



UNIVERSIDAD NACIONAL AUTÓNOMA  
DE MÉXICO

---

---

POSGRADO EN CIENCIAS MATEMÁTICAS  
FACULTAD DE CIENCIAS

PROCESOS NORMALIZADOS COMO DISTRIBUCIONES  
INICIALES EN INFERENCIA BAYESIANA

T E S I S

QUE PARA OBTENER EL GRADO ACADÉMICO DE:

MAESTRA EN CIENCIAS

P R E S E N T A :

KARLA EUGENIA MEDINA DÍAZ

DIRECTOR DE TESIS: DR. RAMSÉS HUMBERTO MENA CHÁVEZ

MÉXICO, D. F.

DICIEMBRE, 2006



Universidad Nacional  
Autónoma de México



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

# Agradecimientos

MADRE muchas gracias por las enseñanzas recibidas y por todo el apoyo que me has otorgado.

HERMANOS, gracias por el cariño.

RAMSES, muchas gracias por el apoyo para la realización de esta tesis.

MAESTROS, gracias por todos los conocimientos que me brindaron.

# Índice general

<b>Resumen</b>	<b>4</b>
<b>Introducción</b>	<b>5</b>
<b>1. Preliminar</b>	<b>10</b>
1.1. Elementos básicos . . . . .	10
1.2. Ejemplos de procesos de Lévy . . . . .	13
Proceso Gamma . . . . .	13
Proceso Estable . . . . .	14
Proceso Gaussiano Inverso . . . . .	14
<b>2. Procesos de distribuciones aleatorias</b>	<b>16</b>
2.1. Familia de distribuciones vía normalización . . . . .	16
Propiedades de las distribuciones N-prior . . . . .	18
Estructura de correlación . . . . .	24
Distribución predictiva . . . . .	31
2.2. Ejemplos de procesos normalizados . . . . .	44
Proceso Dirichlet o Gamma normalizado . . . . .	44
Proceso Estable normalizado . . . . .	47
Proceso Gaussiano Inverso normalizado . . . . .	50
2.3. Esquema de la Urna de Pólya . . . . .	57
2.4. Mezclas de procesos normalizados . . . . .	60

---

2.5. Distribución de $K_n$ . . . . .	62
Proceso Dirichlet o Gamma normalizado . . . . .	64
Proceso Estable normalizado . . . . .	64
Proceso Gaussiano Inverso normalizado . . . . .	66
<b>3. Aplicaciones de los procesos normalizados</b>	<b>68</b>
3.1. Mezcla uniforme de 3 distribuciones normales . . . . .	69
3.2. Mezcla uniforme de 6 distribuciones normales . . . . .	77
3.3. Datos de la galaxias . . . . .	81
<b>Comentarios finales</b>	<b>90</b>
<b>Apéndice</b>	<b>92</b>
<b>Bibliografía</b>	<b>95</b>

# Resumen

La estadística Bayesiana considera la información inicial del fenómeno con la que se cuenta y la información contenida en los datos, resumiendo todo el conocimiento en una distribución de probabilidades sobre las características de interés. En particular, la estadística Bayesiana no paramétrica ha tomado mucha fuerza en los últimos años gracias a los avances computacionales y al estudio reciente de distribuciones iniciales para funciones de distribuciones y otras funciones de interés desconocidas. En este trabajo se presenta un tipo de distribuciones iniciales vía normalización a través de procesos de Lévy, y además se describen algunas características de estas distribuciones como la media, la varianza y la covarianza. Adicionalmente, se da la distribución predictiva a un paso y a un paso para cualquier  $n$ . Finalmente, se dan tres aplicaciones donde se utiliza y se observa la importancia de la distribución del número de componentes de una mezcla. Las dos primeras aplicaciones se realizaron con datos simulados y en la tercera aplicación se usó la información de la velocidad de las galaxias que se ha usado extensamente dentro de la literatura Bayesiana no paramétrica.

# Introducción

La estadística Bayesiana, en comparación con la estadística clásica, permite agregar el conocimiento previo de un fenómeno en las inferencias estadísticas, es decir utiliza la información inicial que tenemos sobre el fenómeno y la información sobre éste contenida en los datos.

Usualmente en la estadística clásica se realizan las inferencias bajo el supuesto de que todas las observaciones de una muestra se obtuvieron bajo las mismas condiciones, es decir que están idénticamente distribuidas. En el caso de la estadística Bayesiana se asume que la muestra es una sucesión de variables **intercambiables**, es decir que el orden de aparición de las observaciones no afecta la asignación de la probabilidad de los datos no observados. La importancia de este supuesto se observa a través del siguiente teorema:

**Teorema de de Finetti 1** *Si  $X_1, X_2, \dots, X_n, \dots$  es una sucesión de variables aleatorias intercambiables con medida de probabilidad  $P$ , entonces existe una medida de probabilidad  $\Pi$  sobre el espacio de funciones de distribución,  $\mathbb{F}$ , tal que la distribución conjunta de  $(X_1, \dots, X_n)$  tiene la forma*

$$P(X_1 \in A_1, \dots, X_n \in A_n) = \int_{\mathbb{F}} \prod_{i=1}^n F(A_i) \Pi(dF),$$

donde  $\Pi(F) = \lim_{n \rightarrow \infty} \Pi(F_n)$  y  $F_n$  es la función de distribución empírica correspondiente de  $(X_1, X_2, \dots, X_n)$ .

Este supuesto ayuda a dar una estructura de dependencia entre las observaciones, para

aprender de observaciones pasadas y así poder hacer predicciones; a su vez también ayuda a justificar la necesidad de una distribución inicial.

En algunas situaciones se pueden realizar algunos supuestos sobre la forma de la distribución que describe las observaciones, así como de la distribución que concentra la información disponible sobre los parámetros que están involucrados en la distribución. Dentro de estos supuestos, se encuentran principalmente la unimodalidad, simetría y cierto comportamiento en las colas. Pero estos supuestos en la mayoría de las ocasiones no son de ayuda ante situaciones donde no se conoce con exactitud la forma de la distribución de los datos y de los parámetros, por lo tanto se necesita una modelación más robusta que sea capaz de asumir diversas formas para la distribución subyacente de los datos.

Esto último lleva a tomar un enfoque no paramétrico y dentro de un paradigma Bayesiano a la modelación de la medida  $\Pi(\cdot)$ . A este enfoque se le conoce como Bayesiano no paramétrico, que tiene como base la construcción de medidas de probabilidad sobre distribuciones aleatorias, así como su uso como información inicial. Una de las características de este enfoque es que permite relajar los supuestos de forma e independencia usados en la estadística clásica, y a su vez es más consistente con el teorema de representación de de Finetti porque no usa un subespacio del espacio de funciones de distribución.

Aparentemente el primer trabajo sobre la construcción de medidas aleatorias se da con Freedman [1963] que considera las distribuciones libres de colas (tailfree), aunque no es hasta que se da a conocer el trabajo de Ferguson [1973] donde se empieza a trabajar con mayor hincapié dentro de la estadística Bayesiana no paramétrica. Cabe aclarar que entre el trabajo de Freedman [1963] y Ferguson [1973] existen otros trabajos no directamente relacionados con el enfoque Bayesiano, pero que a su vez exploran ideas de distribuciones aleatorias.



---

En el trabajo de Ferguson [1973] se formaliza la idea principal del enfoque Bayesiano no paramétrico y se introduce el proceso Dirichlet. A partir de la introducción de este proceso se han generado otras contribuciones en el área, en particular encontramos la representación de Blackwell y MacQueen [1973], la cual forma la base del esquema de la Urna de Pólya, que a su vez sustenta un algoritmo para la simulación de dicho proceso (sección 2.3); el trabajo de Antoniak [1974] donde se aleatorizan los parámetros del proceso Dirichlet y así se obtiene una mezcla de procesos Dirichlet (sección 2.4); el trabajo de Lo [1984] donde se propone otro tipo de mezcla de este proceso, el cual asigna un proceso de Dirichlet a la distribución inicial de los parámetros de un kernel no negativo (sección 2.4). Este último, como se verá mas adelante, se puede ver como un modelo jerárquico.

Otro trabajo de interés es Escobar y West [1995] que une estos dos últimos trabajos junto con la idea del esquema de la Urna de Pólya, y propone un algoritmo para la simulación de mezclas de procesos Dirichlet.

Existen otras perspectivas para obtener medidas aleatorias, entre las que destacan: el análisis de los procesos neutrales por la derecha (Neutral to the Right Process, NRP), hecha por Doksum [1974]; el trabajo de Lenk [1988] donde presenta al proceso Log-Gaussiano Normalizado que es un caso continuo y no discreto como lo es el proceso Dirichlet y el trabajo, que es de mayor interés en esta tesis, de Regazzini, Lijoi, y Prünster [2003] donde se presentan los procesos de Lévy crecientes normalizados (sección 2.1).

El presente trabajo está compuesto por cuatro capítulos. En el primer capítulo, que es el capítulo preliminar, se exponen los elementos básicos necesarios para el desarrollo de las distribuciones N-prior. El segundo capítulo expone las características y propiedades de los procesos normalizados, aquí llamados distribuciones N-Prior. Este capítulo se basa principalmente en el trabajo de Regazzini, Lijoi, y Prünster [2003] y el artículo de James, Lijoi, y Prünster [2006].

Adicionalmente, se presenta un algoritmo para la simulación de estos procesos basado en el esquema de la Urna de Pólya (Escobar y West [1995]) y en el algoritmo presentado por Ishwaran y James [2001], el cual da una forma de poder simular a las distribuciones finales, que a su vez se generalizan a modelos semiparamétricos basado en el proceso Dirichlet. Estos modelos también se exponen en este capítulo, aplicando el trabajo de Escobar y West [1995], sustituyendo los procesos Dirichlet por distribuciones N-prior.

Por último, en este segundo capítulo se expone la importancia de  $K_n$ , que representa el número de valores diferentes en una muestra de tamaño  $n$ , así como su distribución. La distribución de  $K_n$  es de interés ya que en particular se puede usar para determinar los parámetros de ciertas distribuciones N-prior, por ejemplo en el caso del proceso Dirichlet puede ayudar para determinar el valor de la masa total,  $a$ . Además  $K_n$  se puede utilizar en el análisis de conglomerados para determinar el número de grupos existentes en una muestra (Lijoi, Mena, y Prünster [2006b]).

La teoría presentada en el segundo capítulo se ejemplifica considerando tres tipos de procesos de Lévy: proceso Gamma, proceso Estable y proceso Gaussiano Inverso.

El tercer capítulo presenta tres ejemplos donde se aplica la teoría expuesta en el capítulo 2. Los dos primeros ejemplos utilizan datos simulados de mezclas uniformes de distribuciones normales con diferentes medias y varianzas iguales, y se asignan iniciales para  $K_n$  con modas cercanas al número real de elementos de la mezclas, así como modas lejanas. Se obtienen las distribuciones finales de  $K_n$  y la distribución predictiva final.

El tercer ejemplo utiliza los datos de las Galaxias presentados inicialmente por Roeder [1990] y después en el contexto no paramétrico por Escobar y West [1995], donde se aplica la teoría desarrollada, y se compara el número de galaxias encontradas contra las expuestas

en el artículo que dio a conocer los datos. En el último capítulo, se presenta la discusión y los comentarios finales de esta tesis.

# Capítulo 1

## Preliminar

Este capítulo preliminar presenta las definiciones y teoremas necesarios para construir los procesos normalizados, que son el tema principal de esta tesis y que se estudiarán en el capítulo 2.

### 1.1. Elementos básicos

Como parte esencial en la construcción de distribuciones aleatorias discretas, que presentaremos en este capítulo, se requiere la noción de ciertos procesos con incrementos independientes. A continuación se presentan algunas definiciones y teoremas sin demostración (ver Sato [1991]) en las versiones relevantes al trabajo en cuestión.

**Definición 1** *Un proceso estocástico  $\{X_t : t \geq 0\}$  definido en  $\mathbb{R}$  es un **proceso de Lévy creciente o subordinador** si:*

1. Para cualquier  $n \geq 1$  y  $0 \leq t_0 < t_1 < \dots < t_n$ , las variables aleatorias  $X_{t_0}, X_{t_1} - X_{t_0}, X_{t_2} - X_{t_1}, \dots, X_{t_n} - X_{t_{n-1}}$  son independientes (propiedad de incrementos independientes).
2.  $X_0 = 0$  c.s. ( $P[X_0 = 0] = 1$ ).

3. La distribución de  $X_{s+t} - X_s$  no depende de  $s$  (propiedad de incrementos estacionarios).
4. Es continuo estocásticamente, es decir  $\lim_{s \rightarrow t} P[|X_s - X_t| > \epsilon] = 0$  para todo  $t \geq 0$  y  $\epsilon > 0$ .
5. Existe un conjunto  $\Omega_0$  con  $P[\Omega_0] = 1$  tal que, para todo  $\omega \in \Omega_0$ ,  $X_t(\omega)$  es continuo por la derecha en  $t \geq 0$  y tiene límites por la izquierda en  $t > 0$ .
6.  $X_t(\omega)$  es no decreciente c.s. como una función de  $t$ .

Cabe mencionar que en el trabajo de Regazzini, Lijoi, y Prünster [2003] la condición 3 puede ser relajada y así entonces considerar procesos con incrementos independientes que no son necesariamente homogéneos. Sin embargo, en esta revisión nos enfocaremos a procesos de Lévy crecientes, los cuales son homogéneos.

### **Teorema 1 Representación de Lévy-Khintchine**

La función característica de un proceso estocástico continuo con **incrementos independientes** (PII),  $X = \{X_t : t \geq 0\}$  en  $\mathbb{R}$  y  $X_0 = 0$  c.s. está dada por

$$E[e^{-zX_t}] = \exp\{-\psi_t(z)\}, \quad z \in \mathbb{C}, \quad (1.1)$$

donde  $\psi_t$  es llamado el **exponente característico** y está dado por:

$$\psi_t(z) = zd_t + \frac{z^2}{2}G_t + \int_{\mathbb{R}} \{e^{zx} - 1 - x1_{\{|x|<1\}}(x)\} \nu_t(dx), \quad (1.2)$$

donde  $G_t$  es un operador lineal continuo simétrico no negativo con  $G_{t_2} - G_{t_1}$  no negativo para  $t_1 < t_2$  llamado **término Gaussiano**,  $d_t$  es una función continua en  $\mathbb{R}$  llamada **deriva**, y  $\nu_t$  es una medida en  $\mathbb{R}$  que satisface

$$\nu_t(\{0\}) = 0 \quad y \quad \int_{\mathbb{R}} \min(|x|^2, 1) \nu_t(dx) < \infty, \quad (1.3)$$

llamada **medida de Lévy**.

La representación de un PII, llamada canónica, es  $(d_t, G_t, \nu_t)$ , la cual es única (ver Sato [1999]), y donde  $d_t$  es el responsable del desarrollo en promedio del proceso  $X = \{X_t : t \geq 0\}$ ;  $G_t$  define la varianza del componente Gaussiano continuo y  $\nu_t$  es responsable de los brin-  
cos que va dando  $X$ , y exhibe la frecuencia y la magnitud de éstos, en general su intensidad.

Cuando se considera un proceso de Lévy, entonces la representación del PII resulta en  $(d_t, G_t, \nu_t) = (td, tG, t\nu)$  ya que este proceso tiene la propiedad de incrementos estacionarios, además el exponente característico es  $\psi_t(u) = t\psi(u)$ , donde  $\psi$  está dado por (1.2). A este último también se le denominará exponente característico correspondiente a un proceso de Lévy. Además, observemos que cuando  $d = 0$  y  $G = 0$  en un proceso de Lévy tenemos un proceso no decreciente de brin-  
cos puros.

Cuando  $\{X_t\}$  es un subordinador, la transformada de Laplace de su distribución es más conveniente que la función característica. La forma general es

$$E[e^{-uX_t}] = \exp \left[ -td - t \left( \int_{\mathbb{R}^+} (1 - e^{-ux}) \nu(dx) \right) \right] \text{ para } u \geq 0 \quad (1.4)$$

y la condición (1.3) se simplifica a

$$\nu(\{0\}) = 0 \quad \text{y} \quad \int_{\mathbb{R}^+} \min(x, 1) \nu(dx) < \infty. \quad (1.5)$$

En este trabajo se usarán distribuciones con extremidad izquierda,  $d = 0$ , en otras palabras se tiene que  $\inf \{x : Pr(X \leq x) > 0\} = 0$ .

Como veremos más adelante, la construcción de distribuciones aleatorias requiere que  $\nu(\mathbb{R}^+) = \infty$ , de tal forma que la normalización, utilizada en esta tesis, esté bien definida. La siguiente sección presentará tres procesos que cumplen con esta condición.

## 1.2. Ejemplos de procesos de Lévy

En esta sección, con ayuda de las definiciones de la sección anterior y el teorema de Lévy-Khintchine, se verá que el proceso Gamma, el proceso Estable y el proceso Gaussiano Inverso son procesos de Lévy crecientes. Estos procesos se utilizarán en la sección 2.2 para construir procesos normalizados.

### Proceso Gamma

Sea  $\{X_t\}_{t \geq 0}$  un proceso de Lévy con brinco Gamma, es decir  $X_t$  tiene una distribución Gamma con parámetros  $(t, \beta)$  para toda  $t \in \mathbb{R}^+$ .

La función de densidad de la distribución Gamma es

$$f(x) = \frac{\beta^t}{\Gamma(t)} x^{t-1} e^{-\beta x} I(x \geq 0).$$

Entonces la transformada de Laplace de su distribución es

$$E[e^{-uX_t}] = \left(\frac{\beta + u}{\beta}\right)^{-t} = \exp\left\{-t \int_0^\infty (1 - e^{-ux}) x^{-1} e^{-\beta x} dx\right\},$$

lo cual nos lleva a que

$$\nu(dx) = x^{-1} e^{-\beta x} dx. \tag{1.6}$$

A partir de lo anterior se deduce que el exponente característico es igual a

$$\psi(u) = \log\left(\frac{\beta + u}{\beta}\right). \tag{1.7}$$

Dado que  $\nu(0^+) = \lim_{x \downarrow 0} \nu(x) = \infty$  entonces  $\nu(\mathbb{R}^+) = \infty$ ; además se satisface claramente la condición  $\int_{\mathbb{R}^+} \min(x, 1) \nu(dx) < \infty$  ya que en una vecindad que no contiene el cero la medida  $\nu$  es finita.

## Proceso Estable

Sea  $\{X_t\}_{t \geq 0}$  en  $\mathbb{R}$  un proceso de Lévy con brincos  $\gamma$ -Estable,  $\gamma \in (0, 1)$ . En general para este proceso no existe una forma analítica para la función de densidad. La medida de Lévy de una distribución Estable es

$$\nu(dx) = c x^{-(1+\gamma)} dx \quad c > 0, \quad (1.8)$$

de donde se sigue que la transformada de Laplace está dada por

$$E[e^{-uX_t}] = \exp \left\{ -t \frac{c \Gamma(1-\gamma) u^\gamma}{\gamma} \right\}.$$

De esto se obtiene que el exponente característico es

$$\psi(u) = \frac{c \Gamma(1-\gamma) u^\gamma}{\gamma}. \quad (1.9)$$

Observemos que  $\nu(\mathbb{R}^+) = \infty$  ya que  $\nu(0^+) = \lim_{x \downarrow 0} \nu(x) = \infty$  y que, de igual forma que el proceso Gamma,  $\int_{\mathbb{R}^+} \min(x, 1) \nu(dx) < \infty$ .

El proceso de Lévy con brincos  $\gamma$ -Estable también se puede definir como:

1.  $\{X_t\}_{t \geq 0}$  es un proceso de Lévy y
2. para cada  $a > 0$  se satisface

$$\{X_{at}, t \geq 0\} \stackrel{d}{=} \{a^{1/\gamma} X_t, t \geq 0\}.$$

## Proceso Gaussiano Inverso

Sea  $\{X_t\}_{t \geq 0}$  un proceso de Lévy con brincos distribuidos de acuerdo a una ley Gaussiana Inversa (denotado por IG por sus siglas en inglés), es decir  $X_t$  tiene una distribución Gaussiana Inversa con parámetros  $(t\delta, \gamma)$  para toda  $t \in \mathbb{R}^+$ .



La función de densidad de la distribución IG está dada por

$$f_{X_t}(x) = \frac{t\delta}{\sqrt{2\pi}} x^{-\frac{3}{2}} \exp \left\{ -\frac{1}{2} \left( \frac{t^2\delta^2}{x} + \gamma^2 x \right) + t\gamma\delta \right\} I(x \geq 0),$$

con  $\delta > 0$  y  $\gamma > 0$ .

Se puede ver que la transformada de Laplace de la distribución para este proceso se descompone como

$$E[e^{-uX_t}] = \exp \left\{ -t\delta \left( \sqrt{2u + \gamma^2} - \gamma \right) \right\} = \exp \left\{ t \left( \int_0^\infty (e^{-ux} - 1) \frac{\delta}{\sqrt{2\pi}} x^{-\frac{3}{2}} e^{-\frac{\delta^2 x}{2}} dx \right) \right\},$$

lo cual nos lleva a que la medida de Lévy es

$$\nu(dx) = \frac{\delta}{\sqrt{2\pi}} x^{-\frac{3}{2}} e^{-\frac{\delta^2 x}{2}} dx, \quad (1.10)$$

con exponente característico

$$\psi(u) = \delta \left( \sqrt{2ui + \delta^2} - \delta \right). \quad (1.11)$$

Igual que en los dos ejemplos anteriores, se cumple que  $\nu(\mathbb{R}^+) = \infty$  debido a que  $\nu(0^+) = \lim_{x \downarrow 0} \nu(x) = \infty$  y también se cumple que  $\int_{\mathbb{R}^+} \min(x, 1) \nu(dx) < \infty$ .

Los tres ejemplos que hemos presentado constituirán la base de las distribuciones aleatorias estudiadas en este trabajo.

Sean  $X_1, X_2, \dots, X_n$  una sucesión de variables aleatorias definidas sobre un espacio medible  $(\Omega, \sigma)$  y que toman valores en el espacio medible  $(\mathbb{X}, \mathcal{X}, \tilde{P})$ , donde  $\mathcal{X}$  es la  $\sigma$ -álgebra generada por los conjuntos de  $\mathbb{X}$  y  $\tilde{P}$  está definida en el espacio medible  $(\mathbb{F}, \mathcal{F}, \pi)$ , donde  $\mathbb{F}$  es la familia de funciones de distribución y  $\mathcal{F}$  es la  $\sigma$ -álgebra generada por  $\mathbb{F}$ .

De una manera particular, en esta tesis se considerarán  $X_1, X_2, \dots, X_n$  variables aleatorias,  $X_i : (\Omega, \sigma) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$  y  $\tilde{P}$  es una medida aleatoria sobre  $(\mathbb{F}, \mathcal{F})$ .

