$$

esta expresión corresponde a una distribución Gamma inversa con los siguientes parámetros.

$$\sigma^2|\mu,\mathbf{y} \sim GI(\frac{\sum x_i^2\sum(y_i^*)^2-(\sum x_iy_i^*)^2+\kappa_o\beta_o^2(\sum(x_i-y_i/\beta_o)^2)+2\upsilon_o(\kappa_o+\sum x_i^2)}{2(\sum x_i^2+\kappa_o)},\frac{n}{2}+\upsilon_1)$$

donde

$$y_i^* = y_i-\mu_i$$

Además desde las prioris especificadas para β y σ^2 en el modelo de regresión A.1 tenemos que

$$\begin{aligned} p(\beta|\mu,\mathbf{y},\sigma^2) \propto & exp\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i^* - x_i\beta)^2\}exp\{-\frac{\kappa_o}{2\sigma^2}(\beta-\beta_o)^2\} \\ \propto & exp\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i^2+\kappa_o)(\beta - \frac{\sum_{i=1}^{n}x_iy_i^* + \kappa_o\beta_o}{\sum_{i=1}^{n}x_i^2+\kappa_o})^2\} \end{aligned} \tag{A.4}$$

Desde acá se obtiene que la distribución condicional de β dado$\mu_1, ..., \mu_n, y_1, .., y_n, \sigma^2$ es

$$\beta|\mu_1, ..., \mu_n, y_1, .., y_n, \sigma^2 \sim N(\beta_o^*, \sigma^{2*})$$

donde

$$y_i{}^* = y_i - \mu_i$$

$$\beta_o^* = \frac{\sum x_i y_i^* + \kappa_o\beta}{\Sigma x_i^2 + \kappa_o}$$

$$\sigma^{2*} = \frac{\sigma^2}{\Sigma x_i^2 + \kappa_o}$$

Sólo resta determinar la distribución condicional de τ y θ.

$$\begin{aligned} p(\tau|\mu_1,...,\mu_n,\theta) &\propto p(\mu_1,...,\mu_n|\theta,\tau)p(\tau) \\ &\propto \tau^{-n/2}\prod_{i=1}^{n}\{exp(-\frac{1}{2\tau\theta^2}\mu_i^2)I\{\mu_i \geq 0\} + exp(-\frac{1}{2\tau(1-\theta)^2}\mu_i^2)I\{\mu_i < 0\}\} \\ &\tau^{-\lambda_1-1}exp(-\lambda_o/\tau) \\ &\propto \tau^{-n/2-\lambda_1-1}exp\{-\frac{1}{\tau}(\frac{1}{2\theta^2}\sum\mu_i^2 I_{\{\mu_i\geq 0\}} + \frac{1}{2(1-\theta)^2}\sum\mu_i^2 I_{\{\mu_i<0\}} + \lambda_o)\} \end{aligned}$$

Por lo tanto

$$\tau|\theta,\mu_1,...,\mu_n \sim GI(\frac{1}{2\theta^2}\sum_{i=1}^{n}\mu_i^2 I\{\mu_i \geq 0\} + \frac{1}{2(1-\theta)^2}\sum_{i=1}^{n}\mu_i^2 I\{\mu_i < 0\} + \lambda_o, \frac{n}{2} + \lambda_1)$$

Y finalmente para θ nos queda

$$p(\theta|\mu_1,...,\mu_n,\tau) \propto p(\mu_1,...,\mu_n|\theta,\tau)p(\theta)$$
$$\propto exp\{-\frac{1}{2\tau\theta^2}\sum \mu_i^2 I\{\mu_i \geq 0\} - \frac{1}{2\tau(1-\theta)^2}\sum \mu_i^2 I\{\mu_i < 0\}\}\theta^{a_o-1}(1-\theta)^{b_o-1}$$

Bibliografía

Antoniak, C. E. (1974).

Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems.

Ann. Statist., 2:1152–1174.

Arellano-Valle, R. B., Galea-Rojas, M., and Zuazola, P. I. (2000).