# Capítulo 2

## Procesos de distribuciones aleatorias

En este capítulo estudiaremos las distribuciones aleatorias construidas a través de la normalización de procesos de Lévy crecientes. Este tipo de distribuciones fueron inicialmente exploradas de manera general por Regazzini, Lijoi, y Prünster [2003] donde exponen características y propiedades de estos procesos. Cabe mencionar que, como caso particular de este tipo de medidas, se encuentra el proceso Dirichlet que presentó Ferguson [1973], cuyas generalizaciones resultan en modelos que hacen más adecuada la aplicación de la estadística Bayesiana no paramétrica.

### 2.1. Familia de distribuciones vía normalización

En el artículo de Ferguson [1973] se comenta que es deseable que las distribuciones aleatorias cumplan los siguientes dos puntos;

1. El soporte de la distribución inicial debe ser grande -con respecto a alguna topología sobre el espacio de distribuciones de probabilidad.
2. Las distribuciones finales dada una muestra de una distribución de probabilidad debe ser analíticamente tratable.

Como veremos mas adelante, la condición 2 no la cumplen *todas* las distribuciones vía normalización, ya que en general no tienen la propiedad de ser conjugadas; además los casos que se estudiarán en esta tesis cumplen con la condición 1.

Las distribuciones vía normalización fueron dadas a conocer por Regazzini, Lijoi, y Prünster [2003] donde ellos las consideran de importancia por dos razones:

1. Representan una aproximación natural al problema de definir una medida de probabilidad aleatoria.
2. Es posible determinar formas explícitas para la distribución de la media de cualquier medida de probabilidad aleatoria a través de esta aproximación.

En este artículo únicamente se presentó la teoría general. Sin embargo, el estudio de los casos particulares y que a su vez sean viables no siempre es posible de una manera analítica.

Después, Lijoi, Mena, y Prünster [2005a] presentaron un artículo donde exponen de forma general las distribuciones N-prior poniendo interés en el proceso IG normalizado y James, Lijoi, y Prünster [2006] proponen una técnica para derivar funcionales de las distribuciones N-Prior.

En esta sección se exponen las características y propiedades de estas distribuciones que nos serán de ayuda para realizar inferencias posteriormente.

**Definición 2** *Dadas una medida finita no nula  $\alpha$  sobre  $\mathbb{R}$  con una masa total  $a$ ,  $a := \alpha(\mathbb{R})$ , una medida de Lévy  $\nu$  sobre  $\mathbb{R}^+$  que satisface  $\nu(\mathbb{R}^+) = \infty$  y su correspondiente proceso de Lévy  $\xi$ , con  $(d = 0, G = 0, \nu)$ , se define la medida de probabilidad aleatoria*

$$\tilde{P}(\cdot) = \frac{\xi_{\alpha(\cdot)}}{\xi_a},$$

*llamada N-prior con parámetros  $(\alpha, \nu)$ .*

La condición  $\nu(\mathbb{R}^+) = \infty$  se interpreta como que el número de brinco en cualquier intervalo abierto cercano a cero,  $(0, \epsilon)$ , es infinito y por lo tanto evitará que el denominador  $\xi_a$  se vuelva cero.

Sea  $\mathbb{F}$  la clase de distribuciones N-Prior y sea  $\mathcal{F} = \sigma(\mathbb{F})$ , la  $\sigma$ -álgebra relacionada a este espacio.

## Propiedades de las distribuciones N-prior

**Teorema 2** *Si  $\tilde{P}$  se distribuye N-Prior con parámetros  $(\alpha, \nu)$ , entonces*

$$E_\pi [\tilde{P}(B)] = \frac{\alpha(B)}{a},$$

para cualquier conjunto medible  $B \in \mathcal{X}$ .

### Demostración

Sea  $\tilde{P}$  N-prior, entonces  $\tilde{P}(\cdot) = \frac{\xi_{\alpha(\cdot)}}{\xi_a}$  donde  $\{\xi_t\}$  es un proceso de Lévy, y además sabemos que  $a := \alpha(\mathbb{R})$ .

Observemos que

$$\int_0^\infty e^{-u\xi_a} du = \frac{1}{\xi_a}.$$

Considerando la ecuación anterior, calculemos la esperanza de  $\tilde{P}$ ,

$$E_\pi [\tilde{P}(B)] = \int \frac{\xi_{\alpha(B)}}{\xi_a} \tilde{P}_\xi(d\xi) = \int \xi_{\alpha(B)} \int_0^\infty e^{-u\xi_a} du \tilde{P}_\xi(d\xi).$$

Ya que los integrandos son mayores o iguales a cero, podemos aplicar el teorema de Fubini-Tonelli,

$$E_\pi [\tilde{P}(B)] = \int_0^\infty \int e^{-u\xi_a} \xi_{\alpha(B)} \tilde{P}_\xi(d\xi) du = \int_0^\infty E [e^{-u\xi_a} \xi_{\alpha(B)}] du.$$

Sean  $A$  y  $B \in \mathcal{X}$  con  $A \cap B = \emptyset$ , entonces

$$\begin{aligned}
 E \left[ e^{u\xi_{\alpha(A \cup B)}} \right] &= e^{\alpha(A \cup B)} \int_{\mathbb{R}^+} (1 - e^{-uv}) \nu(dv) \\
 &= e^{(\alpha(A) + \alpha(B))} \int_{\mathbb{R}^+} (1 - e^{-uv}) \nu(dv) \\
 &= e^{\alpha(A)} \int_{\mathbb{R}^+} (1 - e^{-uv}) \nu(dv) e^{\alpha(B)} \int_{\mathbb{R}^+} (1 - e^{-uv}) \nu(dv) \\
 &= E \left[ e^{u\xi_{\alpha(A)}} \right] E \left[ e^{u\xi_{\alpha(B)}} \right].
 \end{aligned} \tag{2.1}$$

Podemos definir  $\xi_{\alpha(A \cup B)} = (\xi_{\alpha(A \cup B)} - \xi_{\alpha(A)}) + \xi_{\alpha(A)}$  y por incrementos independientes, tenemos que los términos de la suma son variables aleatorias independientes, entonces

$$\begin{aligned}
 E \left[ e^{u\xi_{\alpha(A \cup B)}} \right] &= E \left[ e^{u((\xi_{\alpha(A \cup B)} - \xi_{\alpha(A)}) + \xi_{\alpha(A)})} \right] \\
 &= E \left[ e^{u(\xi_{\alpha(A \cup B)} - \xi_{\alpha(A)})} \right] E \left[ e^{u\xi_{\alpha(A)}} \right].
 \end{aligned} \tag{2.2}$$

Igualando (2.1) y (2.2)

$$E \left[ e^{u\xi_{\alpha(A)}} \right] E \left[ e^{u\xi_{\alpha(B)}} \right] = E \left[ e^{u(\xi_{\alpha(A \cup B)} - \xi_{\alpha(A)})} \right] E \left[ e^{u\xi_{\alpha(A)}} \right],$$

llegamos a que

$$E \left[ e^{u\xi_{\alpha(B)}} \right] = E \left[ e^{u(\xi_{\alpha(A \cup B)} - \xi_{\alpha(A)})} \right].$$

Esto me lleva a que  $\xi_{\alpha(A \cup B)} = \xi_{\alpha(A)} + \xi_{\alpha(B)}$ , entonces

$$\begin{aligned}
 E \left[ e^{\xi_{\alpha(A \cup B)}} \right] &= E \left[ e^{\xi_{\alpha(A)} + \xi_{\alpha(B)}} \right] \\
 &= E \left[ e^{u\xi_{\alpha(A)}} \right] E \left[ e^{u\xi_{\alpha(B)}} \right]. \text{ ||por (2.1)||}
 \end{aligned}$$

Por lo tanto  $\xi_{\alpha(B)}$  y  $\xi_{\alpha(A)}$  son independientes.

Considerando que a  $\xi_a$  la podemos escribir como  $\xi_a = \xi_{\alpha(B)} + \xi_{\alpha(B^c)}$  con  $B, B^c \in \mathcal{X}$ , el cual es completo y aplicando el resultado anterior a la esperanza de  $\tilde{P}$

$$\begin{aligned}
E_\pi [\tilde{P}(B)] &= \int_0^\infty E \left[ e^{-u(\xi_{\alpha(B)} + \xi_{\alpha(B^c)})} \xi_{\alpha(B)} \right] du \\
&= \int_0^\infty E \left[ e^{-u\xi_{\alpha(B)}} \xi_{\alpha(B)} \right] E \left[ e^{-u\xi_{\alpha(B^c)}} \right] du \\
&= \int_0^\infty E \left[ -\frac{d}{du} e^{-u\xi_{\alpha(B)}} \right] E \left[ e^{-u\xi_{\alpha(B^c)}} \right] du.
\end{aligned}$$

Definamos una función que nos será de utilidad para diversas demostraciones.

Sea

$$\mu_n(u) := \int_{\mathbb{R}^+} v^n e^{-uv} \nu(dv) \quad n = 1, 2, \dots, \text{ para cualquier } u > 0, \quad (2.3)$$

la cual representa los momentos centrales de un proceso de Lévy, para el cual el exponente característico cuando  $d = 0$  es igual a

$$\psi(u) = \int_{\mathbb{R}^+} (1 - e^{-uv}) \nu(dv). \quad (2.4)$$

Al calcular la derivada de (2.4) encontramos que

$$\begin{aligned}
\frac{d\psi(u)}{du} &= \frac{d}{du} \int_{\mathbb{R}^+} (1 - e^{-uv}) \nu(dv) \\
&= \int_{\mathbb{R}^+} \frac{d}{du} (1 - e^{-uv}) \nu(dv) \\
&= \int_{\mathbb{R}^+} v e^{-uv} \nu(dv) = \mu_1(u).
\end{aligned} \quad (2.5)$$

La segunda igualdad se debe a que  $\frac{d}{du} \int_{\mathbb{R}^+} (1 - e^{-uv}) \nu(dv) < \infty$  a pesar de que  $\int_{\mathbb{R}} \nu(dv) = \infty$ .

Usando la transformada de Laplace de la función característica (1.4) tenemos que

$$E \left[ e^{-u\xi_{\alpha(B^c)}} \right] = \exp \left\{ \alpha(B^c) \left( \int_0^\infty (e^{u\xi} - 1) \nu(d\xi) \right) \right\} = \exp \{ -\alpha(B^c) \psi(u) \}. \quad (2.6)$$

Haciendo uso de (1.4) y (2.5) podemos calcular

$$\begin{aligned} E \left[ -\frac{d}{du} e^{-u\xi_{\alpha(B)}} \right] &= -\frac{d}{du} E \left[ e^{-u\xi_{\alpha(B)}} \right] = -\frac{d}{du} e^{-\psi(u)\alpha(B)} \\ &= \alpha(B) e^{-\psi(u)\alpha(B)} \left( -\frac{d}{du} \psi(u) \right) = -\alpha(B) \mu_1(u) e^{-\psi(u)\alpha(B)}. \end{aligned} \quad (2.7)$$

Aplicando (2.6) y (2.7), tenemos

$$\begin{aligned} E_{\pi} \left[ \tilde{P}(B) \right] &= \int_0^{\infty} -\alpha(B) \mu_1(u) e^{-\psi(u)\alpha(B)} e^{-\alpha(B^c)\psi(u)} du \\ &= \alpha(B) \int_0^{\infty} -\mu_1(u) e^{-\psi(u)(\alpha(B)+\alpha(B^c))} du = \alpha(B) \int_0^{\infty} -\mu_1(u) e^{-a\psi(u)} du \\ &= \frac{\alpha(B)}{a} \int_0^{\infty} a(-\mu_1(u)) e^{-a\psi(u)} du = \frac{\alpha(B)}{a} \int_0^{\infty} \left\{ -\frac{d}{du} \psi(u) \right\} a e^{-a\psi(u)} du \\ &= \frac{\alpha(B)}{a} \int_0^{\infty} \left\{ -\frac{d}{du} e^{-a\psi(u)} \right\} du = \frac{\alpha(B)}{a} \left[ e^{a0} - e^{-a\nu(\mathbb{R}^+)} \right] \\ &= \frac{\alpha(B)}{a}. \quad \square \end{aligned} \quad (2.8)$$

Una de la problemáticas a las que se enfrentan los estadísticos Bayesianos es describir la forma inicial de la distribución aleatoria  $\tilde{P}$ , y el resultado anterior nos ayuda a dar esta forma ya que no depende de la medida de Lévy  $\nu$ , sino únicamente de  $\alpha(\cdot)$ .

Más adelante, y con fines comparativos, denotaremos dicha forma inicial por  $G_0 := \frac{\alpha(\cdot)}{a}$ .

**Teorema 3** Si  $\tilde{P}$  se distribuye  $N$ -prior con parámetros  $(\alpha, \nu)$ , entonces

$$\text{Var}_{\pi} \left[ \tilde{P}(B) \right] = \frac{\alpha(B)(a - \alpha(B))}{a^2} I_a,$$

donde  $I_a = a \int_{\mathbb{R}^+} u e^{-a\psi(u)} \int_{\mathbb{R}^+} v^2 e^{-uv} \nu(dv) du$ ,  $\psi(u) = \int_{\mathbb{R}^+} (1 - e^{uv}) \nu(dv)$  y  $B \in \mathcal{X}$ .

**Demostración.**

Usando un argumento análogo al de la demostración del teorema 2 observemos que

$$\int_0^{\infty} u e^{-u\xi_a} du = \frac{1}{\xi_a^2}.$$

Lo que nos lleva a

$$\begin{aligned} E_{\pi} \left[ \left( \tilde{P}(B) \right)^2 \right] &= E \left[ \left( \frac{\xi_{\alpha(B)}}{\xi_a} \right)^2 \right] \\ &= \int_0^{\infty} u E \left[ e^{-u\xi_a} \xi_{\alpha(B)}^2 \right] du \quad \|\text{Fubini-Tonelli}\| \\ &= \int_0^{\infty} u E \left[ e^{-u\xi_{\alpha(B)}} \xi_{\alpha(B)}^2 \right] E \left[ e^{-u\xi_{\alpha(B^c)}} \right] du \quad \|\text{Incrementos Independientes}\| \\ &= \int_0^{\infty} u E \left[ \frac{d^2}{du^2} e^{-u\xi_{\alpha(B)}} \right] E \left[ e^{-u\xi_{\alpha(B^c)}} \right] du \\ &= \int_0^{\infty} u \left\{ \frac{d^2}{du^2} E \left[ e^{-u\xi_{\alpha(B)}} \right] \right\} E \left[ e^{-u\xi_{\alpha(B^c)}} \right] du. \end{aligned} \quad (2.9)$$

Retomando (2.3) y considerando (2.6), tenemos

$$E \left[ e^{-u\xi_{\alpha(B^c)}} \right] = e^{-\alpha(B^c)\psi(u)} \quad y \quad E \left[ e^{-u\xi_{\alpha(B)}} \right] = e^{-\alpha(B)\psi(u)}.$$

Entonces,

$$\begin{aligned} \frac{d^2}{du^2} E \left[ e^{-u\xi_{\alpha(B)}} \right] &= \frac{d^2}{du^2} \left( e^{-\alpha(B)\psi(u)} \right) \\ &= \frac{d}{du} \left[ -\alpha(B) e^{-\alpha(B)\psi(u)} \frac{d\psi(u)}{du} \right] \\ &= \alpha(B) \left[ \alpha(B) e^{-\alpha(B)\psi(u)} \left( \frac{d\psi(u)}{du} \right)^2 - e^{-\alpha(B)\psi(u)} \frac{d^2\psi(u)}{du^2} \right]. \end{aligned}$$

Haciendo algo equivalente a (2.5) tenemos que  $-\frac{d^2\psi(u)}{du^2} = \mu_2(u)$ . Regresando a (2.9)



$$\begin{aligned}
E_{\pi} \left[ \left( \tilde{P}(B) \right)^2 \right] &= \int_0^{\infty} u \alpha(B) \left[ \alpha(B) e^{-\alpha(B)\psi(u)} \left( -\frac{d\psi(u)}{du} \right)^2 - e^{-\alpha(B)\psi(u)} \frac{d^2\psi(u)}{du^2} \right] e^{-\alpha(B^c)\psi(u)} du \\
&= \alpha(B) \int_0^{\infty} u e^{-a\psi(u)} \left[ \alpha(B) \mu_1^2(u) + \mu_2(u) \right] \\
&= \alpha^2(B) \int_0^{\infty} u e^{-a\psi(u)} \left( -\frac{d\psi(u)}{du} \right)^2 du + \frac{\alpha(B)}{a} I_a. \tag{2.10}
\end{aligned}$$

El primer término de la suma de la ecuación (2.10) es igual a

$$\begin{aligned}
\alpha^2(B) \int_0^{\infty} u e^{-a\psi(u)} \left( -\frac{d\psi(u)}{du} \right)^2 du &= \frac{\alpha^2(B)}{a^2} \int_0^{\infty} u \left( -\frac{d\psi(u)}{du} \right) a^2 e^{-a\psi(u)} \left( -\frac{d\psi(u)}{du} \right) du \\
&= \frac{\alpha^2(B)}{a^2} \int_0^{\infty} a e^{-a\psi(u)} (\mu_1(u) - u\mu_2(u)) du \\
&= \frac{\alpha^2(B)}{a^2} \int_0^{\infty} a e^{-a\psi(u)} \mu_1(u) du - \frac{\alpha^2(B)}{a^2} I_a \\
&= \frac{\alpha^2(B)}{a^2} - \frac{\alpha^2(B)}{a^2} I_a. \tag{2.11}
\end{aligned}$$

La segunda igualdad se obtiene al integrar por partes.

Sustituyendo este resultado en (2.10) tenemos

$$\begin{aligned}
E_{\pi} \left[ \left( \tilde{P}(B) \right)^2 \right] &= \frac{\alpha^2(B)}{a^2} - \frac{\alpha^2(B)}{a^2} I_a + \frac{\alpha(B)}{a} I_a \\
&= \frac{\alpha^2(B)}{a^2} + \left[ \frac{\alpha(B)(a - \alpha(B))}{a^2} \right] I_a. \tag{2.12}
\end{aligned}$$

A partir de la ecuación (2.12) y (2.8) obtenemos el resultado del teorema

$$\text{Var}_{\pi} \left[ \tilde{P}(B) \right] = \frac{\alpha(B)(a - \alpha(B))}{a^2} I_a. \quad \square \tag{2.13}$$

La varianza de  $\tilde{P}$  nos sirve para comparar modelos no paramétricos cuando se iguala esta medida entre ellos.

## Estructura de correlación

Los criterios iniciales de la distribución aleatoria con los que se cuenta también pueden estar expresados en términos de la correlación entre probabilidades sobre diferentes subconjuntos de la línea real. El siguiente teorema nos da el principal resultado para realizar este análisis, el cual en algunos casos puede ser de ayuda para especificaciones iniciales.

**Teorema 4** Sean  $B_1$  y  $B_2$  dos subconjuntos medibles de la recta real, y sea  $C = B_1 \cap B_2$ . Si  $\tilde{P}$  se distribuye  $N$ -prior, entonces

$$Cov_{\pi} \left( \tilde{P}(B_1), \tilde{P}(B_2) \right) = \frac{a\alpha(C) - \alpha(B_1)\alpha(B_2)}{a^2} I_a.$$

### Demostración

Sean  $B_1^* = B_1 \setminus C$ ,  $B_2^* = B_2 \setminus C$  y  $C = B_1 \cap B_2$ . Entonces,

$$\tilde{P}(B_1) \tilde{P}(B_2) = \tilde{P}(B_1^*) \tilde{P}(B_2^*) + \tilde{P}(C) \tilde{P}(B_2^*) + \tilde{P}(C) \tilde{P}(B_1^*) + \tilde{P}^2(C),$$

lo cual nos lleva a que

$$\begin{aligned} E_{\pi} \left[ \tilde{P}(B_1) \tilde{P}(B_2) \right] &= E_{\pi} \left[ \tilde{P}(B_1^*) \tilde{P}(B_2^*) \right] + E_{\pi} \left[ \tilde{P}(C) \tilde{P}(B_2^*) \right] \\ &\quad + E_{\pi} \left[ \tilde{P}(C) \tilde{P}(B_1^*) \right] + E_{\pi} \left[ \tilde{P}^2(C) \right]. \end{aligned} \quad (2.14)$$

Desarrollaremos el primer término de la suma, y el resto de los términos se desarrollan de una manera similar.

$$\begin{aligned}
E_\pi \left[ \tilde{P}(B_1^*) \tilde{P}(B_2^*) \right] &= E \left[ \frac{\xi_{\alpha(B_1^*)} \xi_{\alpha(B_2^*)}}{\xi_a^2} \right] \\
&= \int_0^\infty u E \left[ e^{-u\xi_a} \xi_{\alpha(B_1^*)} \xi_{\alpha(B_2^*)} \right] du \quad \|\text{Fubini-Tonelli}\| \\
&= \int_0^\infty u E \left[ e^{-u\xi_{\alpha(B_1^*)}} \xi_{\alpha(B_1^*)} \right] E \left[ e^{-u\xi_{\alpha(B_2^*)}} \xi_{\alpha(B_2^*)} \right] \\
&\quad E \left[ e^{-u\xi_{(\alpha(B_1^*) \cup \alpha(B_2^*))^c}} \right] du \quad \|\text{Incrementos Independientes}\| \\
&= \int_0^\infty u E \left[ -\frac{d}{du} e^{-u\xi_{\alpha(B_1^*)}} \right] E \left[ -\frac{d}{du} e^{-u\xi_{\alpha(B_2^*)}} \right] E \left[ e^{-u\xi_{(\alpha(B_1^*) \cup \alpha(B_2^*))^c}} \right] du \\
&= \alpha(B_1^*) \alpha(B_2^*) \int_0^\infty u e^{a\psi(u)} \mu_1^2(u) du \quad \|\text{por (2.7)}\| \\
&= \frac{\alpha(B_1^*) \alpha(B_2^*)}{a^2} - \frac{\alpha(B_1^*) \alpha(B_2^*)}{a^2} I_a. \quad \|\text{por (2.11)}\|
\end{aligned}$$

Sustituyendo el resultado anterior y (2.12) en (2.14),

$$\begin{aligned}
E_\pi \left[ \tilde{P}(B_1) \tilde{P}(B_2) \right] &= \left[ \frac{\alpha(B_1^*) \alpha(B_2^*) + \alpha(C) \alpha(B_2^*) + \alpha(B_1^*) \alpha(C) + \alpha^2(C)}{a^2} \right] \\
&\quad - \left[ \frac{\alpha(B_1^*) \alpha(B_2^*) + \alpha(C) \alpha(B_2^*) + \alpha(B_1^*) \alpha(C) - \alpha(C)(a - \alpha(C))}{a^2} \right] I_a \\
&= \left[ \frac{(\alpha(B_1^*) + \alpha(C))(\alpha(B_2^*) + \alpha(C))}{a^2} \right] \\
&\quad - \left[ \frac{(\alpha(B_1^*) + \alpha(C))(\alpha(B_2^*) + \alpha(C)) - a\alpha(C)}{a^2} \right] I_a \\
&= \frac{\alpha(B_1) \alpha(B_2)}{a^2} + \frac{a\alpha(C) - \alpha(B_1) \alpha(B_2)}{a^2} I_a. \tag{2.15}
\end{aligned}$$

Entonces,

$$\begin{aligned}
Cov_{\pi} \left( \tilde{P}(B_1), \tilde{P}(B_2) \right) &= E_{\pi} \left[ \tilde{P}(B_1) \tilde{P}(B_2) \right] - E_{\pi} \left[ \tilde{P}(B_1) \right] E_{\pi} \left[ \tilde{P}(B_2) \right] \\
&= \frac{a\alpha(C) - \alpha(B_1)\alpha(B_2)}{a^2} I_a. \quad \square
\end{aligned} \tag{2.16}$$

Si  $B_1 \cap B_2 = \emptyset$  entonces el resultado del teorema anterior es

$$Cov_{\pi} \left( \tilde{P}(B_1), \tilde{P}(B_2) \right) = \frac{-\alpha(B_1)\alpha(B_2)}{a^2} I_a,$$

de donde el coeficiente de correlación vale

$$\begin{aligned}
\rho_{\pi} \left( \tilde{P}(B_1), \tilde{P}(B_2) \right) &= \frac{Cov_{\pi} \left( \tilde{P}(B_1), \tilde{P}(B_2) \right)}{\sqrt{Var_{\pi} \left( \tilde{P}(B_1) \right) Var_{\pi} \left( \tilde{P}(B_2) \right)}} \\
&= -\sqrt{\frac{\alpha(B_1)\alpha(B_2)}{(a - \alpha(B_1))(a - \alpha(B_2))}}.
\end{aligned}$$

De aquí, se observa que las distribuciones N-prior presentan una estructura de correlación negativa entre subconjuntos disjuntos.

Aparte de las características antes presentadas de las distribuciones N-prior, también se podría estar interesado en obtener el valor esperado de alguna función de  $\tilde{P}$ .

### Definición 3

$$Varianza, V_{\tilde{P}} = \int x^2 \tilde{P}(dx) - \left( \int x \tilde{P}(dx) \right)^2.$$

$$Diferencia Promedio de Ginni \Delta_{\tilde{P}} = \int |x - y| \tilde{P}(dx) \tilde{P}(dy).$$

El siguiente teorema nos da la esperanza de estas funcionales antes presentadas de  $\tilde{P}$ .

**Teorema 5** Si  $\tilde{P}$  se distribuye N-prior con parámetros  $(\alpha, \nu)$  y  $\int x^2 \alpha(dx) < \infty$ , entonces

$$E_\pi [V_{\tilde{P}}] = (1 - I_a) V_{\alpha/a}$$

y

$$E_\pi [\Delta_{\tilde{P}}] = (1 - I_a) \Delta_{\alpha/a},$$

donde  $V_{\alpha/a}$  y  $\Delta_{\alpha/a}$  son la varianza y la diferencia promedio de Ginni, respectivamente, de la medida de probabilidad  $\alpha/a$ .

### Demostración.

Tomemos una partición para  $m \geq 1$  de  $\mathbb{R}$  en intervalos  $B_{m,0}, B_{m,1}, \dots, B_{m,k_m+1}$  donde  $B_{m,0} = (-\infty, -R_m]$ ,  $B_{m,k_m+1} = (R_m, \infty)$  y  $\{B_{m,i} : i = 1, \dots, k_m\}$  es una partición de  $[-R_m, R_m]$ .

Sea  $R_m$  tal que  $\lim_{m \rightarrow \infty} R_m = \infty$  y sea  $\max_{1 \leq i \leq k_m} |B_{m,i}| \xrightarrow{m \rightarrow \infty} 0$  donde  $|B|$  es el tamaño del intervalo  $B$ .

Definamos una medida discretizada como

$$\alpha_m = \sum_{i=0}^{k_m+1} \alpha(B_{m,i}) \delta_{y_{m,i}},$$

donde  $y_{m,i}$  es cualquier punto interior de  $B_{m,i}$  para  $i = 0, \dots, k_m + 1$ .

Entonces la distribución N-prior con parámetros  $(\alpha_m, \nu)$  está dada por

$$\tilde{P}_m = \sum_{i=0}^{k_m+1} \tilde{P}(B_{m,i}) \delta_{m,i}.$$

Con esto obtenemos que la varianza es

$$\begin{aligned} V_{\tilde{P}_m} &= \int_{\mathbb{R}} x^2 \tilde{P}_m(dx) - \left( \int_{\mathbb{R}} x \tilde{P}_m(dx) \right)^2 \\ &= \sum_{i=0}^{k_m+1} y_{m,i}^2 \tilde{P}(B_{m,i}) - \left( \sum_{i=0}^{k_m+1} y_{m,i} \tilde{P}(B_{m,i}) \right)^2. \end{aligned}$$

Al tomar la esperanza tenemos

$$E_{\pi} [V_{\tilde{P}_m}] = \sum_{i=0}^{k_m+1} y_{m,i}^2 E_{\pi} [\tilde{P}(B_{m,i})] - E_{\pi} \left[ \left( \sum_{i=0}^{k_m+1} y_{m,i} \tilde{P}(B_{m,i}) \right)^2 \right].$$

Considerando (2.12) para el segundo término de la resta anterior

$$\begin{aligned} E_{\pi} \left[ \left( \sum_{i=0}^{k_m+1} y_{m,i} \tilde{P}(B_{m,i}) \right)^2 \right] &= E_{\pi} \left[ \sum_{i=0}^{k_m+1} y_{m,i}^2 \tilde{P}^2(B_{m,i}) + 2 \sum_i \sum_{j \neq i} y_{m,i} y_{m,j} \tilde{P}(B_{m,i}) \tilde{P}(B_{m,j}) \right] \\ &= E_{\pi} \left[ \sum_{i=0}^{k_m+1} y_{m,i}^2 \tilde{P}^2(B_{m,i}) \right] \\ &\quad + 2E \left[ \sum_i \sum_{j \neq i} y_{m,i} y_{m,j} \tilde{P}(B_{m,i}) \tilde{P}(B_{m,j}) \right]. \end{aligned} \quad (2.17)$$

El primer término de la suma anterior es

$$\begin{aligned} E_{\pi} \left[ \sum_{i=0}^{k_m+1} y_{m,i}^2 \tilde{P}^2(B_{m,i}) \right] &= \sum_{i=0}^{k_m+1} y_{m,i}^2 E_{\pi} [\tilde{P}^2(B_{m,i})] \\ &= \sum_{i=0}^{k_m+1} y_{m,i}^2 \frac{\alpha^2(B_{m,i})}{a^2} + I_a \sum_{i=0}^{k_m+1} y_{m,i}^2 \frac{\alpha(B_{m,i})(a - \alpha(B_{m,i}))}{a^2} \\ &= \sum_{i=0}^{k_m+1} y_{m,i}^2 \frac{\alpha^2(B_{m,i})}{a^2} + I_a \sum_{i=0}^{k_m+1} y_{m,i}^2 \frac{\alpha(B_{m,i})}{a} \\ &\quad - I_a \sum_{i=0}^{k_m+1} y_{m,i}^2 \frac{\alpha^2(B_{m,i})}{a^2}. \end{aligned} \quad (2.18)$$

El segundo término de la ecuación (2.17) es igual a

$$\begin{aligned} E \left[ \sum_i \sum_{j \neq i} y_{m,i} y_{m,j} \tilde{P}(B_{m,i}) \tilde{P}(B_{m,j}) \right] &= \sum_i \sum_{j \neq i} y_{m,i} y_{m,j} E_{\pi} [\tilde{P}(B_{m,i}) \tilde{P}(B_{m,j})] \\ &= \sum_i \sum_{j \neq i} y_{m,i} y_{m,j} \left( \frac{\alpha(B_{m,i}) \alpha(B_{m,j})}{a^2} - \frac{\alpha(B_{m,i}) \alpha(B_{m,j})}{a^2} I_a \right). \end{aligned} \quad (2.19)$$

La última igualdad se debe a (2.15). Sustituyendo (2.18) y (2.19) en (2.17) queda

$$\begin{aligned}
E_\pi \left[ \left( \sum_{i=0}^{k_m+1} y_{m,i} \tilde{P}(B_{m,i}) \right)^2 \right] &= (1 - I_a) \sum_{i=0}^{k_m+1} y_{m,i}^2 \frac{\alpha^2(B_{m,i})}{a^2} + I_a \sum_{i=0}^{k_m+1} y_{m,i}^2 \frac{\alpha(B_{m,i})}{a} \\
&+ (1 - I_a) 2 \sum_i \sum_{j \neq i} y_{m,i} y_{m,j} \frac{\alpha(B_{m,i}) \alpha(B_{m,j})}{a^2} \\
&= (1 - I_a) \left( \sum_{i=0}^{k_m+1} y_{m,i}^2 \frac{\alpha^2(B_{m,i})}{a^2} + 2 \sum_i \sum_{j \neq i} y_{m,i} y_{m,j} \frac{\alpha(B_{m,i}) \alpha(B_{m,j})}{a^2} \right) \\
&+ I_a \sum_{i=0}^{k_m+1} y_{m,i}^2 \frac{\alpha(B_{m,i})}{a} \\
&= (1 - I_a) \left[ \sum_{i=0}^{k_m+1} y_{m,i} \frac{\alpha(B_{m,i})}{a} \right]^2 + I_a \sum_{i=0}^{k_m+1} y_{m,i}^2 \frac{\alpha(B_{m,i})}{a}.
\end{aligned}$$

Aplicando este resultado a la esperanza de la función denominada varianza

$$\begin{aligned}
E_\pi [V_{\tilde{P}_m}] &= \sum_{i=0}^{k_m+1} y_{m,i}^2 \frac{\alpha(B_{m,i})}{a} - (1 - I_a) \left[ \sum_{i=0}^{k_m+1} y_{m,i} \frac{\alpha(B_{m,i})}{a} \right]^2 \\
&- I_a \sum_{i=0}^{k_m+1} y_{m,i}^2 \frac{\alpha(B_{m,i})}{a} \\
&= (1 - I_a) \left\{ \sum_{i=0}^{k_m+1} y_{m,i}^2 \frac{\alpha(B_{m,i})}{a} - \left[ \sum_{i=0}^{k_m+1} y_{m,i} \frac{\alpha(B_{m,i})}{a} \right]^2 \right\}.
\end{aligned}$$

Si obtenemos el límite cuando  $m \rightarrow \infty$ , tenemos

$$\lim_{m \rightarrow \infty} E_\pi [V_{\tilde{P}_m}] = (1 - I_a) V_{\frac{\alpha}{a}}.$$

Se puede demostrar que  $E_\pi [V_{\tilde{P}}] = \lim_{m \rightarrow \infty} E_\pi [V_{\tilde{P}_m}]$ .

Ahora, pasemos a la segunda igualdad del teorema. Aplicando la medida discretizada a la diferencia promedio de Ginni,

$$\begin{aligned}\Delta_{\tilde{P}_m} &= \int_{\mathbb{R} \times \mathbb{R}} |x - y| \tilde{P}_m(dx) \tilde{P}_m(dy) \\ &= \sum_{i=0}^{k_m+1} \sum_{j=0}^{k_m+1} |y_{m,i} - y_{m,j}| \tilde{P}(B_{m,i}) \tilde{P}(B_{m,j}).\end{aligned}$$

Considerando (2.15) tenemos:

$$\begin{aligned}E_\pi [\Delta_{\tilde{P}_m}] &= E_\pi \left[ \sum_{i=0}^{k_m+1} \sum_{j=0}^{k_m+1} |y_{m,i} - y_{m,j}| \tilde{P}(B_{m,i}) \tilde{P}(B_{m,j}) \right] \\ &= \sum_{i=0}^{k_m+1} \sum_{j=0}^{k_m+1} |y_{m,i} - y_{m,j}| E_\pi \left[ \tilde{P}(B_{m,i}) \tilde{P}(B_{m,j}) \right] \\ &= \sum_{i=0}^{k_m+1} \sum_{j=0}^{k_m+1} |y_{m,i} - y_{m,j}| \left( \frac{\alpha(B_{m,i}) \alpha(B_{m,j})}{a^2} + \frac{a\alpha(\emptyset) - \alpha(B_{m,i}) \alpha(B_{m,j})}{a^2} I_a \right) \\ &= (1 - I_a) \sum_{i=0}^{k_m+1} \sum_{j=0}^{k_m+1} |y_{m,i} - y_{m,j}| \left( \frac{\alpha(B_{m,i})}{a} \frac{\alpha(B_{m,j})}{a} \right).\end{aligned}$$

Entonces,  $\lim_{m \rightarrow \infty} E_\pi [\Delta_{\tilde{P}_m}] = (1 - I_a) \Delta_{\frac{\alpha}{a}}$ . Y se puede demostrar que  $E_\pi [\Delta_{\tilde{P}}] = \lim_{m \rightarrow \infty} E_\pi [\Delta_{\tilde{P}_m}]$   $\square$

El modelo con el cual se trabajará en esta tesis considera eventos intercambiables tal y como se hace en estadística Bayesiana. Este modelo consiste en:

Sean  $X_1, \dots, X_n$  variables aleatorias intercambiables con valores en  $(\mathbb{X}, \mathcal{X})$  que tienen una distribución común  $\tilde{P}$ , la cual a su vez es aleatoria, definida en  $(\mathbb{F}, \mathcal{F})$  y tal que tiene una distribución N-Prior, es decir

$$X_i | \tilde{P} \stackrel{\text{ind}}{\sim} \tilde{P} \quad i = 1, \dots, n,$$

$$\tilde{P} \sim \text{N-Prior}.$$



A esta estructura se le conoce como modelo jerárquico, usualmente se utiliza para comprender la estructura de dependencia en el modelo de una forma “más clara”.

## Distribución predictiva

Una de las metas importantes de la inferencia es la predicción de valores futuros de una cantidad aleatoria basada en una muestra obtenida. Bajo las distribuciones N-prior se podrá observar, con los siguientes teoremas, que las distribuciones predictivas tienen formas intuitivas, dado que consisten en una combinación lineal de la media inicial del proceso y de una versión ponderada de la función de distribución empírica.

**Teorema 6** Si  $\tilde{P}$  se distribuye N-prior con parámetros  $(\alpha, \nu)$ , entonces

$$Pr(X_2 \in B | X_1) = \frac{\alpha(B)}{a} [a - I_a] + \delta_{X_1}(B) I_a$$

para cualquier conjunto medible  $B$ , donde  $I_a = a \int_{\mathbb{R}^+} u e^{-a\psi(u)} \int_{\mathbb{R}^+} v^2 e^{-uv} \nu(dv) du$ , que se interpreta como  $Pr(X_2 = X_1)$ .

### Demostración

Supóngase que  $X_1 = x_1$  y sea  $C_\epsilon = (x_1 - \epsilon, x_1]$  con  $\epsilon > 0$ . Dado que  $\nu(\mathbb{R}^+) = \infty$  entonces  $\tilde{P}(C_\epsilon) > 0$  c.s.

Seleccionemos un  $\epsilon > 0$  tal que  $C_\epsilon \subset B$  o  $C_\epsilon \cap B = \emptyset$ , con lo cual tenemos que,

$$\begin{aligned} Pr(X_2 \in B | X_1 \in C_\epsilon) &= \frac{E_\pi \left[ \tilde{P}(B) \tilde{P}(C_\epsilon) \right]}{E_\pi \left[ \tilde{P}(C_\epsilon) \right]} \\ &= \frac{(\alpha(B) \alpha(C_\epsilon) + a \alpha(C_\epsilon \cap B) I_a - \alpha(C_\epsilon) \alpha(B) I_a) / a^2}{(\alpha(C_\epsilon)) / a} \\ &= \frac{\alpha(B) \alpha(C_\epsilon) + [a \alpha(C_\epsilon \cap B) - \alpha(C_\epsilon) \alpha(B)] I_a}{a \alpha(C_\epsilon)}. \end{aligned}$$

Entonces,

$$\begin{aligned} Pr(X_2 \in B | X_1 \in C_\epsilon) &= \begin{cases} \frac{\alpha(B)}{a} (1 - I_a) + I_a & \text{si } C_\epsilon \subset B \text{ (es decir } x_1 \in B) \\ \frac{\alpha(B)}{a} (1 - I_a) & \text{si } C_\epsilon \cap B = \emptyset \text{ (es decir } x_1 \notin B) \end{cases} \\ &= \frac{\alpha(B)}{a} (1 - I_a) + \delta_{X_1}(B) I_a. \quad \square \end{aligned}$$

El teorema anterior nos da una característica interesante de la predicción a un paso ya que da una idea de las expresiones que se usan para calcular las propiedades de segundo orden de las distribuciones N-prior.

Antes de dar una expresión general para la distribución predictiva, definamos a

$$\prod_k^{(n)}(n_1, \dots, n_k) := \int_{\mathbb{R}^+} u^{n-1} e^{-\alpha\psi(u)} \mu_{n_1}(u) \cdots \mu_{n_k}(u) du, \quad (2.20)$$

donde  $k$  denota el número de observaciones distintas en la muestra de tamaño  $n$ ,  $n_j$  es la cardinalidad del conjunto  $j$ -ésimo de observaciones que coinciden ( $j = 1, \dots, k$ ), y  $n = n_1 + \cdots + n_k$  es el número total de observaciones en la muestra.

La notación en (2.20) se hereda de Pitman [1995], en donde se le da la interpretación de probabilidades sobre particiones  $n_1, n_2, \dots, n_k$  de un entero  $n$  y juegan un papel importante es la teoría de Fragmentación y Coagulación, ver Bertoin y Yor [2001]

El siguiente teorema nos dará la forma de obtener la predicción a un paso para todo  $n$  al considerar las distribuciones N-Prior.

**Teorema 7** *Si  $\tilde{P}$  se distribuye N-prior con parámetros  $(\alpha, \nu)$  y  $\alpha$  es no nulo, entonces las distribuciones predictivas son de la forma*

$$Pr(X_{n+1} \in B | X_1, \dots, X_n) = w^{(n)} \frac{\alpha(B)}{a} + \frac{1}{n} \sum_{j=1}^k w_j^{(n)} \delta_{X_j^*}(B),$$

donde  $X_1^*, \dots, X_k^*$  denotan los  $k$  distintos valores dentro de la muestra,  $n_j > 0$  denota al número de observaciones que son iguales a  $X_j^*$  para  $j = 1, \dots, k$ , y

$$w^{(n)} = \frac{a \prod_{k+1}^{(n+1)} (n_1, \dots, n_k, 1)}{n \prod_k^{(n)} (n_1, \dots, n_k)}, \quad (2.21)$$

$$w_j^{(n)} = \frac{\prod_k^{(n+1)} (n_1, \dots, n_j + 1, \dots, n_k)}{\prod_k^{(n)} (n_1, \dots, n_k)}. \quad (2.22)$$

A las cantidades  $w^{(n)}$  y  $w_j^{(n)}$ , las llamaremos los pesos correspondientes a la predictiva de la distribución N-prior. Más adelante veremos que una especificación en particular de la medida de Lévy conlleva a simplificaciones de  $w^{(n)}$  y  $w_j^{(n)}$ .

**Demostración** Sean  $C_1, C_2, \dots, C_k \subset \mathbb{R}$  disjuntos. Entonces

$$\begin{aligned} Pr(X_{n+1} \in B | (X_1, X_2, \dots, X_n) \in \times_{j=1}^k C_j^{n_j}) &= \frac{Pr(X_{n+1} \in B, (X_1, X_2, \dots, X_n) \in \times_{j=1}^k C_j^{n_j})}{Pr((X_1, X_2, \dots, X_n) \in \times_{j=1}^k C_j^{n_j})} \\ &= \frac{E \left[ \frac{\xi_{\alpha(B)}}{\xi_a} \frac{\xi_{\alpha_1}^{n_1}}{\xi_a} \dots \frac{\xi_{\alpha_k}^{n_k}}{\xi_a} \right]}{E \left[ \frac{\xi_{\alpha_1}^{n_1}}{\xi_a} \dots \frac{\xi_{\alpha_k}^{n_k}}{\xi_a} \right]} \\ &= \frac{E [\xi_a^{-n-1} \xi_{\alpha(B)} \xi_{\alpha_1}^{n_1} \dots \xi_{\alpha_k}^{n_k}]}{E [\xi_a^{-n} \xi_{\alpha_1}^{n_1} \dots \xi_{\alpha_k}^{n_k}]}, \end{aligned} \quad (2.23)$$

donde  $\alpha_j = \alpha(C_j)$ .

Sea  $C_{k+1} = \left( \bigcup_{j=1}^k C_j \right)^c$  y observemos que  $\xi_{\alpha(B)} = \sum_{j=1}^k \xi_{\alpha(B \cap C_j)} + \xi_{\alpha(B \cap C_{k+1})}$ . Considerando (2.3) y (2.6), el numerador de (2.23) resulta en

$$\begin{aligned}
& \frac{1}{n} \left\{ \int \xi_a^{-n-1} \xi_{\alpha(B)} \xi_{\alpha_1}^{n_1} \xi_{\alpha_2}^{n_2} \cdots \xi_{\alpha_k}^{n_k} P_\xi d\xi \right\} \\
&= \frac{1}{n} \left\{ \int \xi_a^{-n-1} \xi_{\alpha(B \cap C_{k+1})} \xi_{\alpha_1}^{n_1} \xi_{\alpha_2}^{n_2} \cdots \xi_{\alpha_k}^{n_k} P_\xi d\xi + \sum_{j=1}^k \int \xi_a^{-n-1} \xi_{\alpha(B \cap C_j)} \xi_{\alpha_1}^{n_1} \xi_{\alpha_2}^{n_2} \cdots \xi_{\alpha_k}^{n_k} P_\xi d\xi \right\} \\
&= \frac{1}{n} \left\{ \int_0^\infty u^n E \left[ e^{-u\xi_a} \xi_{\alpha(B \cap C_{k+1})} \xi_{\alpha_1}^{n_1} \xi_{\alpha_2}^{n_2} \cdots \xi_{\alpha_k}^{n_k} \right] du \right. \\
&\quad \left. + \sum_{j=1}^k \int_0^\infty u^n E \left[ e^{-u\xi_a} \xi_{\alpha(B \cap C_j)} \xi_{\alpha_1}^{n_1} \xi_{\alpha_2}^{n_2} \cdots \xi_{\alpha_k}^{n_k} \right] du \right\}.
\end{aligned} \tag{2.24}$$

Definamos dos funciones que nos serán de ayuda para esta demostración.