Bayesian sensitivity analysis in elliptical linear regression models.

J. Statist. Plann. Inference, 86(1):175–199.

Arellano-Valle, R. B. and Genton, M. G. (2005).

On fundamental skew distributions.

J. Multivariate Anal., 96(1):93–116.

Arellano-Valle, R. B., Gómez, H. W., and Quintana, F. A. (2005).

Statistical inference for a general class of asymmetric distributions.

J. Statist. Plann. Inference, 128(2):427–443.

Arellano-Valle, R. B., Iglesias, P. L., and Vidal, I. (2003).

Bayesian inference for elliptical linear models: conjugate analysis and model comparison.

In *Bayesian statistics, 7 (Tenerife, 2002)*, pages 3–24. Oxford Univ. Press, New York.

With discussions by Judith Rousseau and Manuel Mendoza and a reply by the authors.

Azzalini, A. (1986).

Further results on a class of distributions which includes the normal ones.

Statistica (Bologna), 46(2):199–208.

Barron, A., Schervish, M. J., and Wasserman, L. (1999).

The consistency of posterior distributions in nonparametric problems.

Ann. Statist., 27(2):536–561.

Blackwell, D. and MacQueen, J. B. (1973).

Ferguson distributions via Pólya urn schemes.

Ann. Statist., 1:353–355.

Box, G. E. P. and Tiao, G. C. (1975).

Intervention analysis with applications to economic and environmental problems.

J. Amer. Statist. Assoc., 70:70–79.

Branco, M. D. and Dey, D. K. (2002).

Regression model under skew elliptical error distribution.

J. Math. Sci. (N.S.) (Delhi), 1:151–168.

Brunner, L. J. (1992).

Bayesian nonparametric methods for data from a unimodal density.

Statist. Probab. Lett., 14(3):195–199.

Brunner, L. J. (1995a).

Bayesian linear regression with error terms that have symmetric unimodal densities.

J. Nonparametr. Statist., 4(4):335–348.

Brunner, L. J. (1995b).

Bayesian linear regression with error terms that have symmetric unimodal densities.

J. Nonparametr. Statist., 4(4):335–348.

Brunner, L. J. and Lo, A. Y. (1989).

Bayes methods for a symmetric unimodal density and its mode.

Ann. Statist., 17(4):1550–1566.

Burr, D. and Doss, H. (2005).

A Bayesian semiparametric model for random-effects meta-analysis.

J. Amer. Statist. Assoc., 100(469):242–251.

Bush, C.A., M. S.Ñ. (1993).

A semiparametric bayesian model for randomized block designs.

ISDS Discussion Paper 93 A22 . Duke University.

Dalal, S. R. (1979).

Dirichlet invariant processes and applications to nonparametric estimation of symmetric distribution functions.

Stochastic Process. Appl., 9(1):99–107.

Damien, P. and Walker, S. (2002).

A Bayesian non-parametric comparison of two treatments.

Scand. J. Statist., 29(1):51–56.

Dey, D. K. and Sinha, D. (1998).

Bayesian model determination in lifetime data analysis.

Rebrape, 12(2):113–130.

Doss, H. (1984).

Bayesian estimation in the symmetric location problem.

Z. Wahrsch. Verw. Gebiete, 68(2):127–147.

Doss, H. (1985).

Bayesian nonparametric estimation of the median. I. Computation of the estimates.

Ann. Statist., 13(4):1432–1444.

Dunson, D. B. and Taylor, J. A. (2005).

Approximate Bayesian inference for quantiles.

J. Nonparametr. Stat., 17(3):385–400.

Escobar, M. D. (1994).

Estimating normal means with a Dirichlet process prior.

J. Amer. Statist. Assoc., 89(425):268–277.

Escobar, M. D. and West, M. (1995a).

Bayesian density estimation and inference using mixtures.

J. Amer. Statist. Assoc., 90(430):577–588.

Escobar, M. D. and West, M. (1995b).

Bayesian density estimation and inference using mixtures.