Para cualquier  $\rho > 0$ , sea

$$V_\rho^{(n)}(u) = \left\{ (-1)^n \frac{d^n}{du^n} e^{-\rho\psi(u)} \right\} e^{\rho\psi(u)}, \quad n = 1, 2, \dots,$$

donde  $V_\rho^{(0)}(u) := 1$ , y sea

$$\Delta_\rho^{(n)}(u) = \sum_{i=0}^{n-1} \binom{n-1}{i} \mu_{n-i}(u) V_\rho^{(i)}(u). \tag{2.25}$$

Entonces encontramos que

$$V_\rho^{(n)}(u) = \rho \sum_{i=0}^{n-1} \binom{n-1}{i} \mu_{n-i}(u) V_\rho^{(i)}(u) = \rho \Delta_\rho^{(n)}(u), \tag{2.26}$$

lo cual se demostrará al final. Desarrollando el primer término de (2.24).

$$\begin{aligned}
& \int_0^\infty u^n E \left[ e^{-u\xi_a} \xi_{\alpha(B \cap C_{k+1})} \xi_{\alpha_1}^{n_1} \xi_{\alpha_2}^{n_2} \cdots \xi_{\alpha_k}^{n_k} \right] du \\
&= \int_0^\infty u^n E \left[ e^{-u\xi_{\sum_{i=1}^k \alpha_i + \alpha_{k+1}}} \xi_{\alpha(B \cap C_{k+1})} \xi_{\alpha_1}^{n_1} \xi_{\alpha_2}^{n_2} \cdots \xi_{\alpha_k}^{n_k} \right] du \\
&= \int_0^\infty u^n E \left[ e^{-u\xi_{\alpha_1}} \xi_{\alpha_1}^{n_1} \right] E \left[ e^{-u\xi_{\alpha_2}} \xi_{\alpha_2}^{n_2} \right] \cdots E \left[ e^{-u\xi_{\alpha_k}} \xi_{\alpha_k}^{n_k} \right] E \left[ e^{-u\xi_{\alpha_{k+1}}} \xi_{\alpha(B \cap C_{k+1})} \right] du \\
&= \int_0^\infty u^n E \left[ (-1)^{n_1} \frac{d^{n_1}}{du^{n_1}} e^{-u\xi_{\alpha_1}} \right] E \left[ (-1)^{n_2} \frac{d^{n_2}}{du^{n_2}} e^{-u\xi_{\alpha_2}} \right] \\
&\quad \cdots E \left[ (-1)^{n_k} \frac{d^{n_k}}{du^{n_k}} e^{-u\xi_{\alpha_k}} \right] E \left[ e^{-u\xi_{\alpha_{k+1}}} \xi_{\alpha(B \cap C_{k+1})} \right] du.
\end{aligned}$$

La segunda igual se debe a que se tienen incrementos independientes.

Aplicando (2.6) a la ecuación anterior,

$$\begin{aligned}
& \int_0^\infty u^n E \left[ e^{-u\xi_a} \xi_{\alpha(B \cap C_{k+1})} \xi_{\alpha_1}^{n_1} \xi_{\alpha_2}^{n_2} \cdots \xi_{\alpha_k}^{n_k} \right] du \\
&= \int_0^\infty u^n (-1)^{n_1} \frac{d^{n_1}}{du^{n_1}} \left( e^{-\alpha_1 \psi(u)} \right) (-1)^{n_2} \frac{d^{n_2}}{du^{n_2}} \left( e^{-\alpha_2 \psi(u)} \right) \\
&\quad \cdots (-1)^{n_k} \frac{d^{n_k}}{du^{n_k}} \left( e^{\alpha_k \psi(u)} \right) E \left[ e^{-u\xi_{\alpha_{k+1}}} \xi_{\alpha(B \cap C_{k+1})} \right] du.
\end{aligned}$$

Al aplicar (2.26) se tiene

$$\begin{aligned}
& \int_0^\infty u^n E \left[ e^{-u\xi_a} \xi_{\alpha(B \cap C_{k+1})} \xi_{\alpha_1}^{n_1} \xi_{\alpha_2}^{n_2} \cdots \xi_{\alpha_k}^{n_k} \right] du \\
&= \int_0^\infty u^n \left( \prod_{i=1}^k \alpha_i \right) \prod_{j=1}^k \Delta_{\alpha_j}^{(n_j)} e^{-\psi(u) \sum_{i=1}^k \alpha_i} E \left[ e^{-u\xi_{\alpha_{k+1}}} \xi_{\alpha(B \cap C_{k+1})} \right] du \\
&= \int_0^\infty u^n \left( \prod_{i=1}^k \alpha_i \right) \prod_{j=1}^k \Delta_{\alpha_j}^{(n_j)} \alpha(B \cap C_{k+1}) \mu_1(u) e^{-\psi(u) \sum_{i=1}^k \alpha_i} e^{-\alpha_{k+1} \psi(u)} du \\
&= \alpha(B \cap C_{k+1}) \left( \prod_{i=1}^k \alpha_i \right) \int_0^\infty u^n e^{-\alpha \psi(u)} \mu_1(u) \prod_{j=1}^k \Delta_{\alpha_j}^{(n_j)} du. \tag{2.27}
\end{aligned}$$

La segunda desigualdad de la ecuación anterior se obtiene al aplicar (2.7).

Ahora, pasemos al desarrollo del segundo término de (2.24) de la misma manera en que lo hicimos para el primer término.

$$\begin{aligned}
& \sum_{j=1}^k \int_0^{\infty} u^n E \left[ e^{-u\xi_a} \xi_{\alpha(B \cap C_j)} \xi_{\alpha_1}^{n_1} \xi_{\alpha_2}^{n_2} \cdots \xi_{\alpha_k}^{n_k} \right] du \\
&= \sum_{j=1}^k \int_0^{\infty} u^n \left( \prod_{i \neq j} \alpha_i \right) e^{-(a-\alpha_j)\psi(u)} \left( \prod_{i \neq j} \Delta_{\alpha_i}^{(n_i)}(u) \right) E \left[ e^{-u\xi_{\alpha_j}} \xi_{\alpha_j}^{n_j} \xi_{\alpha(B \cap C_j)} \right] du \\
&= \sum_{j=1}^k \int_0^{\infty} u^n \left( \prod_{i \neq j} \alpha_i \right) e^{-(a-\alpha_j)\psi(u)} \left( \prod_{i \neq j} \Delta_{\alpha_i}^{(n_i)}(u) \right) \\
& \quad E \left[ e^{-u\xi_{\alpha(B \cap C_j) + \alpha(B^c \cap C_j)}} \xi_{\alpha(B \cap C_j) + \alpha(B^c \cap C_j)}^{n_j} \xi_{\alpha(B \cap C_j)} \right] du.
\end{aligned}$$

Dado que tenemos incrementos independientes podemos desarrollar a  $\xi_{\alpha(B \cap C_j) + \alpha(B^c \cap C_j)}^{n_j}$  como

$$\xi_{\alpha(B \cap C_j) + \alpha(B^c \cap C_j)}^{n_j} = \sum_{i=0}^{n_j} \binom{n_j}{i} \xi_{\alpha(B \cap C_j)}^i \xi_{\alpha(B^c \cap C_j)}^{n_j - i}.$$

Aplicando la igualdad anterior,

$$\begin{aligned}
& \sum_{j=1}^k \int_0^{\infty} u^n E \left[ e^{-u\xi_a} \xi_{\alpha(B \cap C_j)} \xi_{\alpha_1}^{n_1} \xi_{\alpha_2}^{n_2} \cdots \xi_{\alpha_k}^{n_k} \right] du \\
&= \sum_{j=1}^k \int_0^{\infty} u^n \left( \prod_{i \neq j} \alpha_i \right) e^{-(a-\alpha_j)\psi(u)} \left( \prod_{i \neq j} \Delta_{\alpha_i}^{(n_i)}(u) \right) \\
& \quad \sum_{i=0}^{n_j} \binom{n_j}{i} E \left[ e^{-u\xi_{\alpha(B \cap C_j)}} \xi_{\alpha(B \cap C_j)}^{i+1} \right] E \left[ e^{-u\xi_{\alpha(B^c \cap C_j)}} \xi_{\alpha(B^c \cap C_j)}^{n_j - i} \right]
\end{aligned}$$

$$\begin{aligned}
&= \sum_{j=1}^k \int_0^\infty u^n \left( \prod_{i \neq j} \alpha_i \right) e^{-(a-\alpha_j)\psi(u)} \left( \prod_{i \neq j} \Delta_{\alpha_i}^{(n_i)}(u) \right) \\
&\quad \sum_{i=0}^{n_j} \binom{n_j}{i} (-1)^{i+1} \frac{d^{i+1}}{du^{i+1}} e^{-\psi(u)\alpha(B \cap C_j)} (-1)^{n_j-i} \frac{d^{n_j-i}}{du^{n_j-i}} e^{-\psi(u)\alpha(B^c \cap C_j)}.
\end{aligned}$$

Usando las funciones definidas en (2.25) y (2.26), obtenemos

$$\begin{aligned}
&\sum_{j=1}^k \int_0^\infty u^n E \left[ e^{-u\xi_a} \xi_{\alpha(B \cap C_j)} \xi_{\alpha_1}^{n_1} \xi_{\alpha_2}^{n_2} \dots \xi_{\alpha_k}^{n_k} \right] du \\
&= \sum_{j=1}^k \int_0^\infty u^n \left( \prod_{i \neq j} \alpha_i \right) e^{-(a-\alpha_j)\psi(u)} \left( \prod_{i \neq j} \Delta_{\alpha_i}^{(n_i)}(u) \right) \\
&\quad \sum_{i=0}^{n_j} \binom{n_j}{i} \alpha(B \cap C_j) \Delta_{\alpha(B \cap C_j)}^{(i+1)}(u) e^{-\alpha(B \cap C_j)\psi(u)} \alpha(B^c \cap C_j) \Delta_{\alpha(B^c \cap C_j)}^{(n_j-i)}(u) e^{-\alpha(B^c \cap C_j)\psi(u)} \\
&= \sum_{j=1}^k \int_0^\infty u^n \left( \prod_{i \neq j} \alpha_i \right) e^{-a\psi(u)} \left( \prod_{i \neq j} \Delta_{\alpha_i}^{(n_i)}(u) \right) \alpha(B \cap C_j) \sum_{i=0}^{n_j} \binom{n_j}{i} \Delta_{\alpha(B \cap C_j)}^{(i+1)}(u) V_{\alpha(B^c \cap C_j)}^{(n_j-i)}.
\end{aligned} \tag{2.28}$$

Si desarrollamos de la misma manera el denominador de (2.23), y lo unimos a (2.27) y (2.28), obtenemos

$$\begin{aligned}
&Pr(X_{n+1} \in B | (X_1, X_2, \dots, X_n) \in \times_{j=1}^k C_j^{n_j}) \\
&= \frac{\alpha(B \cap C_{k+1}) \int_0^\infty u^n e^{-a\psi(u)} \mu_1(u) \prod_{j=1}^k \Delta_{\alpha_j}^{(n_j)}(u) du}{n \int_0^\infty u^{n-1} e^{-a\psi(u)} \prod_{j=1}^k \Delta_{\alpha_j}^{(n_j)}(u) du} \\
&\quad + \frac{1}{n} \sum_{j=1}^k \left( \frac{\alpha(B \cap C_j) \sum_{i=0}^{n_j} \binom{n_j}{i} \int_0^\infty u^n e^{-a\psi(u)} \left( \prod_{i \neq j} \Delta_{\alpha_i}^{(n_i)}(u) \right) \Delta_{\alpha(B \cap C_j)}^{(i+1)}(u) V_{\alpha(B^c \cap C_j)}^{(n_j-i)}(u) du}{\alpha_j \sum_{i=0}^{n_j} \binom{n_j}{i} \int_0^\infty u^{n-1} e^{-a\psi(u)} \prod_{j=1}^k \Delta_{\alpha_j}^{(n_j)}(u) du} \right).
\end{aligned}$$

Sea  $C_j = C_{j,\epsilon} = (X_j^* - \epsilon, X_j^*]$  y dado que  $\alpha$  es no nula:

$$\Delta_{\alpha_i}^{(n_i)} = \mu_{n_i} + o(\alpha_i) \quad \text{si } \epsilon \rightarrow 0$$

y

$$\delta_{X_j^*}(B) V_{\alpha(B^c \cap C_j)}^{(n_j-i)}(u) = 0 \quad i = 0, 1, \dots, n_j - 1.$$

Entonces, si  $\epsilon \rightarrow 0$ ,

$$\begin{aligned} & Pr(X_{n+1} \in B | (X_1, X_2, \dots, X_n) \in \times_{j=1}^k C_j^{n_j}) \\ &= \frac{\alpha(B)}{n} \frac{\int_0^\infty u^n e^{-\alpha\psi(u)} \mu_1(u) \prod_{j=1}^k \mu_{n_j}(u) du}{\int_0^\infty u^n e^{-\alpha\psi(u)} \prod_{j=1}^k \mu_{n_j}(u) du} \\ & \quad + \frac{1}{n} \sum_{j=1}^k \delta_{X_j^*}(B) \frac{\int_0^\infty u^n e^{-\alpha\psi(u)} \prod_{i \neq j} \mu_{n_i}(u) \mu_{n_j+1}(u) du}{\int_0^\infty u^n e^{-\alpha\psi(u)} \prod_{j=1}^k \mu_{n_j}(u) du} \\ &= \frac{\alpha(B)}{a} \left[ \frac{a \prod_{k+1}^{(n+1)} (n_1, \dots, n_k, 1)}{n \prod_k^{(n)} (n_1, \dots, n_k)} \right] \\ & \quad + \frac{1}{n} \sum_{j=1}^k \delta_{X_j^*}(B) \frac{\prod_k^{(k+1)} (n_1, n_2, \dots, n_j + 1, \dots, n_k)}{\prod_k^{(n)} (n_1, \dots, n_k)}. \end{aligned}$$

Ahora, solo nos queda demostrar (2.26), lo cual se hará por inducción.

- Para  $n = 1$

$$V_\rho^{(1)}(u) = \left\{ -\frac{d}{du} e^{-\rho\psi(u)} \right\} e^{\rho\psi(u)} = \rho \frac{d}{du} \psi(u) = \rho \mu_1(u),$$



$$\rho \Delta_{\rho}^{(1)}(u) = \rho \binom{0}{0} \mu_1(u) V_{\rho}^{(0)}(u) = \rho \mu_1(u).$$

Por lo tanto se cumple para  $n = 1$ .

- Supóngase que se cumple para  $n = k$ , es decir

$$V_{\rho}^{(k)}(u) = \rho \sum_{i=0}^{k-1} \binom{k-1}{i} \mu_{k-1}(u) V_{\rho}^{(i)}(u) = \rho \Delta_{\rho}^{(k)}(u).$$

- Para  $n = k + 1$

$$\begin{aligned} V_{\rho}^{(k+1)}(u) &= \left\{ (-1)^{k+1} \frac{d^{k+1}}{du^{k+1}} e^{-\rho\psi(u)} \right\} e^{\rho\psi(u)} = \left\{ (-1)^k \frac{d^k}{du^k} \left[ \rho e^{-\rho\psi(u)} \frac{d}{du} \psi(u) \right] \right\} e^{\rho\psi(u)} \\ &= \rho \left\{ (-1)^k \frac{d^k}{du^k} \left[ e^{-\rho\psi(u)} \frac{d}{du} \psi(u) \right] \right\} e^{\rho\psi(u)} \\ &= \rho \left\{ \sum_{i=0}^k \binom{k}{i} \left[ (-1)^i \frac{d^i}{du^i} e^{-\rho\psi(u)} \right] \left[ (-1)^{k-i} \frac{d^{k-i}}{du^{k-i}} \psi(u) \right] \right\} e^{\rho\psi(u)} \\ &= \rho \sum_{i=0}^k \binom{k}{i} \left[ (-1)^i \frac{d^i}{du^i} e^{-\rho\psi(u)} \right] e^{\rho\psi(u)} \mu_{k-i}(u) \\ &= \rho \sum_{i=0}^k \binom{k}{i} V_{\rho}^{(i)}(u) \mu_{k-i}(u) = \rho \Delta_{\rho}^{(k+1)}(u). \quad \square \end{aligned}$$

Con esto queda demostrado el teorema de predicción a un paso para todo  $n$ .

El proceso Dirichlet es el único miembro de las distribuciones N-Prior que es conjugado, lo cual se demuestra en el siguiente teorema.

**Teorema 8** Sea  $\mathbb{F}$  la clase de Distribuciones N-Prior y sea  $\tilde{P} \in \mathbb{F}$ . Entonces la distribución final de  $\tilde{P}$ , dada una muestra  $X_1, \dots, X_n$ , está en  $\mathbb{F}$  si y sólo si  $\tilde{P}$  es un proceso Dirichlet.

### Demostración

Supóngase que la distribución final de  $\tilde{P}$  dado  $X_1$ , es N-Prior con parámetros  $(\alpha_1, \nu_1)$ .

Sea  $\tilde{P}_1$  la distribución final, entonces por el teorema 2

$$E_{\pi} \left[ \tilde{P}_1(B) \right] = \frac{\alpha_1(B)}{a_1},$$

donde  $a_1 = \alpha_1(\mathbb{R})$ .

Por el teorema 6 tenemos

$$E_{\pi} \left[ \tilde{P}(B) | X_1 \right] = \frac{\alpha(B)}{a} (a - I_a) + \delta_{X_1}(B) I_a = E \left[ \tilde{P}_1 \right].$$

Entonces,

$$\frac{\alpha_1(B)}{a_1} = \frac{\alpha(B)}{a} (1 - I_a) + \delta_{X_1}(B) I_a$$

es decir,

$$\alpha_1 = a_1 \left[ \frac{\alpha}{a} (1 - I_a) + \delta_{X_1} I_a \right].$$

Aplicando de nuevo el teorema 6

$$\begin{aligned} E_{\pi} \left[ \tilde{P} | X_1, X_2 \right] &= \frac{\alpha_1(B)}{a_1} (1 - I_{a_1}) + \delta_{X_2} I_{a_1} \\ &= \left( \frac{\alpha(B)}{a} (1 - I_a) + \delta_{X_1}(B) I_a \right) (1 - I_{a_1}) + \delta_{X_2}(B) I_{a_1} \\ &= \frac{\alpha(B)}{a} (1 - I_a) (1 - I_{a_1}) + \delta_{X_1} I_a (1 - I_{a_1}) + \delta_{X_2}(B) I_{a_1}. \end{aligned}$$

Sea  $\tilde{P}_2$  una distribución final, entonces

$$E_{\pi} \left[ \tilde{P}_2(B) \right] = \frac{\alpha_2(B)}{a_2}.$$

De la ecuación anterior se tiene

$$\alpha_2 = a_2 \left[ \frac{\alpha(B)}{a} (1 - I_a) (1 - I_{a_1}) + \delta_{X_1}(B) I_a (1 - I_{a_1}) + \delta_{X_2}(B) I_{a_1} \right].$$

Si este proceso lo hiciéramos iterativamente  $n$  veces obtendríamos que

$$E_\pi \left[ \tilde{P}(B) | X_1, X_2, \dots, X_n \right] = \frac{\alpha(B)}{a} \prod_{i=0}^{n-1} (1 - I_{a_i}) + \sum_{i=1}^n \left( \delta_{X_i}(B) I_{a_{i-1}} \prod_{k=i}^{n-1} (1 - I_{a_k}) \right).$$

Usando la igualdad

$$I_{a_{n-1}} = \frac{I_a}{1 + (n-1)I_a} \text{ para } n \geq 1, \quad (2.29)$$

que se demostrará al final, se obtiene

$$E_\pi \left[ \tilde{P}(B) | X_1, X_2, \dots, X_n \right] = \frac{1}{1 + (n-1)I_a} \left\{ \frac{\alpha(B)}{a} (1 - I_a) + I_a \sum_{i=1}^n \delta_{x_i}(B) \right\}.$$

Observemos que  $I_a$  es una constante en  $[0, 1]$ , por lo que existe una constante,  $c \geq 0$  tal que  $I_a = \frac{1}{1+c}$ . Entonces

$$E_\pi \left[ \tilde{P}(B) | X_1, X_2, \dots, X_n \right] = \frac{c}{c+n} \frac{\alpha(B)}{a} + \frac{1}{c+n} \sum_{i=1}^n \delta_{x_i}(B).$$

Así como lo demostraron Regazzini (1978) y Lo (1991), la ecuación anterior también caracteriza el proceso Dirichlet, así que  $c = a$  y la demostración queda completa.

Ahora demostraremos (2.29) por inducción. Empezaremos demostrando que es cierto para  $n = 2$ .

Sean  $\tilde{P}_{x_1}$  y  $\tilde{P}_{x_2}$  dos distribuciones iniciales que siguen un proceso Dirichlet con parámetros  $(\alpha_{x_1}, \nu_1)$  y  $(\alpha_{x_2}, \nu_2)$  respectivamente. Además tenemos por el teorema 2 y el teorema 6 que

$$E_\pi \left[ \tilde{P}_{x_i}(B) | X_j \right] = \frac{\alpha_{x_1}}{a_1} = \frac{\alpha(B)}{a} (1 - I_a) + \delta_{x_i}(B) I_a; \quad i = 1, 2; \quad j \neq i,$$

dado que  $X_1$  y  $X_2$  son intercambiables. Entonces, aplicando el teorema 6

$$\begin{aligned}
E_\pi \left[ \tilde{P}_{x_1}(B) | X_2 \right] &= \frac{\alpha_{x_1}}{a_{x_1}} (1 - I_{a_1}) + \delta_{X_2}(B) I_{x_1} \\
&= \frac{\alpha(B)}{a} (1 - I_a) (1 - I_{a_1}) + \delta_{X_1}(B) I_a (1 - I_{a_1}) + \delta_{X_2} I_{a_1} \\
&= E_\pi \left[ \tilde{P}_{x_2}(B) | X_1 \right] = \frac{\alpha_{x_2}}{a_2} (1 - I_{a_2}) + \delta_{X_1}(B) I_{x_2} \\
&= \frac{\alpha(B)}{a} (1 - I_a) (1 - I_{a_2}) + \delta_{X_2}(B) I_a (1 - I_{a_2}) + \delta_{X_1} I_{a_2}.
\end{aligned}$$

Entonces, para que esta desigualdad sea cierta es necesario que  $I_{a_1} = I_{a_2} = \frac{I_a}{1+I_a}$ .

Con lo que demostramos (2.29) para  $n = 2$ .

Supóngase que,  $I_{a_j} = \frac{I_a}{1+(j-1)I_a}$  para  $j = 1, \dots, n-2$  con  $n \geq 2$  y sea

$$I_{a_{(n-2),x_n}} := a^* \int_0^\infty u e^{-a^* \phi^*(u)} \left( \int_{\mathbb{R}^+} v^2 e^{uv} \nu^*(dv) \right) du,$$

donde  $a^* = \alpha^*(\mathbb{R})$  y  $\alpha^*$  y  $\nu^*$  son el parámetro y la medida de Lévy de un proceso Dirichlet, respectivamente. Así, por el teorema 2 y el teorema 6,

$$E_\pi \left[ \tilde{P}(B) | X_1, \dots, X_{n-2}, X_n, X_{n-1} \right] = \frac{\alpha^*(B)}{a^*} [1 - I_{a_{(n-2),x_n}}] + \delta_{x_{n-1}}(B) I_{a_{(n-2),x_n}}.$$

Ahora, si  $\tilde{P}_{x_{n-2}}$  es un proceso Dirichlet con parámetros  $(\alpha_{n-2}, \nu_{n-2})$ , se tiene

$$\begin{aligned}
\frac{\alpha^*(B)}{a^*} &= E_\pi \left[ \tilde{P}_{x_{n-2},x_n}(B) | X_1, \dots, X_{n-2}, X_n, X_{n-1} \right] = \frac{\alpha_{n-2}(B)}{a_{n-2}} [1 - I_{a_{n-2}}] + \delta_{X_n}(B) I_{a_{n-2}} \\
&= \left\{ \frac{\alpha(B)}{a} \prod_{j=1}^{n-2} [1 - I_{a_{n-2}}] + \sum_{k=1}^{n-2} \delta_{X_k}(B) I_{a_{k-1}} \prod_{j=k}^{n-3} [1 - I_{a_j}] \right\} \\
&\quad [1 - I_{a_{n-2}}] + \delta_{X_n}(B) I_{a_{n-2}} \\
&= \frac{\alpha(B)}{a} \frac{1 - I_a}{1 + (n-2)I_a} + \left( \sum_{k=1}^{n-2} \delta_{x_k}(B) + \delta_{X_n}(B) \right) \frac{I_a}{1 + (n-2)I_a}.
\end{aligned}$$

Entonces,

$$\begin{aligned}
E_\pi \left[ \tilde{P}(B) \mid X_1, \dots, X_{n-2}, X_n, X_{n-1} \right] &= \frac{\alpha(B)}{a} \frac{(1 - I_a)}{1 + (n - 2) I_a} [1 - I_{x_{n-2}, x_n}] \\
&+ \left( \sum_{k=1}^{n-2} \delta_{x_k}(B) + \delta_{X_n}(B) \right) \frac{I_a}{1 + (n - 2) I_a} [1 - I_{x_{n-2}, x_n}] + \delta_{X_{n-1}}(B) I_{a_{(n-2), x_n}} \\
&= E_\pi \left[ \tilde{P}(B) \mid X_1, \dots, X_{n-2}, X_{n-1}, X_n \right] \\
&= \frac{\alpha(B)}{a} \frac{(1 - I_a)}{1 + (n - 2) I_a} [1 - I_{a_{n-1}}] \\
&+ \left( \sum_{k=1}^{n-1} \delta_{X_k}(B) \right) \frac{I_a}{1 + I_a} [1 - I_{a_{n-1}}] + \delta_{X_n}(B) I_{a_{n-1}}.
\end{aligned}$$

Esto nos lleva a que  $I_{x_{n-2}, x_n} = I_{a_{n-1}}$  y por lo tanto

$$\frac{I_a [1 - I_{a_{n-1}}]}{1 + (n - 2) I_a} = I_{a_{n-1}}$$

$$\frac{I_a - I_a I_{a_{n-1}}}{1 + (n - 2) I_a} = I_{a_{n-1}}$$

$$I_a = I_{a_{n-1}} (1 + (n - 1) I_a)$$

$$I_{a_{n-1}} = \frac{I_a}{1 + (n - 1) I_a}.$$

Con esto queda demostrado totalmente el teorema.  $\square$

Con este resultado podemos ver una nueva caracterización del proceso Dirichlet y también podemos observar que no todos los procesos N-prior son conjugados, aunque la estructura de la media, de correlación, de la varianza y de la distribución predictiva de las distribuciones N-Prior es análoga a las estructuras del proceso Dirichlet.

Cabe destacar que la distribución final de las distribuciones N-Prior existe, pero no se presenta en esta tesis.

## 2.2. Ejemplos de procesos normalizados

Los teoremas presentados en la sección anterior serán aplicados a los tres procesos vistos en la sección 1.2, lo que servirá para mostrar la similitud de las estructuras entre los tres procesos.

### Proceso Dirichlet o Gamma normalizado

Considerando el exponente característico y la medida de Lévy definidos por (1.7) y (1.6) respectivamente, tenemos

#### *Esperanza del proceso Dirichlet*

Aplicando el teorema 2,

$$E_{\pi} [\tilde{P}(B)] = \frac{\alpha(B)}{a}.$$

#### *Varianza del proceso Dirichlet*

Aplicando el teorema 3,

$$\text{Var}_{\pi} [\tilde{P}(B)] = \frac{\alpha(B)(a - \alpha(B))}{a^2} I_a,$$

donde,

$$\begin{aligned} I_a &= a \int_0^{\infty} u e^{-a\psi(u)} \int_0^{\infty} v^2 e^{-uv} \nu(dv) du \\ &= a \int_0^{\infty} u e^{-a \log(\frac{\beta+u}{\beta})} \int_0^{\infty} v e^{-v(u+\beta)} dv du \\ &= a \int_0^{\infty} u \frac{(\beta+u)^{-a}}{\beta^{-a}} \frac{1}{(u+\beta)^2} du \\ &= a\beta^a \int_0^{\infty} u (u+\beta)^{-(a+2)} du = \frac{1}{1+a}. \end{aligned}$$

Entonces,

$$\text{Var}_\pi [\tilde{P}(B)] = \frac{\alpha(B)(a - \alpha(B))}{a^2(1+a)}.$$

### *Covarianza del proceso Dirichlet*

Aplicando el teorema 4,

$$\text{Cov}_\pi (\tilde{P}(B_1), \tilde{P}(B_2)) = \frac{a\alpha(C) - \alpha(B_1)\alpha(B_2)}{a^2} I_a = \frac{a\alpha(C) - \alpha(B_1)\alpha(B_2)}{a^2(1+a)}.$$

### *Distribución predictiva a un paso del proceso Dirichlet*

Aplicando el teorema 6,

$$\begin{aligned} \text{Pr}(X_2 \in B | X_1) &= \frac{\alpha(B)}{a} [1 - I_a] + \delta_{X_1}(B) I_a \\ &= \frac{\alpha(B)}{a} \left[ 1 - \frac{1}{a+1} \right] + \delta_{X_1}(B) \frac{1}{a+1} \\ &= \frac{\alpha(B)}{a} \frac{a}{1+a} + \delta_{X_1}(B) \frac{1}{a+1}. \end{aligned}$$

### *Distribución predictiva del proceso Dirichlet*

Aplicando el teorema 7,

$$\text{Pr}(X_{n+1} \in B | X_1, \dots, X_n) = w^{(n)} \frac{\alpha(B)}{a} + \frac{1}{n} \sum_{j=1}^k w_j^{(n)} \delta_{X_j^*}(B).$$

Primero obtengamos (2.3) para este proceso,

$$\begin{aligned}
\mu_n(u) &= \int_0^\infty v^n e^{-uv} \nu(dv) = \int_0^\infty v^{n-1} e^{-v(u+\beta)} dv \\
&= \frac{(n-1)}{(u+\beta)} \int_0^\infty v^{n-2} e^{-v(u+\beta)} dv \\
&= \frac{(n-1)(n-2)}{(u+\beta)^2} \int_0^\infty v^{n-3} e^{-v(u+\beta)} dv \\
&= \frac{(n-1)!}{(u+\beta)^n}.
\end{aligned}$$

Ahora,

$$\begin{aligned}
\prod_{k+1}^{(n+1)} (n_1, \dots, n_k, 1) &= \prod_{i=1}^k (n_i - 1)! \int_0^\infty u^n \left( \frac{u+\beta}{\beta} \right)^{-a} (u+\beta)^{-n-1} du \\
&= \prod_{i=1}^k (n_i - 1)! \beta^a \int_0^\infty u^n (u+\beta)^{-(a+n+1)} du \\
&= \prod_{i=1}^k (n_i - 1)! \frac{\Gamma(a) \Gamma(n+1)}{\Gamma(n+a+1)} \tag{2.30}
\end{aligned}$$

$$\begin{aligned}
\prod_k^{(n)} (n_1, \dots, n_k) &= \prod_{i=1}^k (n_i - 1)! \beta^a \int_0^\infty u^{n-1} (u+\beta)^{-(a+n)} du \\
&= \prod_{i=1}^k (n_i - 1)! \frac{\Gamma(a) \Gamma(n)}{\Gamma(n+a)}. \tag{2.31}
\end{aligned}$$

Con las ecuaciones (2.30) y (2.31) tenemos que (2.21) es

$$w^{(n)} = \frac{a}{n} \frac{\Gamma(n+1) \Gamma(n+a)}{\Gamma(n+a+1) \Gamma(n)} = \frac{a}{n+a}.$$

Para obtener  $w_j^{(n)}$ , observemos que



$$\begin{aligned}
\prod_k^{(n+1)} (n_1, \dots, n_j + 1, \dots, n_k) &= n_j \prod_{i=1}^k (n_i - 1)! \beta^a \int_0^\infty u^n (u + \beta)^{-(a+n+1)} \\
&= n_j \prod_{i=1}^k (n_i - 1)! \frac{\Gamma(a) \Gamma(n+1)}{\Gamma(n+a+1)}. \tag{2.32}
\end{aligned}$$

Con las ecuaciones (2.32) y (2.31) tenemos que (2.22)

$$w_j^{(n)} = \frac{n_j \Gamma(n+1) \Gamma(n+a)}{\Gamma(n+a+1) \Gamma(n)} = n_j \frac{n}{n+a}.$$

Entonces,

$$Pr(X_{n+1} \in B | X_1, \dots, X_n) = \frac{a}{n+a} \frac{\alpha(B)}{a} + \frac{1}{n+a} \sum_{j=1}^k n_j \delta_{X_j^*}(B).$$

## Proceso Estable normalizado

Considerando el exponente característico y la medida de Lévy definidos por (1.9) y (1.8), tenemos

### *Esperanza del proceso Estable normalizado*

Con el teorema 2 tenemos

$$E_\pi [\tilde{P}(B)] = \frac{\alpha(B)}{a}.$$

### *Varianza del proceso Estable normalizado*

Aplicando el teorema 3 resulta

$$Var_\pi [\tilde{P}(B)] = \frac{\alpha(B) (a - \alpha(B))}{a^2} I_a,$$

donde

$$\begin{aligned}
I_a &= a \int_0^\infty u e^{-a\psi(u)} \int_0^\infty v^2 e^{-uv} \nu(dv) du \\
&= a \int_0^\infty u e^{-a \frac{c\Gamma(1-\gamma)u^\gamma}{\gamma}} \int_0^\infty v^{1-\gamma} e^{-uv} dv du \\
&= a c \Gamma(2-\gamma) \int_0^\infty u^{\gamma-1} e^{-a \frac{c\Gamma(1-\gamma)u^\gamma}{\gamma}} du \\
&= 1 - \gamma.
\end{aligned}$$

Entonces,

$$Var_\pi [\tilde{P}(B)] = \frac{\alpha(B)(a - \alpha(B))}{a^2} (1 - \gamma).$$

### ***Covarianza del proceso Estable normalizado***

Con el teorema 4, tenemos

$$\begin{aligned}
Cov_\pi (\tilde{P}(B_1), \tilde{P}(B_2)) &= \frac{a \alpha(C) - \alpha(B_1) \alpha(B_2)}{a^2} I_a \\
&= \frac{a \alpha(C) - \alpha(B_1) \alpha(B_2)}{a^2} (1 - \gamma).
\end{aligned}$$

### ***Distribución predictiva a un paso del proceso Estable normalizado***

Con el teorema 6, tenemos

$$\begin{aligned}
Pr(X_2 \in B | X_1) &= \frac{\alpha(B)}{a} [1 - I_a] + \delta_{X_1}(B) I_a \\
&= \frac{\alpha(B)}{a} [1 - (1 - \gamma)] + \delta_{X_1}(B) (1 - \gamma) \\
&= \frac{\alpha(B)}{a} \gamma + \delta_{X_1}(B) (1 - \gamma).
\end{aligned}$$

*Distribución predictiva del proceso Estable normalizado*

Con el teorema 7, tenemos

$$Pr(X_{n+1} \in B | X_1, \dots, X_n) = w^{(n)} \frac{\alpha(B)}{a} + \frac{1}{n} \sum_{j=1}^k w_j^{(n)} \delta_{X_j^*}(B).$$

La ecuación (2.3) para este proceso es

$$\begin{aligned} \mu_n(u) &= \int_0^\infty v^n e^{-uv} \nu(dv) \\ &= c \int_0^\infty v^{n-1-\gamma} e^{-uv} dv \\ &= cu^{-n+\gamma} \Gamma(n-\gamma). \end{aligned}$$

Para obtener  $w^{(n)}$ , observemos que

$$\prod_{k+1}^{(n+1)} (n_1, \dots, n_k, 1) = c^{k+1} \Gamma(1-\gamma) \prod_{i=1}^k \Gamma(n_i - \gamma) \int_0^\infty u^{(k+1)\gamma-1} e^{-a \frac{c\Gamma(1-\gamma)u^\gamma}{\gamma}} du \quad (2.33)$$

$$\prod_k^{(n)} (n_1, \dots, n_k) = c^k \prod_{i=1}^k \Gamma(n_i - \gamma) \int_0^\infty u^{k\gamma-1} e^{-a \frac{c\Gamma(1-\gamma)u^\gamma}{\gamma}} du. \quad (2.34)$$

Con las ecuaciones (2.33) y (2.34) tenemos que

$$\begin{aligned} w^{(n)} &= \frac{a}{n} \frac{c^{k+1} \Gamma(1-\gamma) \prod_{i=1}^k \Gamma(n_i - \gamma) \int_0^\infty u^{(k+1)\gamma-1} e^{-a \frac{c\Gamma(1-\gamma)u^\gamma}{\gamma}} du}{c^k \prod_{i=1}^k \Gamma(n_i - \gamma) \int_0^\infty u^{k\gamma-1} e^{-a \frac{c\Gamma(1-\gamma)u^\gamma}{\gamma}} du} \\ &= \frac{k}{n} \gamma. \end{aligned}$$

Ahora, para el cálculo de  $w_j^{(n)}$ , encontramos que

$$\prod_k^{(n+1)} (n_1, \dots, n_j + 1, \dots, n_k) = c^k (n_j - 1) \prod_{i=1}^k \Gamma(n_i - \gamma) \int_0^\infty u^{k\gamma-1} e^{-a \frac{c\Gamma(1-\gamma)u^\gamma}{\gamma}} du. \quad (2.35)$$

Con las ecuaciones (2.35) y (2.34) tenemos que

$$\begin{aligned} w_j^{(n)} &= \frac{c^k (n_j - 1) \prod_{i=1}^k \Gamma(n_i - \gamma) \int_0^\infty u^{k\gamma-1} e^{-a \frac{c\Gamma(1-\gamma)u^\gamma}{\gamma}} du}{c^k \prod_{i=1}^k \Gamma(n_i - \gamma) \int_0^\infty u^{k\gamma-1} e^{-a \frac{c\Gamma(1-\gamma)u^\gamma}{\gamma}} du} \\ &= (n_j - \gamma). \end{aligned}$$

Entonces,

$$Pr(X_{n+1} \in B | X_1, \dots, X_n) = \gamma \frac{k}{n} \frac{\alpha(B)}{a} + \frac{1}{n} \sum_{j=1}^k (n_j - \gamma) \delta_{X_j^*}(B).$$

Nótese que para este proceso el término correspondiente a la distribución empírica se ve afectada por  $\gamma$ .

## Proceso Gaussiano Inverso normalizado

Considerando el exponente característico y la medida de Lévy definidos por (1.11) y (1.10), tenemos

### *Esperanza del proceso Gaussiano Inverso normalizado*

Aplicando el teorema 2,

$$E_\pi \left[ \tilde{P}(B) \right] = \frac{\alpha(B)}{a}.$$

*Varianza del proceso Gaussiano Inverso normalizado*

Aplicando el teorema 3,

$$\text{Var}_\pi [\tilde{P}(B)] = \frac{\alpha(B)(a - \alpha(B))}{a^2} I_a,$$

con

$$\begin{aligned} I_a &= a \int_0^\infty u e^{-a\psi(u)} \int_0^\infty v^2 e^{-uv} \nu(dv) du \\ &= a \int_0^\infty u e^{-a\delta(\sqrt{2u+\gamma^2}-\gamma)} \int_0^\infty \frac{\delta}{\sqrt{2\pi}} v^{1/2} e^{-v\left(u+\frac{\gamma^2}{2}\right)} dv du \\ &= a \int_0^\infty u e^{-a\delta(\sqrt{2u+\gamma^2}-\gamma)} \delta(2u+\gamma^2)^{-3/2} du \\ &= a \delta e^{a\delta\gamma} \int_0^\infty u (2u+\gamma^2)^{-3/2} e^{-a\delta\sqrt{2u+\gamma^2}} du. \end{aligned}$$

Sea  $y = \sqrt{2u+\gamma^2}$ , entonces,

$$\begin{aligned} I_a &= \frac{a\delta}{2} e^{a\delta\gamma} \int_\gamma^\infty y^{-2} (y^2 - \gamma^2) e^{-a\delta y} dy \\ &= \frac{a\delta}{2} e^{a\delta\gamma} \left[ \frac{e^{-a\delta y} (1 - a\delta y)}{a\delta} + a\delta\gamma^2 \Gamma(0, a\delta y) \right] \\ &= \frac{1}{2} - \frac{a\delta\gamma}{2} + \frac{(a\delta\gamma)^2}{2} e^{a\delta\gamma} \Gamma(0, a\delta\gamma), \end{aligned}$$

donde  $\Gamma(a, x) = \int_x^\infty t^{a-1} e^{-t} dt$ , llamada la función Gamma Incompleta, la cual cumple con

$$\Gamma(a+1, x) = a \Gamma(a, x) + x^a e^{-x}.$$

Por lo tanto

$$\begin{aligned} \Gamma(0, a\delta\gamma) &= -\Gamma(-1, a\delta\gamma) + (a\delta\gamma)^{-1} e^{-a\delta\gamma} \text{ y} \\ \frac{(a\delta\gamma)^2}{2} e^{a\delta\gamma} \Gamma(0, a\delta\gamma) &= -\frac{(a\delta\gamma)^2}{2} e^{a\delta\gamma} \Gamma(-1, a\delta\gamma) + \frac{a\delta\gamma}{2}, \end{aligned}$$

lo cual nos lleva a que

$$I_a = \frac{1}{2} - \frac{(a\delta\gamma)^2}{2} e^{a\delta\gamma} \Gamma(-1, a\delta\gamma).$$

Entonces llegamos a que la varianza es

$$\text{Var}_\pi [\tilde{P}(B)] = \frac{\alpha(B)(a - \alpha(B))}{a^2} \left( \frac{1}{2} - \frac{(a\delta\gamma)^2}{2} e^{a\delta\gamma} \Gamma(-1, a\delta\gamma) \right).$$

### ***Covarianza del proceso Gaussiano Inverso normalizado***

Aplicando el teorema 4,

$$\begin{aligned} \text{Cov}_\pi \left( \tilde{P}(B_1), \tilde{P}(B_2) \right) &= \frac{a\alpha(C) - \alpha(B_1)\alpha(B_2)}{a^2} I_a \\ &= \frac{a\alpha(C) - \alpha(B_1)\alpha(B_2)}{a^2} \left( \frac{1}{2} - \frac{(a\delta\gamma)^2}{2} e^{a\delta\gamma} \Gamma(-1, a\delta\gamma) \right). \end{aligned}$$

### ***Distribución predictiva a un paso del proceso Gaussiano Inverso normalizado***

Aplicando el teorema 6,

$$\begin{aligned} \text{Pr}(X_2 \in B | X_1) &= \frac{\alpha(B)}{a} [1 - I_a] + \delta_{X_1}(B) I_a \\ &= \frac{\alpha(B)}{a} \left[ 1 - \frac{1}{2} + \frac{(a\delta\gamma)^2}{2} e^{a\delta\gamma} \Gamma(-1, a\delta\gamma) \right] \\ &\quad + \delta_{X_1}(B) \left( \frac{1}{2} - \frac{(a\delta\gamma)^2}{2} e^{a\delta\gamma} \Gamma(-1, a\delta\gamma) \right) \\ &= \frac{\alpha(B)}{a} \left[ \frac{1}{2} + \frac{(a\delta\gamma)^2}{2} e^{a\delta\gamma} \Gamma(-1, a\delta\gamma) \right] \\ &\quad + \delta_{X_1}(B) \left( \frac{1}{2} - \frac{(a\delta\gamma)^2}{2} e^{a\delta\gamma} \Gamma(-1, a\delta\gamma) \right). \end{aligned}$$

*Distribución predictiva del proceso Gaussiano Inverso normalizado*

Aplicando el teorema 7,

$$Pr(X_{n+1} \in B | X_1, \dots, X_n) = w^{(n)} \frac{\alpha(B)}{a} + \frac{1}{n} \sum_{j=1}^k w_j^{(n)} \delta_{X_j^*}(B).$$

Para este proceso. (2.3) es

$$\begin{aligned} \mu_n(u) &= \int_0^\infty v^n e^{-uv} \nu(dv) \\ &= \frac{\delta}{\sqrt{2\pi}} \int_0^\infty v^{n-\frac{3}{2}} e^{-v\left(u+\frac{\gamma^2}{2}\right)} dv \\ &= \frac{\delta}{\sqrt{2\pi}} \left(u + \frac{\gamma^2}{2}\right)^{\frac{1}{2}-n} \Gamma\left(n - \frac{1}{2}\right). \end{aligned}$$

Para el cálculo de  $w^{(n)}$ , usamos el resultado anterior y llegamos a que

$$\begin{aligned} &\prod_{k+1}^{(n+1)} (n_1, \dots, n_k, 1) \\ &= \left(\frac{\delta}{\sqrt{\pi}}\right)^{k+1} \frac{\left(\prod_{i=1}^k \Gamma\left(n_i - \frac{1}{2}\right)\right) \Gamma\left(\frac{1}{2}\right)}{2^k} e^{a\delta\gamma} \int_0^\infty u^n (2u + \gamma^2)^{\frac{k+1}{2}-n-1} e^{-a\delta\sqrt{2u+\gamma^2}} du. \end{aligned}$$