J. Amer. Statist. Assoc., 90(430):577–588.

Fang, K. T., Kotz, S., and Ng, K. W. (1990).

Symmetric multivariate and related distributions, volume 36 of *Monographs on Statistics and Applied Probability.*

Chapman and Hall Ltd., London.

Feller, W. (1971).

An introduction to probability theory and its applications. Vol. II.

Second edition. John Wiley & Sons Inc., New York.

Ferguson, T. S. (1973).

A Bayesian analysis of some nonparametric problems.

Ann. Statist., 1:209–230.

Ferguson, T. S. (1983).

Bayesian density estimation by mixtures of normal distributions.

In *Recent advances in statistics*, pages 287–302. Academic Press, New York.

Fernández, C. and Steel, M. F. J. (1998).

Reference priors for non-normal two-sample problems.

Test, 7(1):179–205.

Ferreira, J. T. A. S. and Steel, M. F. J. (2004).

Bayesian multivariate skewed regression modeling with an application to firm size.

In *Skew-elliptical distributions and their applications*, pages 175–190. Chapman & Hall/CRC, Boca Raton, FL.

Ferreira, J. T. A. S. and Steel, M. F. J. (2005).

Modelling directional dispersion through hyperspherical log-splines.

J. R. Stat. Soc. Ser. B Stat. Methodol., 67(4):599–616.

Freedman, D. A. and Diaconis, P. (1982).

On inconsistent M-estimators.

Ann. Statist., 10(2):454–461.

Gasparini, M. (1996).

Bayesian density estimation via Dirichlet density processes.

J. Nonparametr. Statist., 6(4):355–366.

Gaudard, M. and Hadwin, D. (1989).

Sigma-algebras on spaces of probability measures.

Scand. J. Statist., 16(2):169–175.

Gelfand, A. E. (1999).

Approaches for semiparametric Bayesian regression.

In *Asymptotics, nonparametrics, and time series*, volume 158 of *Statist. Textbooks Monogr.*, pages 615–638. Dekker, New York.

Gelfand, A. E. and Kottas, A. (2003).

Bayesian semiparametric regression for median residual life.

Scand. J. Statist., 30(4):651–665.

Genton, M. G. (2004).

Skew-symmetric and generalized skew-elliptical distributions.

In *Skew-elliptical distributions and their applications*, pages 81–100. Chapman & Hall/CRC, Boca Raton, FL.

Geweke, J. and Terui, N. (1993).

Bayesian threshold autoregressive models for nonlinear time series.

J. Time Ser. Anal., 14(5):441–454.

Ghosh, J. K. and Ramamoorthi, R. V. (2003).

Bayesian nonparametrics.

Springer Series in Statistics. Springer-Verlag, New York.

Hannum, R. and Hollander, M. (1983).

Correction: "Robustness of Ferguson's Bayes estimator of a distribution function".

Ann. Statist., 11(4):1267.

Ishwaran, H. and James, L. F. (2003).

Generalized weighted Chinese restaurant processes for species sampling mixture models.

Statist. Sinica, 13(4):1211–1235.

Johnson, N. L. and Kotz, S. (1977).

Urn models and their application.

John Wiley & Sons, New York-London-Sydney.

An approach to modern discrete probability theory, Wiley Series in Probability and Mathematical Statistics.

Kottas, A. and Gelfand, A. E. (2001).

Bayesian semiparametric mdeian regression modeling.

J. Amer. Statist. Assoc., 96(456):1458–1468.

Lijoi, A., Prünster, I., and Walker, S. G. (2005).

On consistency of nonparametric normal mixtures for Bayesian density estimation.

J. Amer. Statist. Assoc., 100(472):1292–1296.

Liseo, B. (2004).

Skew-elliptical distributions in Bayesian inference.

In *Skew-elliptical distributions and their applications*, pages 153–171. Chapman & Hall/CRC, Boca Raton, FL.

Lo, A. Y. (1984).

On a class of Bayesian nonparametric estimates. I. Density estimates.

Ann. Statist., 12(1):351–357.