Sea  $y = \sqrt{2u + \gamma^2}$ , entonces nos queda

$$\begin{aligned} &\prod_{k+1}^{(n+1)} (n_1, \dots, n_k, 1) \\ &= \left(\frac{\delta}{\sqrt{\pi}}\right)^{k+1} \frac{\left(\prod_{i=1}^k \Gamma\left(n_i - \frac{1}{2}\right)\right) \Gamma\left(\frac{1}{2}\right)}{2^k} e^{a\delta\gamma} \int_\gamma^\infty (y^2 - \gamma^2)^n y^{k-2n} e^{-a\delta y} dy. \end{aligned}$$

Aplicando el teorema del binomio a  $(y^2 - \gamma^2)^n$ ,

$$\begin{aligned}
&= \left( \frac{\delta}{\sqrt{\pi}} \right)^{k+1} \frac{\left( \prod_{i=1}^k \Gamma \left( n_i - \frac{1}{2} \right) \right) \Gamma \left( \frac{1}{2} \right)}{2^k} e^{a\delta\gamma} \int_{\gamma}^{\infty} \sum_{i=0}^n \binom{n}{i} (-\gamma^2)^{n-i} y^{2i+k-2n} e^{-a\delta y} dy \\
&= \left( \frac{\delta}{\sqrt{\pi}} \right)^{k+1} \frac{\left( \prod_{i=1}^k \Gamma \left( n_i - \frac{1}{2} \right) \right) \Gamma \left( \frac{1}{2} \right)}{2^k} e^{a\delta\gamma} \sum_{i=0}^n \binom{n}{i} (-\gamma^2)^{n-i} \int_{\gamma}^{\infty} y^{2i+k-2n} e^{-a\delta y} dy \\
&= \left( \frac{\delta}{\sqrt{\pi}} \right)^{k+1} \frac{\left( \prod_{i=1}^k \Gamma \left( n_i - \frac{1}{2} \right) \right) \Gamma \left( \frac{1}{2} \right)}{2^k} e^{a\delta\gamma} \\
&\quad \sum_{i=0}^n \binom{n}{i} (-\gamma^2)^{n-i} (a\delta)^{2n-2i-k-1} \Gamma(1+2i+k-2n, a\delta\gamma). \tag{2.36}
\end{aligned}$$

Análogamente,

$$\begin{aligned}
&\prod_k^{(n)} (n_1, \dots, n_k) \\
&= \left( \frac{\delta}{\sqrt{\pi}} \right)^k \frac{\left( \prod_{i=1}^k \Gamma \left( n_i - \frac{1}{2} \right) \right)}{2^{k-1}} e^{a\delta\gamma} \int_0^{\infty} u^{n-1} (2u + \gamma^2)^{\frac{k}{2}-n} e^{-a\delta\sqrt{2u+\gamma^2}} du.
\end{aligned}$$

Haciendo de nuevo  $y = \sqrt{2u + \gamma^2}$  entonces

$$\begin{aligned}
&\prod_k^{(n)} (n_1, \dots, n_k) \\
&= \left( \frac{\delta}{\sqrt{\pi}} \right)^k \frac{\left( \prod_{i=1}^k \Gamma \left( n_i - \frac{1}{2} \right) \right)}{2^{k-1}} e^{a\delta\gamma} \int_{\gamma}^{\infty} (y^2 - \gamma^2)^{n-1} y^{k-2n+1} e^{-a\delta y} dy
\end{aligned}$$



y aplicando el teorema binomial, tenemos

$$\begin{aligned}
& \prod_k^{(n)} (n_1, \dots, n_k) \\
&= \left( \frac{\delta}{\sqrt{\pi}} \right)^k \frac{\left( \prod_{i=1}^k \Gamma \left( n_i - \frac{1}{2} \right) \right)}{2^{k-1}} e^{a\delta\gamma} \int_{\gamma}^{\infty} \sum_{i=0}^{n-1} \binom{n-1}{i} (-\gamma^2)^{n-1-i} y^{2i+k-2n+1} e^{-a\delta y} dy \\
&= \left( \frac{\delta}{\sqrt{\pi}} \right)^k \frac{\left( \prod_{i=1}^k \Gamma \left( n_i - \frac{1}{2} \right) \right)}{2^{k-1}} e^{a\delta\gamma} \sum_{i=0}^{n-1} \binom{n-1}{i} (-\gamma^2)^{n-i-1} \int_{\gamma}^{\infty} y^{2i+k-2n+1} e^{-a\delta y} dy \\
&= \left( \frac{\delta}{\sqrt{\pi}} \right)^k \frac{\left( \prod_{i=1}^k \Gamma \left( n_i - \frac{1}{2} \right) \right)}{2^{k-1}} e^{a\delta\gamma} \\
& \quad \sum_{i=0}^{n-1} \binom{n-1}{i} (-\gamma^2)^{n-i-1} (a\delta)^{2n-2i-k-2} \Gamma(2+2i+k-2n, a\delta\gamma). \tag{2.37}
\end{aligned}$$

Con las ecuaciones (2.36) y (2.37) tenemos que

$$w^{(n)} = \frac{(a\delta)^2 (-\gamma^2) \sum_{i=0}^n \binom{n}{i} (-\gamma^2)^{-i} (a\delta)^{-2i} \Gamma(1+2i+k-2n, a\delta\gamma)}{2n \sum_{i=0}^{n-1} \binom{n-1}{i} (-\gamma^2)^{-i} (a\delta)^{-2i} \Gamma(2+2i+k-2n, a\delta\gamma)}.$$

Para obtener  $w_j^{(n)}$ , tenemos que

$$\begin{aligned}
& \prod_k^{(n+1)} (n_1, \dots, n_j + 1, \dots, n_k) \\
&= \left( \frac{\delta}{\sqrt{\pi}} \right)^k \frac{\left( \prod_{i=1}^k \Gamma \left( n_i - \frac{1}{2} \right) \right) (n_j - \frac{1}{2})}{2^{k-1}} e^{a\delta\gamma} \int_0^{\infty} u^n (2u + \gamma^2)^{\frac{k}{2}-n-1} e^{-a\delta\sqrt{2u+\gamma^2}} du.
\end{aligned}$$

Igual que en las dos ecuaciones anteriores, sea  $y = \sqrt{y^2 - \gamma^2}$  y aplicando el teorema

binomial,

$$\begin{aligned}
& \prod_k^{(n+1)} (n_1, \dots, n_j + 1, \dots, n_k) \\
&= \left( \frac{\delta}{\sqrt{\pi}} \right)^k \frac{\left( \prod_{i=1}^k \Gamma \left( n_i - \frac{1}{2} \right) \right) \left( n_j - \frac{1}{2} \right)}{2^{k-1}} e^{a\delta\gamma} \int_{\gamma}^{\infty} (y^2 - \gamma^2)^n y^{k-2n-1} e^{-a\delta y} dy \\
&= \left( \frac{\delta}{\sqrt{\pi}} \right)^k \frac{\left( \prod_{i=1}^k \Gamma \left( n_i - \frac{1}{2} \right) \right) \left( n_j - \frac{1}{2} \right)}{2^{k-1}} e^{a\delta\gamma} \int_{\gamma}^{\infty} \sum_{i=0}^n \binom{n}{i} (-\gamma^2)^{n-i} y^{2i+k-2n-1} e^{-a\delta y} dy \\
&= \left( \frac{\delta}{\sqrt{\pi}} \right)^k \frac{\left( \prod_{i=1}^k \Gamma \left( n_i - \frac{1}{2} \right) \right) \left( n_j - \frac{1}{2} \right)}{2^{k-1}} e^{a\delta\gamma} \sum_{i=0}^n \binom{n}{i} (-\gamma^2)^{n-i} \int_{\gamma}^{\infty} y^{2i+k-2n-1} e^{-a\delta y} dy \\
&= \left( \frac{\delta}{\sqrt{\pi}} \right)^k \frac{\left( \prod_{i=1}^k \Gamma \left( n_i - \frac{1}{2} \right) \right) \left( n_j - \frac{1}{2} \right)}{2^{k-1}} e^{a\delta\gamma} \\
& \quad \sum_{i=0}^n \binom{n}{i} (-\gamma^2)^{n-i} (a\delta)^{2n-2i-k} \Gamma(2i+k-2n, a\delta\gamma). \tag{2.38}
\end{aligned}$$

Con las ecuaciones (2.38) y (2.37) tenemos que

$$w_j^{(n)} = (a\delta)^2 (-\gamma^2) \left( n_j - \frac{1}{2} \right) \frac{\sum_{i=0}^n \binom{n}{i} (-\gamma^2)^{-i} (a\delta)^{-2i} \Gamma(2i+k-2n, a\delta\gamma)}{\sum_{i=0}^{n-1} \binom{n}{i} (-\gamma^2)^{-i} (a\delta)^{-2i} \Gamma(2+2i+k-2n, a\delta\gamma)}.$$

Entonces,

$$\begin{aligned}
& Pr(X_{n+1} \in B | X_1, \dots, X_n) \\
&= \frac{(a\delta)^2 (-\gamma^2)^{\sum_{i=0}^n \binom{n}{i} (-\gamma^2)^{-i} (a\delta)^{-2i} \Gamma(1+2i+k-2n, a\delta\gamma)}}{2n \sum_{i=0}^{n-1} \binom{n-1}{i} (-\gamma^2)^{-i} (a\delta)^{-2i} \Gamma(2+2i+k-2n, a\delta\gamma)} \frac{\alpha(B)}{a} \\
&+ \frac{1}{n} \sum_{j=1}^k (a\delta)^2 (-\gamma^2)^{\left(n_j - \frac{1}{2}\right)} \\
&\frac{\sum_{i=0}^n \binom{n}{i} (-\gamma^2)^{-i} (a\delta)^{-2i} \Gamma(2i+k-2n, a\delta\gamma)}{\sum_{i=0}^{n-1} \binom{n-1}{i} (-\gamma^2)^{-i} (a\delta)^{-2i} \Gamma(2+2i+k-2n, a\delta\gamma)} \delta_{X_j^*}(B).
\end{aligned}$$

Observemos que la diferencia principal de los pesos entre este proceso y el proceso Estable normalizado es  $(n_j - \frac{1}{2})$  y  $(n_j - \gamma)$ .

En estos tres ejemplos se puede ver claramente que las estructuras de la media, varianza, y covarianza, así como de la distribución predictiva son semejantes y que todo depende de los parámetros de cada proceso.

### 2.3. Esquema de la Urna de Pólya

En esta sección se expone un algoritmo para simular los procesos normalizados a partir del esquema de la Urna de Pólya. El esquema de la Urna de Pólya relacionado con el proceso Dirichlet fue presentado por Blackwell y MacQueen [1973] que sustenta un algoritmo de simulación de realizaciones del proceso Dirichlet. Basándose en este esquema, Escobar y West [1995] dan una aplicación con mezclas de procesos Dirichlet. Ishwaran y James [2001] mencionan que se puede aplicar la Urna de Pólya a cualquier proceso con distribución predictiva.

**Definición 4** (Blackwell y MacQueen [1973]). Una sucesión  $\{X_n : n \geq 1\}$  de variables aleatorias con valores en  $(\mathbb{X}, \mathcal{X})$ , es una **sucesión de Pólya con parámetro  $\alpha$**  si, para todo  $B \subset \mathcal{X}$ ,

$$P(X_1 \in B) = \frac{\alpha(B)}{\alpha(\mathcal{X})}$$

y

$$P(X_{n+1} \in B | X_1, \dots, X_n) = \frac{\alpha_n(B)}{\alpha_n(\mathcal{X})},$$

donde  $\alpha_n(\cdot) = \alpha(\cdot) + \sum_{i=1}^n \delta_{X_i}(\cdot)$ .

Observemos que, para un espacio  $\mathcal{X}$  finito, la sucesión  $\{X_n\}$  se puede ver como los resultados de extracciones sucesivas de una urna que inicialmente tiene  $\alpha(x)$  bolas de color  $x$  y donde después de cada extracción, la bola es remplazada y se agrega otra bola del mismo color a la urna.

Para las distribuciones N-prior tenemos por el teorema 7 que la distribución predictiva es

$$Pr(X_{n+1} \in B | X_1, \dots, X_n) = w^{(n)} G_0(B) + \frac{1}{n} \sum_{j=1}^k w_j^{(n)} \delta_{X_j^*}(B),$$

la cual se puede caracterizar por el esquema de la Urna de Pólya.

Considerando el teorema de Blackwell y MacQueen [1973] se puede obtener una muestra de una función de distribución. En este caso sería para las distribuciones N-Prior.

Sea  $Z_1, \dots, Z_n$  una muestra de observaciones independientes e idénticamente distribuidas de  $G_0$ .

Sea

$$X_1 = Z_1,$$

y para  $i = 2, 3, \dots, n$

$$(X_i|X_1, \dots, X_{i-1}) = \begin{cases} Z_i & \text{con probabilidad } w^{(i-1)} \\ X_{i-1,j}^* & \text{con probabilidad } \frac{w_j^{(i-1)}}{i-1}, j = 1, \dots, k_i, \end{cases} \quad (2.39)$$

donde  $k_i$  representa el número de observaciones distintas en una muestra de tamaño  $i - 1$  denotadas por  $X_{i-1,1}^*, \dots, X_{i-1,k_i}^*$ , y  $w^{(i-1)}$  y  $w_j^{(i-1)}$  están dados por la ecuaciones (2.21) y (2.22) respectivamente.

En el artículo de Ishwaran y James [2001] los autores generalizan este método caracterizando la distribución conjunta de  $X_1, X_2, \dots, X_n$ . Este mismo algoritmo se puede utilizar para los procesos normalizados.

Sea una serie observada,  $X_1, \dots, X_n$ , y  $Z_1, \dots, Z_n$  una muestra de observaciones independientes e idénticamente distribuidas de  $G_0$ . Sea

$$X_1 = Z_1,$$

y para  $i = 2, 3, \dots, n$

$$(X_i|X_{-i}) = \begin{cases} Z_i & \text{con probabilidad } w^{(n-1)} \\ X_{i,j}^* & \text{con probabilidad } \frac{w_j^{(n-1)}}{n-1}, j = 1, \dots, k_i, \end{cases} \quad (2.40)$$

donde  $X_{-i}$  denota la muestra de tamaño  $n$  sin la observación  $i$ .

La Urna de Pólya para los procesos N-prior tiene la tendencia de obtener valores nuevos de  $X^*$  muy lentamente si los valores  $w_j^{(n)}$  son mucho más grandes que el valor de  $w^{(n)}$ . Es decir, se tienen que realizar muchas iteraciones antes de obtener un nuevo valor de  $X^*$ , y mientras tanto se irán obteniendo muestras de los valores previos  $X_1^*, \dots, X_k^*$ .

Esto conlleva a que las realizaciones del proceso se estancuen en ciertos valores, aunque para algunos procesos N-prior este estancamiento es menor, por ejemplo el proceso Estable.

Hay dos formas de abordar el problema del estancamiento de las realizaciones del proceso; una es con las mezclas de procesos normalizados que se definirán en la sección 2.4 y la otra es haciendo un remuestreo de  $X_1^*, \dots, X_k^*$  como lo proponen Ishwaran y James [2001].

Sea  $C = (C_1, \dots, C_n)$ , donde cada  $C_i$  contiene los índices de las  $X$ 's tales que  $X_j^*$  corresponde al valor  $X_i$ , es decir,  $X_i = X_j^*$  si y sólo si  $C_i = j$ . Entonces hay que muestrear de la distribución condicional para  $(X_j^*|C, \theta)$  para  $j = 1, \dots, k$ .

Como veremos más adelante, este trabajo tiene como finalidad la comparación de algunos modelos dentro de la familia N-prior. Por esta razón, el paso de remuestreo no se llevará acabo ya que afectaría dicha comparación.

## 2.4. Mezclas de procesos normalizados

Los procesos normalizados no consideran todas las posibles situaciones que se tienen en el análisis Bayesiano, por ejemplo los problemas con observaciones continuas. Esta es una de las principales motivaciones del estudio de mezcla de procesos normalizados.

Antoniak [1974] definió la mezcla de procesos Dirichlet, aleatorizando el parámetro  $\alpha$ .

$\tilde{P}$  es una Mezcla de Procesos Dirichlet, con distribución de mezcla  $H$ , si para todo  $k = 1, 2, \dots$  y cualquier partición  $B_1, \dots, B_k$  del espacio de parámetros se tiene que

$$Pr\left(\tilde{P}(B_1) \leq x_1, \dots, \tilde{P}(B_k) \leq x_k\right) = \int_U D(x_1, \dots, x_k | \alpha(u, B_1), \dots, \alpha(u, B_k)) dH(u),$$

donde  $D(x_1, \dots, x_k | \alpha_1, \dots, \alpha_k)$  denota la función de distribución de la distribución Dirichlet con parámetros  $(\alpha_1, \dots, \alpha_k)$ . Esto se puede denotar como

$$u \sim H \quad P|u \sim DP(\alpha_u) \quad \text{con} \quad P = \int_U DP(\alpha_u) dH(u).$$

Por otro lado, Lo [1984] propone el modelo donde la distribución inicial de los parámetros de un kernel no negativo sigue un proceso Dirichlet.

Sea  $K(x; u)$  un kernel no negativo tal que, para cada  $u$ ,  $\int K(x; u) dx = 1$  y, para cada  $x$ ,  $\int K(x; u) \alpha(du) < \infty$ . Se define la función de densidad  $f(x|G) = \int K(x; u) G(du)$  donde  $G$  sigue un proceso Dirichlet.

Escobar y West [1995] unieron ambos trabajos usando la densidad normal como kernel y asignando a la distribución de los parámetros de la normal un proceso Dirichlet.

Esta última idea se puede aplicar a los procesos normalizados, dando como resultado la siguiente definición.

**Definición 5** Sean  $Y_1, \dots, Y_n$  condicionalmente independientes y tales que  $(Y_j|X_i, \theta) \sim F_i(X_i, \theta)$ , donde  $(X_i, \theta)$  son los parámetros de la distribución  $F_i$ , con  $i = 1, \dots, n$ . Si la distribución de  $\theta$  es  $F(\theta)$  y la distribución de  $X_i$  es incierta y modelada como un proceso normalizado entonces se dice que los datos vienen de una **mezcla de procesos normalizados**

Denotado esto último, como un modelo jerárquico se tiene

$$\begin{aligned} Y_i|X_i, \theta &\stackrel{\text{iid}}{\sim} F_i(Y_i|X_i, \theta) \quad i = 1, \dots, n \\ X_i &\stackrel{\text{iid}}{\sim} \tilde{P} \\ \tilde{P} &\sim \text{N-prior} \\ \theta &\sim F(\theta) \end{aligned}$$

Ishwaran y James [2001] aplican el algoritmo del muestreo de Gibbs para las mezclas de procesos Dirichlet y para el proceso Pitman-Yor. Dicho algoritmo se puede generalizar al

caso de los procesos normalizados.

Sea  $X_{-i}$  el subvector de  $X$  al quitar la coordenada  $i$ -ésima. Entonces, para la simulación de la distribución final  $f(X, \theta|Y)$  del modelo anterior, se deben simular valores iterativamente de la distribución condicional de  $(X_i|X_{-i}, \theta, Y)$   $i = 1, \dots, n$  y de  $(\theta|X, Y)$ .

Para simular de la primera distribución condicional se tiene que

$$P(X_i \in \cdot | X_{-i}, \theta, Y) = q_0^* P(X_i \in \cdot | \theta, Y) + \sum_{j=1}^n q_j^* \delta_{X_j^*}(\cdot),$$

donde

$$q_0^* \propto w^{(n)} \int f(Y_i | X, \theta) G_0(dX)$$

y

$$q_j^* \propto w_j^{(n)} f(Y_i | X_j^*, \theta).$$

Estos valores están sujetos a que  $\sum_{j=0}^m q_j^* = 1$ .

En las aplicaciones del capítulo 3 se utilizará este algoritmo.

## 2.5. Distribución de $K_n$

Ahora, observemos que la tasa a la cual aparecerán nuevos valores diferentes en los procesos normalizados presentados en este trabajo sólo depende de  $a$  o de  $\gamma$  y no de la forma de  $\alpha$ . Para los procesos normalizados presentados aquí tendríamos lo siguiente.

### Proceso Dirichlet



$$Pr(X_{n+1} \notin (X_1^*, \dots, X_k^*) | X_1, \dots, X_n) = \frac{a}{n+a}.$$

### Proceso Estable normalizado

$$Pr(X_{n+1} \notin (X_1^*, \dots, X_k^*) | X_1, \dots, X_n) = \frac{k}{n} \gamma.$$

### Proceso Gaussiano Inverso normalizado

$$Pr(X_{n+1} \notin (X_1^*, \dots, X_k^*) | X_1, \dots, X_n) = \frac{1}{2n} (a\delta)^2 (-\gamma^2) \frac{\sum_{i=0}^n \binom{n}{i} (-\gamma^2)^{-i} (a\delta)^{-2i} \Gamma(1+2i+k-2n, a\delta\gamma)}{\sum_{i=0}^{n-1} \binom{n-1}{i} (-\gamma^2)^{-i} (a\delta)^{-2i} \Gamma(2+2i+k-2n, a\delta\gamma)}.$$

Observemos que, en el caso del proceso Dirichlet,  $Pr(X_{n+1} \notin (X_1^*, \dots, X_k^*) | X_1, \dots, X_n)$  no depende de  $k$ , sólo de la magnitud de  $a$  y del tamaño de la muestra  $n$ . Para los otros dos procesos dicha probabilidad depende de los tres parámetros;  $a, k$  y  $n$ .

Además, observemos que  $k$  es importante en estas distribuciones porque con base en este parámetro se puede obtener la probabilidad de obtener un nuevo valor; si  $a$  y  $\gamma$  son fijos, entre más grande sea  $k$ , más cercana a uno será la probabilidad de obtener un nuevo valor fuera de  $(X_1^*, \dots, X_k^*)$ .

Lo interesante aquí es saber cuál es la distribución (tanto de forma inicial como de forma final) de los valores distintos que tenemos en una muestra de tamaño  $n$ .

Sea  $K_n$  la variable aleatoria que denota el número de valores distintos en una muestra discreta de tamaño  $n$ . Entonces al elegir una distribución N-prior se induce una distribución sobre  $K_n$ , la cual no es aleatoria y no siempre se simplifica de una forma razonable.

Aquí presentaremos dicha distribución para los casos de distribuciones N-prior aquí analizadas.