MacEachern, S.Ñ. (1998).

Computational methods for mixture of Dirichlet process models.

In *Practical nonparametric and semiparametric Bayesian statistics*, volume 133 of *Lecture Notes in Statist.*, pages 23–43. Springer, New York.

Mudholkar, G. S. and Hutson, A. D. (2000).

The epsilon-skew-normal distribution for analyzing near-normal data.

J. Statist. Plann. Inference, 83(2):291–309.

Müller, P., Erkanli, A., and West, M. (1996).

Bayesian curve fitting using multivariate normal mixtures.

Biometrika, 83(1):67–79.

Müller, P. and MacEachern, S. (1998).

Estimating mixture of dirichlet process models.

Journal of Computationalñ and Graphical Statistics, (7):223–338.

Müller, P. and Quintana, F. A. (2004).

Nonparametric Bayesian data analysis.

Statist. Sci., 19(1):95–110.

Neal, R. M. (2000).

Markov chain sampling methods for Dirichlet process mixture models.

J. Comput. Graph. Statist., 9(2):249–265.

Osiewalski, J. and Steel, M. F. J. (1993).

Bayesian marginal equivalence of elliptical regression models.

J. Econometrics, 59(3):391–403.

Quintana, F. A. and Iglesias, P. L. (2003).

Bayesian clustering and product partition models.

J. R. Stat. Soc. Ser. B Stat. Methodol., 65(2):557–574.

Ramgopal, P., Laud, P. W., and Smith, A. F. M. (1994).

Correction to: "Nonparametric Bayesian bioassay with prior constraints on the shape of the potency curve" [Biometrika **80** (1993), no. 3, 489–498; MR1248016 (94g:62089)].

Biometrika, 81(3):631.

Richardson, S. and Green, P. J. (1997).

On Bayesian analysis of mixtures with an unknown number of components.

J. Roy. Statist. Soc. Ser. B, 59(4):731–792.

Roeder, K. (1992).

Semiparametric estimation of normal mixture densities.

Ann. Statist., 20(2):929–943.

Roeder, K. and Wasserman, L. (1997).

Practical Bayesian density estimation using mixtures of normals.

J. Amer. Statist. Assoc., 92(439):894–902.

Sahu, S. K., Dey, D. K., and Branco, M. D. (2003).

A new class of multivariate skew distributions with applications to Bayesian regression models.

Canad. J. Statist., 31(2):129–150.

Sethuraman, J. (1994).

A constructive definition of Dirichlet priors.

Statist. Sinica, 4(2):639–650.

Stephens, M. (2000).

Bayesian analysis of mixture models with an unknown number of components—an alternative to reversible jump methods.

Ann. Statist., 28(1):40–74.

Walker, S. G., Damien, P., Laud, P. W., and Smith, A. F. M. (1999).

Bayesian nonparametric inference for random distributions and related functions.

J. R. Stat. Soc. Ser. B Stat. Methodol., 61(3):485–527.

With discussion and a reply by the authors.

West, M., Müller, P., and Escobar, M. D. (1994a).

Hierarchical priors and mixture models, with application in regression and density estimation.

In *Aspects of uncertainty*, Wiley Ser. Probab. Math. Statist. Probab. Math. Statist., pages 363–386. Wiley, Chichester.

West, M., Müller, P., and Escobar, M. D. (1994b).

Hierarchical priors and mixture models, with application in regression and density estimation.

In *Aspects of uncertainty*, Wiley Ser. Probab. Math. Statist. Probab. Math. Statist., pages 363–386. Wiley, Chichester.

Yu, K. and Moyeed, R. A. (2001).

Bayesian quantile regression.

Statist. Probab. Lett., 54(4):437–447.

Yu, K., Van Kerm, P., and Zhang, J. (2005).

Bayesian quantile regression: an application to the wage distribution in 1990s Britain.

Sankhyā, 67(2):359–377.

Zellner, A. (1975).

Bayesian analysis of regression error terms.

J. Amer. Statist. Assoc., 70:138–144.