## Proceso Dirichlet o Gamma normalizado

Antoniak [1974] encuentra la distribución de  $K_n$ , llegando a

$$Pr(K_n = k|n) = c_n(k) a^k \frac{\Gamma(a)}{\Gamma(n+a)}, \quad (2.41)$$

donde  $c_n(k)$  es el valor absoluto de un número de Stirling de primera clase.

Esta fórmula también se conoce como la fórmula de muestreo de Ewens [1972] y se utiliza mucho en biología matemática.

La figura 2.1 muestra la distribución inicial de  $K_n$  para algunos valores de  $a$  y  $n = 60$  y se puede observar que conforme  $a$  va creciendo, la densidad de  $K_n$  va teniendo mayor dispersión hasta llegar a un valor de  $a$  tal que la distribución vuelve a tener menor varianza.

## Proceso Estable normalizado

En el artículo de Lijoi, Mena, y Prünster [2006b] se obtiene la distribución de  $K_n$ , la cual no depende de  $a$ :

$$Pr(K_n = k|n) = \frac{1}{\gamma \Gamma(n)} \sum_{j=0}^k (-1)^j \binom{k}{j} (-j\gamma)_n, \quad (2.42)$$

donde  $(b)_n = b(b+1)\cdots(b+n-1)$  denota el factorial creciente. Esta expresión conocida como la fórmula de Pashammer.

La figura 2.2 muestra algunos ejemplos de la distribución inicial de  $K_n$ .

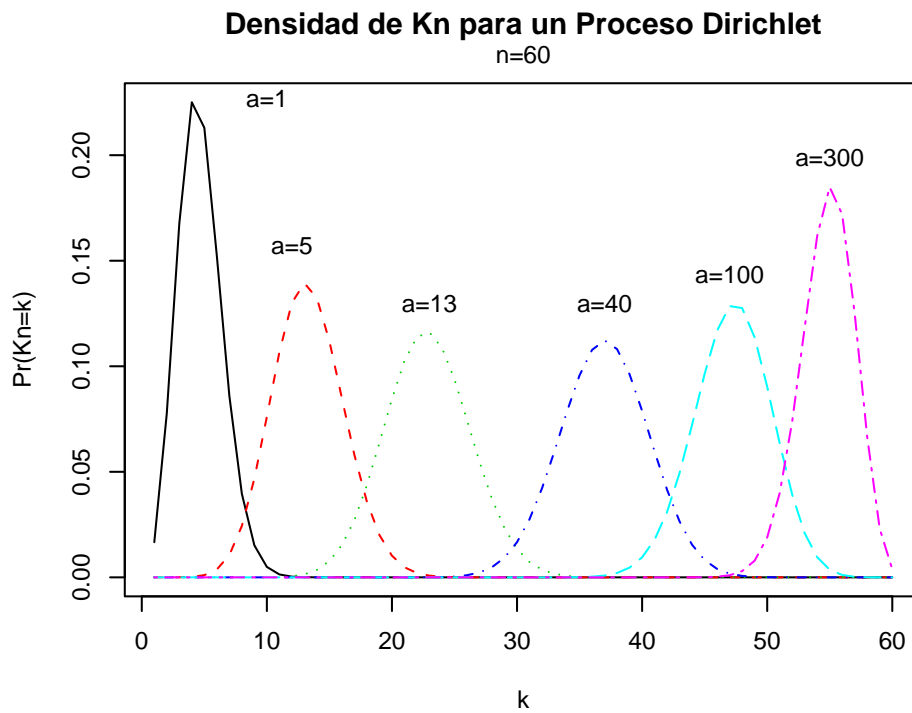


Figura 2.1: Ejemplos de  $\Pr(K_n = k)$  considerando un proceso Dirichlet.

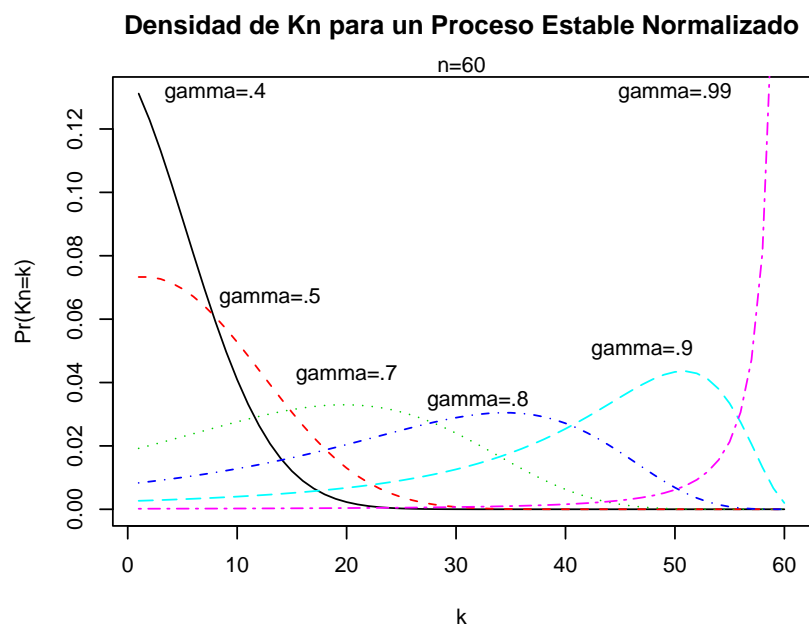


Figura 2.2: Ejemplos de  $\Pr(K_n = k)$  considerando un proceso Estable normalizado.

En esta figura se tomó un tamaño de muestra igual a 60, y con esto se observa que con valores de  $\gamma$  menores de 0.4 y mayores a 0.9, la moda es 0 y 60 respectivamente. Sobre el

resto de las distribuciones se puede decir que son planas en comparación con las mostradas en el caso del proceso Dirichlet.

Además, observemos que el valor máximo de la densidad es menor a 0.12 para las distribuciones donde  $\gamma \neq 0.4$  y  $\gamma \neq 0.99$ , mientras que con el proceso Dirichlet es mayor a 0.20 (figura 2.3).

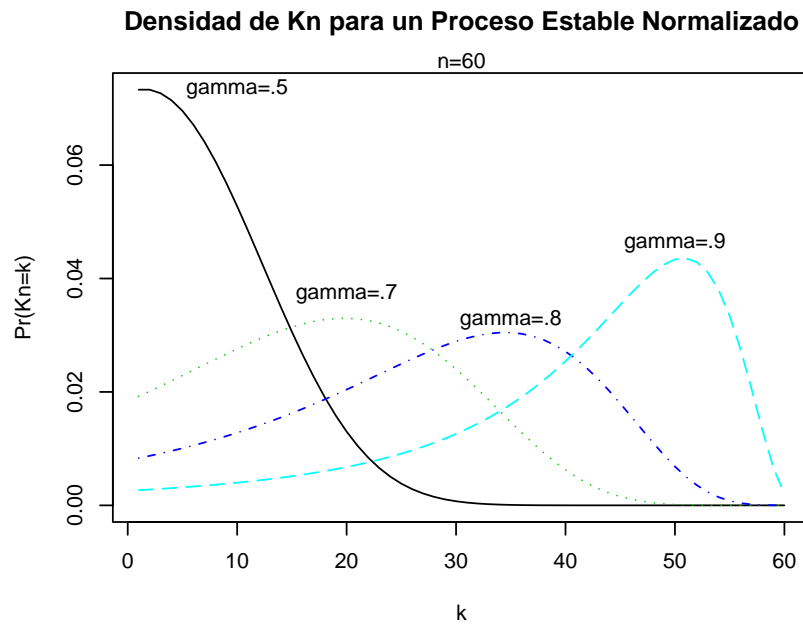


Figura 2.3: Ejemplos de  $Pr(K_n = k)$  considerando un proceso Estable normalizado.

## Proceso Gaussiano Inverso normalizado

Lijoi, Mena, y Prünster [2005b] demuestran que la distribución inicial de  $K_n$  es

$$Pr(K_n = k|n) = \binom{2n-k-1}{n-1} \frac{e^a (-a^2)^{n-1}}{2^{2n-k-1} \Gamma(k)} \sum_{r=0}^{n-1} \binom{n-1}{r} (-a^2)^{-r} \Gamma(k+2+2r-2n; a). \quad (2.43)$$

La figura 2.4 muestra algunos ejemplos de la distribución inicial de  $K_n$ .

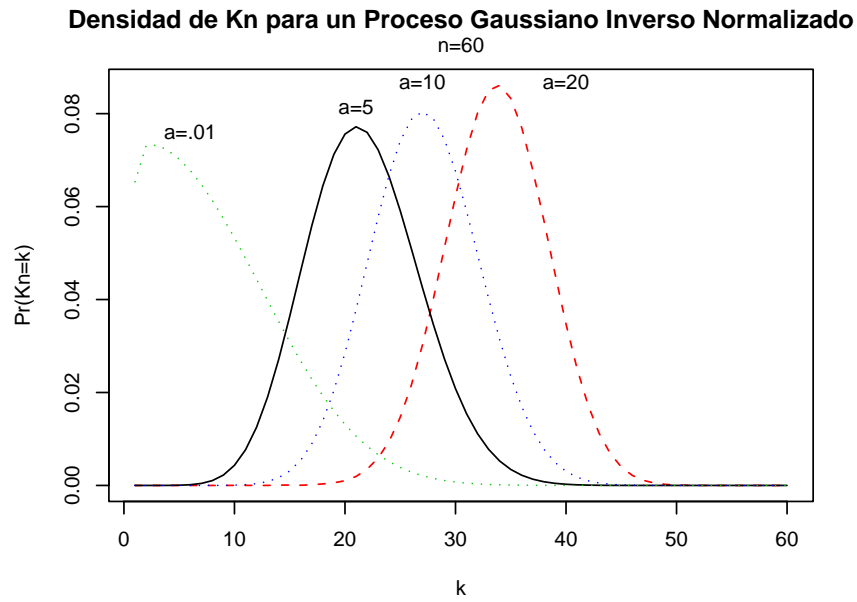


Figura 2.4: Ejemplos de  $Pr(K_n = k)$  considerando un proceso de Gaussiano Inverso normalizado

Observemos que, con este proceso, los valores máximos de las densidades no pasan de 0.09, como lo hace el proceso Dirichlet. Además ninguna de las densidades graficadas tiene un valor en su moda menor a 0.06, a diferencia del proceso Estable normalizado. Adicionalmente, este último proceso tiene densidades cuyo valor en la moda llega a alrededor de 0.03, tal y como se muestra en la figura 2.3.

Es importante mencionar que  $Pr(K_n = k|n)$  es de utilidad para la especificación inicial de la distribución N-prior. Para los procesos N-prior aquí presentados, a través de la distribución de  $K_n$  se pueden obtener los valores de  $a$  o de  $\gamma$ .

# Capítulo 3

## Aplicaciones de los procesos normalizados

Las primeras dos aplicaciones que presentaremos en este capítulo son ejemplos sencillos donde se consideran mezclas de distribuciones normales con varianzas iguales entre sí para no afectar las comparaciones con otro tipo de ruido.

El modelo jerárquico que se tiene para estas dos aplicaciones es

$$\begin{aligned} Y_i | X_i, 1 &\stackrel{\text{iid}}{\sim} N(Y_i | X_i, 1) \quad i = 1, \dots, n \\ X_i &\stackrel{\text{iid}}{\sim} \tilde{P} \\ \tilde{P} &\sim \text{N-prior.} \end{aligned}$$

Considerando como media inicial del proceso,  $G_0 = N(X_i | \bar{Y}, t^3)$ , donde,  $t^3 = (\text{máx}(Y_i) - \text{mín}(Y_i))^3$ .

Los algoritmos para realizar estas aplicaciones se implementaron en R y se puede ver mayor detalle en el apéndice.

### 3.1. Mezcla uniforme de 3 distribuciones normales

En esta subsección se replica el ejemplo de Lijoi, Mena, y Prünster [2005b], donde se toma una muestra de tamaño 100 de una mezcla uniforme de tres normales con varianza igual a 1 y medias -4, 0, y 8 respectivamente.

Cabe mencionar que en el trabajo antes mencionado solo hace comparación entre el proceso Dirichlet y el proceso IG normalizado, aquí adicionalmente se hará la comparación con el proceso Estable normalizado.

El histograma de la muestra de la mezcla se encuentra en la figura 3.1.

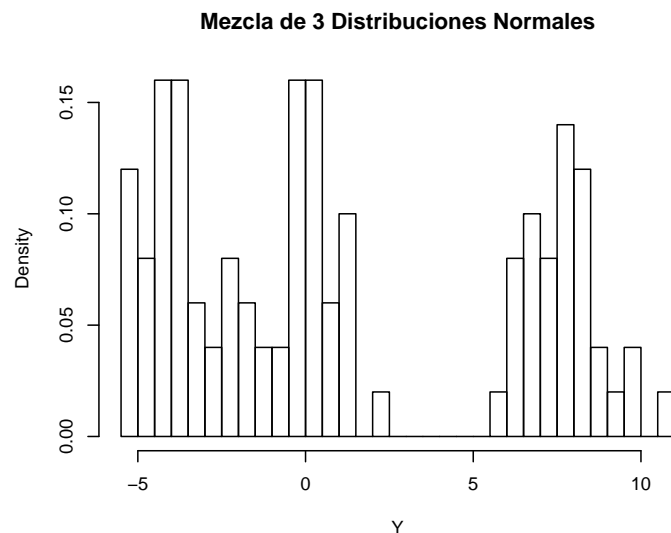


Figura 3.1: Muestra de una mezcla de tres distribuciones normales.

Para determinar los parámetros de la distribución N-Prior, aparte de  $G_0$ , se consideraron dos modelos; el primero asigna una distribución inicial para  $K_n$  con moda en  $k = 3$  (el valor verdadero del número de grupos) y el segundo modelo tiene moda en  $k = 8$ .

Para el primer modelo observamos que los valores de los parámetros de cada proceso para

obtener una moda en  $k = 3$  son:

Proceso	Parámetros
Dirichlet	$a = 0.4316$
Estable normalizado	$\gamma = 0.52$
Gaussiano Inverso normalizado	$a = 0.0109$

Las gráficas de las distribuciones iniciales de  $K_n$  para cada uno de los tres procesos normalizados están en la figura 3.2. En ella se puede observar que el proceso Dirichlet le da a la distribución inicial de  $K_n$  menos dispersión, en cambio para los otros dos procesos se observa que la distribución es casi plana para los primeros 10 valores de  $K_n$ .

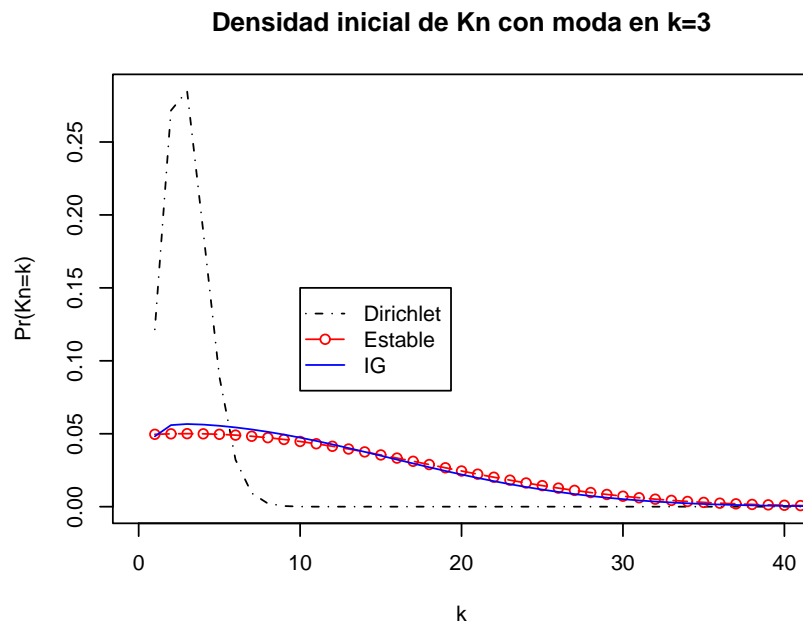


Figura 3.2: Distribuciones iniciales de  $K_n$  para los tres procesos normalizados.

Sin considerar a  $a$  ó  $\gamma$  como variables aleatorias, una densidad alta en un cierto número de grupos,  $k$ , implica que la distribución aleatoria, N-prior, tendrá un porcentaje alto de realizaciones con  $k$  grupos.

Al ajustar el modelo jerárquico para obtener la distribución predictiva final a través del



esquema de la Urna de Pólya (algoritmo de Ishwaran y James [2001]), considerando un período de calentamiento de 2,000 iteraciones y generando 10,000 muestras de tamaño 100, tenemos que las gráficas de esta distribución con los tres procesos son muy parecidos al histograma de la muestra, y únicamente presenta una pequeña diferencia cuando se aplica el proceso Dirichlet, lo cual se puede observar en la figura 3.3.

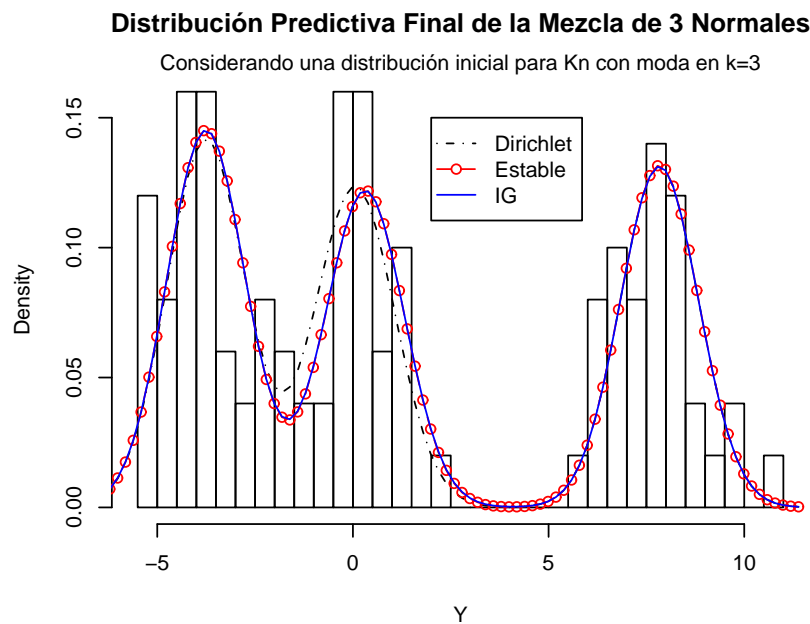


Figura 3.3: Distribución predictiva final para los tres procesos normalizados, considerando una moda en  $k = 3$  para la distribución inicial de  $K_n$ .

Con respecto a la distribución final de  $K_n$ , podemos observar que los tres procesos tienen una moda en 3 y en particular el proceso Dirichlet da una distribución final más dispersa en comparación con los otros 2 procesos (figura 3.4).

Además el proceso Estable normalizado le da una probabilidad igual a 1 al evento  $k = 3$ . De hecho, se observó que este proceso desde la primera iteración después del período de calentamiento arrojaba que el número de valores distintos de  $X's$  era 3.

El cuadro 3.1 muestra las probabilidades de la distribución inicial y distribución final para

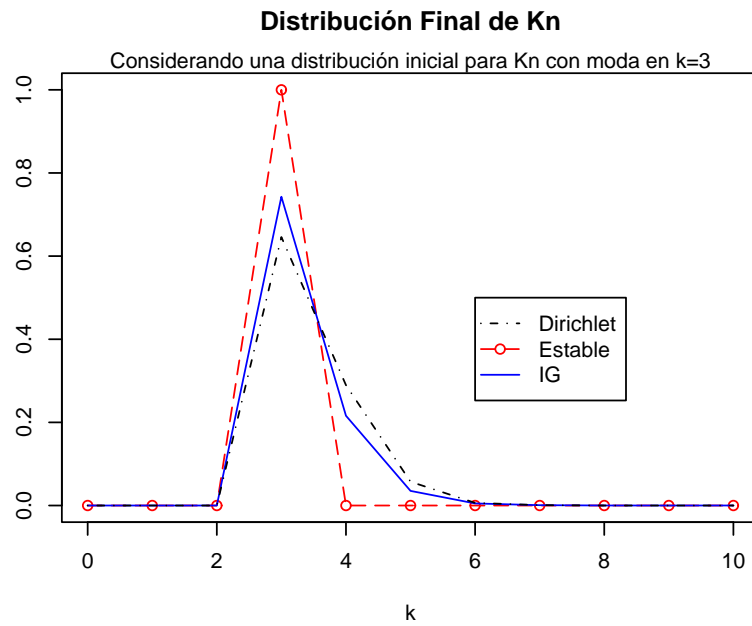


Figura 3.4: Distribución final de  $K_n$  para los tres procesos normalizados, considerando una moda en  $k = 3$  para la distribución inicial de  $K_n$ .

diferentes valores de  $K_n$ .

Como se puede ver en el cuadro 3.1, el proceso Estable normalizado es el que le da mayor probabilidad a  $k = 3$  siguiéndole el proceso IG normalizado y por último el proceso Dirichlet.

k	Distribución inicial			Distribución final		
	Dirichlet	Estable Normalizado	IG Normalizado	Dirichlet	Estable Normalizado	IG Normalizado
1	0.121552	0.049619	0.048197	0.000000	0.000000	0.000000
2	0.271616	0.049947	0.055922	0.000000	0.000000	0.000000
3	<b>0.284961</b>	<b>0.050060</b>	<b>0.056663</b>	<b>0.645900</b>	<b>1.000000</b>	<b>0.742800</b>
4	0.188596	0.049952	0.056271	0.289600	0.000000	0.216000
5	0.089082	0.049619	0.055459	0.057200	0.000000	0.035200
6	0.032170	0.049063	0.054337	0.006400	0.000000	0.005000
7	0.009283	0.048285	0.052939	0.000900	0.000000	0.000800
8	0.002208	0.047293	0.051291	0.000000	0.000000	0.000100
9	0.000443	0.046095	0.049414	0.000000	0.000000	0.000000
10	0.000076	0.044705	0.047336	0.000000	0.000000	0.000100

Cuadro 3.1: Distribución inicial y final de  $K_n$  al considerar una inicial con moda en  $k = 3$ .

Pasando al segundo modelo, donde se da una distribución inicial de  $K_n$  con moda en  $k = 8$ , los parámetros son:

Proceso	Parámetros
Dirichlet	$a = 1.9$
Estable normalizado	$\gamma = 0.57$
Gaussiano Inverso normalizado	$a = 0.2365$

Las distribuciones iniciales de  $K_n$  para los 3 procesos están en la figura 3.5.

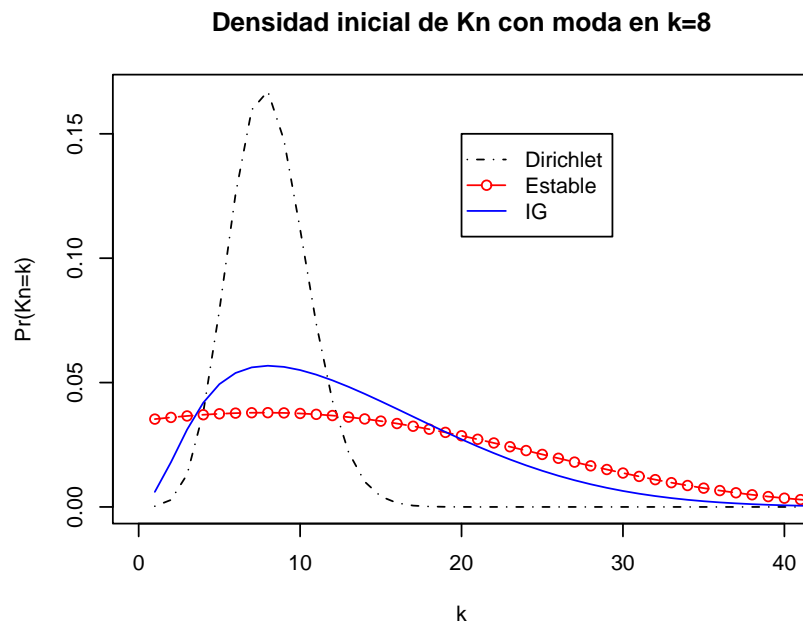


Figura 3.5: Distribución inicial de  $K_n$  con moda en  $k = 8$ .

La figura 3.5 muestra claramente que el proceso Estable normalizado es el que tiene una mayor dispersión en comparación de los otros dos procesos.

La distribución predictiva final para cada uno de los tres procesos se presenta en la figura 3.6, donde se puede ver que el proceso Dirichlet presenta un desfase a la izquierda en la segunda moda en comparación con los otros dos procesos, este desfase no es muy grande. También, como en el primer caso, el proceso Estable normalizado y IG nor-

malizado le dan una menor densidad al valle que se encuentra entre la primera y segunda moda, en comparación con el proceso Dirichlet.

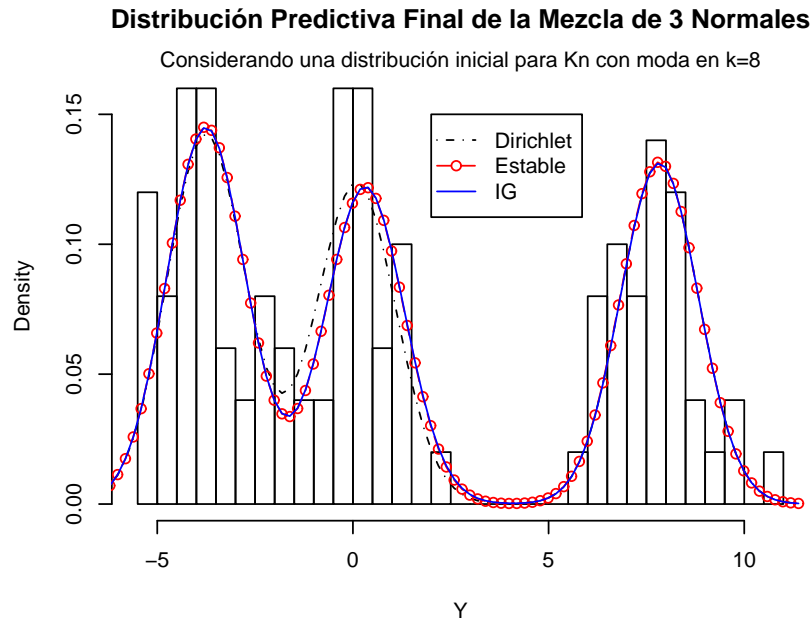


Figura 3.6: Distribución predictiva final para los tres procesos normalizados considerando una distribución inicial de  $K_n$  con moda en  $k = 8$ .

Con respecto a la distribución final de  $K_n$ , podemos observar que los tres procesos tienen una moda en 3 y que el proceso Estable normalizado le da una probabilidad igual a 1 a este valor de  $K_n$  (figura 3.7).

En el cuadro 3.2 se puede observar que el proceso Estable normalizado es el que le da una mayor probabilidad a  $k = 3$ , siguiéndole el proceso IG normalizado y por último el Dirichlet. También cabe mencionar que el proceso Estable normalizado desde la primera iteración asignó  $k = 3$  y que el proceso Dirichlet le da una densidad alta a  $k = 4$  en comparación con los otros dos procesos.

En esta aplicación se puede observar que si se asigna una distribución inicial de  $K_n$  con una moda lejana por arriba del número de componentes de la mezcla original o igual a ésta, la

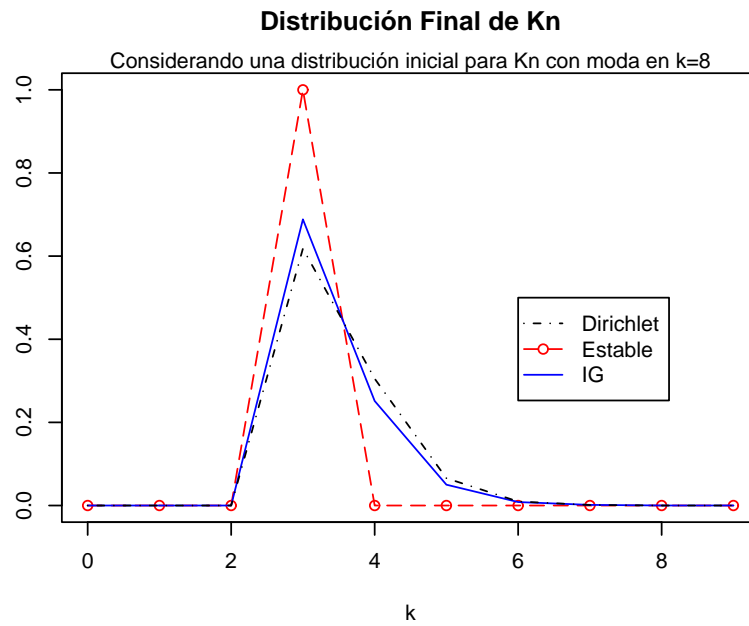


Figura 3.7: Distribución final de  $K_n = k$  considerando una distribución inicial con moda en  $k = 8$ .

k	Distribución inicial			Distribución final		
	Dirichlet	Estable Normalizado	IG Normalizado	Dirichlet	Estable Normalizado	IG Normalizado
1	0.000287	0.035316	0.006057	0.000000	0.000000	0.000000
2	0.002825	0.035989	0.017985	0.000000	0.000000	0.000000
3	0.013046	0.036575	0.031101	<b>0.618100</b>	<b>1.000000</b>	<b>0.688400</b>
4	0.038011	0.037063	0.041903	0.305200	0.000000	0.251500
5	0.079040	0.037446	0.049365	0.065700	0.000000	0.050100
6	0.125653	0.037717	0.053856	0.009700	0.000000	0.008300
7	0.159619	0.037871	0.056108	0.001300	0.000000	0.001500
8	<b>0.167105</b>	<b>0.037902</b>	<b>0.056767</b>	0.000000	0.000000	0.000100
9	0.147503	0.037807	0.056304	0.000000	0.000000	0.000100

Cuadro 3.2: Distribución inicial y final de  $K_n$  al considerar una inicial con moda en  $k = 8$ .

distribución final de  $K_n$  será muy cercana a la moda original. Además el proceso Estable normalizado le da una probabilidad igual a 1 al número de componentes de la mezcla en la distribución final de  $K_n$ . En cambio los otros dos procesos le dan una mayor densidad sin llegar nunca a 1.

En la siguiente subsección veremos qué pasa cuando se asigna una moda inicial lejana por abajo del número de componentes de la mezcla uniforme de donde provienen los datos.

### 3.2. Mezcla uniforme de 6 distribuciones normales

En esta aplicación se tomó una muestra de tamaño 180 de una mezcla de seis distribuciones normales igualmente probables con varianza 1 y medias -10, -5, 0, 5, 10, y 15 respectivamente, y se le asignó una distribución inicial de  $K_n$  con moda en  $k = 4$ . El histograma de esta muestra se presenta en la figura 3.8.

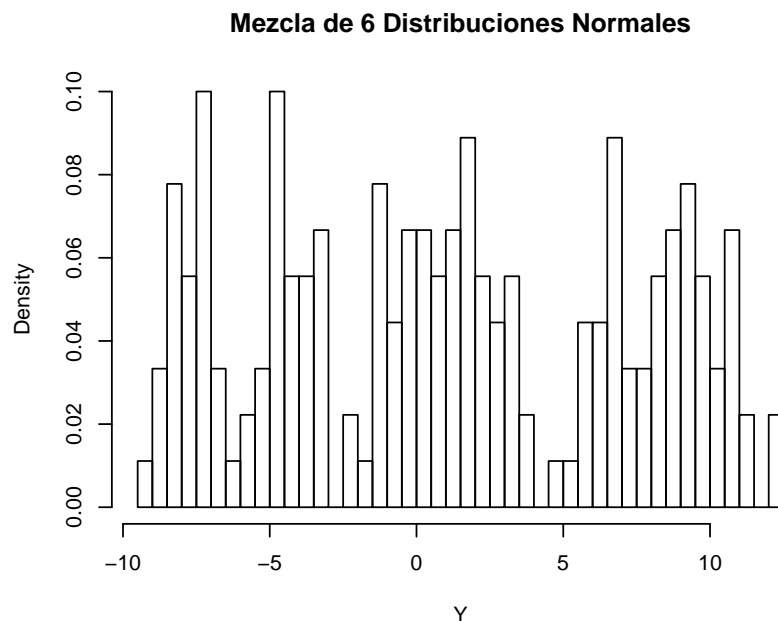


Figura 3.8: Muestra de una mezcla de seis Distribuciones Normales.

Para lograr que la distribución inicial de  $K_n$  para cada proceso tuviera una moda en  $k = 4$

se tomaron los siguiente parámetros;

Proceso	Parámetros
Dirichlet	$a = 0.5911$
Estable normalizado	$\gamma = 0.528$
Gaussiano Inverso normalizado	$a = 0.2365$

Las gráficas de las distribuciones iniciales están en la figura 3.9, donde se puede observar que las distribuciones del proceso Estable normalizado y IG normalizado son muy parecidas, y la distribución correspondiente al proceso Dirichlet tiene poca varianza en comparación con los otros dos procesos.

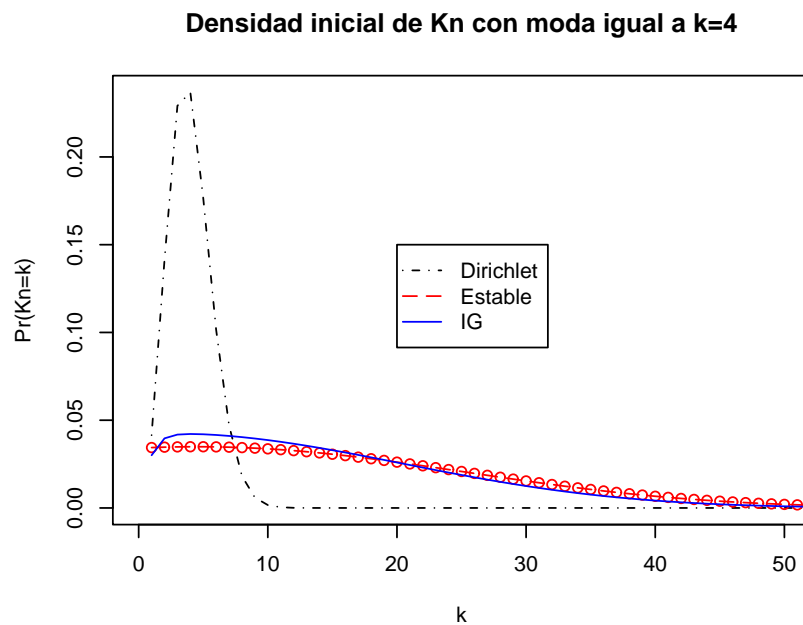


Figura 3.9: Distribuciones iniciales para  $K_n$  para los tres procesos normalizados con moda en  $k = 4$

Al ajustar el modelo jerárquico para obtener la distribución predictiva final a través del esquema de la Urna de Pólya con un período de calentamiento de 2,000 muestras y considerando 10,000 muestras, tenemos que las distribuciones predictivas finales del proceso Dirichlet y IG normalizado son muy parecidas entre sí, mientras que el proceso Estable



normalizado difiere de los otros dos procesos (ver figura 3.10).

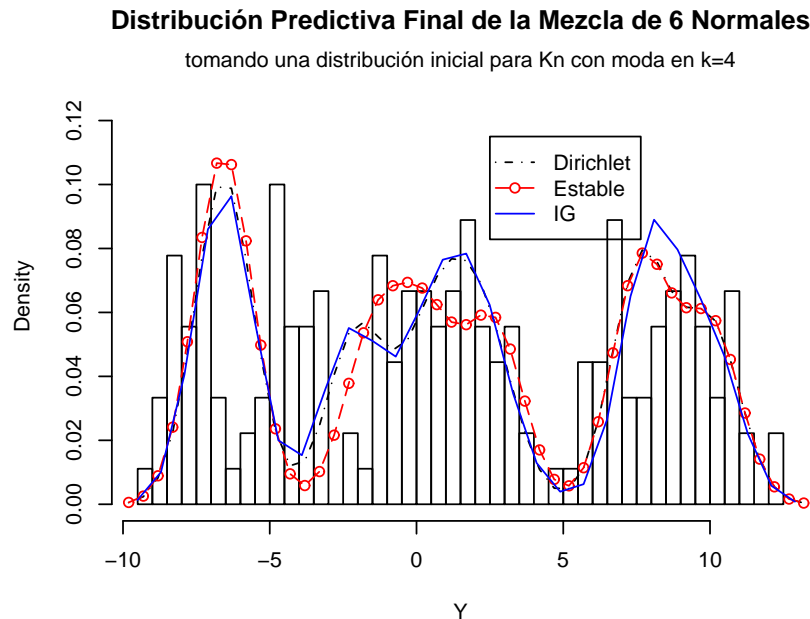


Figura 3.10: Distribución predictiva final para los tres procesos normalizados considerando una distribución inicial de  $K_n$  con moda en  $k=4$ .

La distribución final de  $K_n$  para los tres procesos se presenta en la figura 3.11, la cual muestra que tanto el proceso Estable normalizado como el proceso Dirichlet le dan una mayor probabilidad a  $k = 6$  y el proceso IG normalizado se lo da a  $k = 7$ . Esto también se puede ver en el cuadro 3.3.

De esta aplicación se puede concluir que, aunque se asigne una moda para la distribución inicial de  $K_n$  por abajo del número de componentes de la mezcla de donde provienen los datos, en el caso de los procesos Dirichlet y Estable normalizado se llega al número de componentes. Por otro lado para el proceso IG normalizado se tiene un valor aproximado de este valor.

Además, en el caso del proceso Estable normalizado bastaron menos de las 100 primeras iteraciones del período de calentamiento para que empezara a arrojar que la moda de la

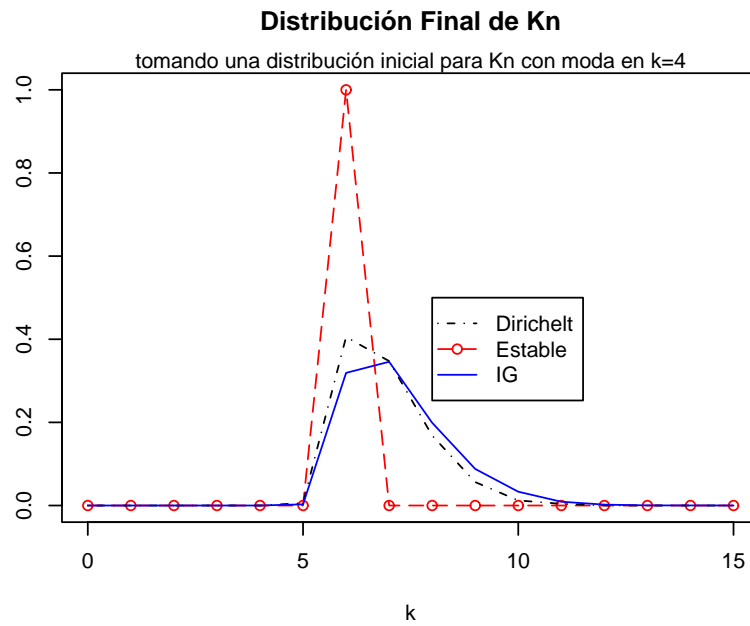


Figura 3.11: Distribución final de  $K_n = k$  considerando una distribución inicial con moda en  $k = 4$

k	Distribución inicial			Distribución final		
	Dirichlet	Estable Normalizado	IG Normalizado	Dirichlet	Estable Normalizado	IG Normalizado
1	0.041479	0.034425	0.030018	0.000000	0.000000	0.000000
2	0.141406	0.034651	0.039666	0.000000	0.000000	0.000000
3	0.229156	0.034803	0.041811	0.000000	0.000000	0.000000
4	<b>0.236839</b>	<b>0.034878</b>	<b>0.042152</b>	0.000000	0.000000	0.000000
5	0.176444	0.034876	0.041984	0.006000	0.000000	0.002900
6	0.101433	0.034795	0.041588	<b>0.405400</b>	<b>1.000000</b>	0.319100
7	0.047005	0.034634	0.041039	0.347400	0.000000	<b>0.345800</b>
8	0.018103	0.034395	0.040361	0.168300	0.000000	0.199100
9	0.005926	0.034077	0.039570	0.056400	0.000000	0.088000

Cuadro 3.3: Distribución inicial y final de  $K_n$  al considerar una inicial con moda en  $k = 4$ .

distribución final de  $K_n$  es 6. Entonces con estas dos aplicaciones podemos decir que el proceso Estable normalizado converge rápidamente al número de componentes de la mezcla,  $K_n$ .

Se realizaron tres aplicaciones adicionales; una de ellas considero una mezcla no uniforme de distribuciones normales con medias distantes entre ellas y se obtuvieron resultados semejantes a las dos aplicaciones anteriores. En las otras dos aplicaciones se consideraron distribuciones normales con medias muy cercanas, por ejemplo 1, 1.1, 1.3, y diferían en que una consideraba una mezcla uniforme y la otra no. La moda de la distribución final de  $K_n$  en ambos casos fue igual a uno, es decir estos tres procesos no distinguen el número de componentes en una mezcla cuando se tienen medias muy cercanas.

La siguiente aplicación será con datos reales que se han usado en literatura Bayesiana no paramétrica.

### 3.3. Datos de la galaxias

En esta aplicación se consideró la base de datos de la medidas de velocidad (1000 km/s) de 82 galaxias que divergen a nuestra galaxia. A partir de la teoría del Big Bang se sabe que existen puntos de atracción, y varios astrónomos han demostrado que existen conglomerados que son formados por estos puntos, pero hay datos que sugieren que existen superconglomerados, tema que aún no se demuestra.

Estos datos fueron dados a conocer en Roeder [1990] y han sido utilizados en diversos artículos para ilustrar diferentes modelos de mezclas. Entre estos trabajos destacan Escobar y West [1995], donde concluyen que hay entre 3 y 7 conglomerados, y Lijoi, Mena, y Prünster [2006a], donde aplican las N-Priors a estos datos y mencionan que los resultados dentro de la literatura arrojan resultados entre 3 y 8 para el número de conglomerados.

El modelo que se utilizó en esta aplicación es un modelo de mezclas de normales con media y varianza desconocidas, es decir,

$$Y_i | m_i, V_i \stackrel{\text{ind}}{\sim} N(Y_i | m_i, V_i) \quad i = 1, \dots, 82$$

$$m_i, V_i \stackrel{\text{iid}}{\sim} \tilde{P}$$

$$\tilde{P} \sim \text{N-prior}$$

Considerando a la media inicial del proceso,  $G_0 = N(m_i | \mu, \frac{V_i}{\tau}) Ga(V_i^{-1} | 2, 1)$  y

$$\mu \sim N(\mu | 0, 0.001)$$

$$\tau \sim Ga(\tau | 1, 100)$$

El histograma de estos datos se presenta en la figura 3.12.

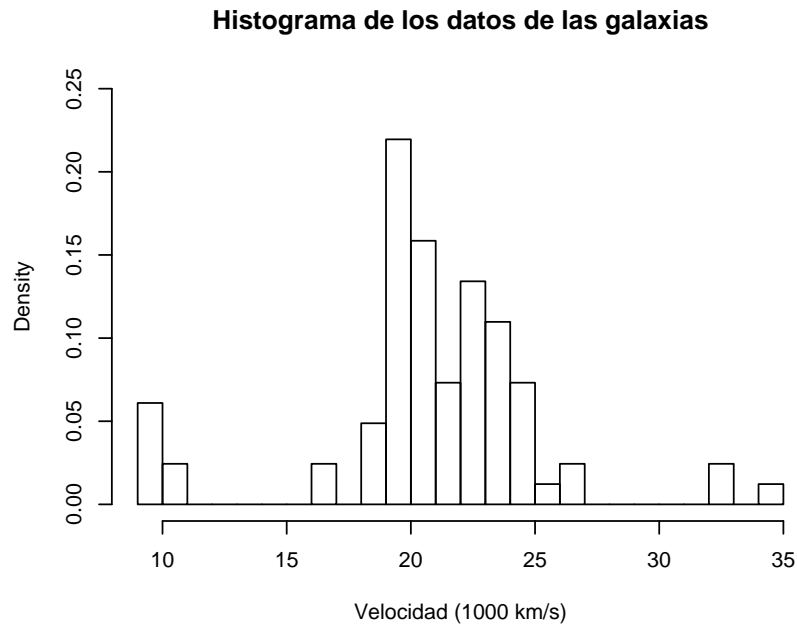


Figura 3.12: Datos de las Galaxias.

Se consideraron dos distribuciones iniciales para  $K_n$ , la primera con una moda en  $k = 5$  y la segunda con  $k = 25$ . Además, sólo se consideraron los procesos Dirichlet y el proceso Estable normalizado; el proceso IG normalizado lo excluimos ya que entre éste y el proceso Estable normalizado, este último se aproximó mejor al número de componentes de las

mezclas en las dos aplicaciones anteriores.

Para lograr que las distribuciones iniciales de  $K_n$  para cada proceso tuviera una moda en  $k = 5$  se tomaron los siguiente parámetros;

Proceso	Parámetros
Dirichlet	$a = 1.003$
Estable normalizado	$\gamma = 0.5489$

En la figura 3.13 (distribuciones iniciales), se puede observar que el proceso Estable normalizado tiene una distribución mas “plana”, y el proceso Dirichlet le da una mayor densidad a  $k = 5$ .

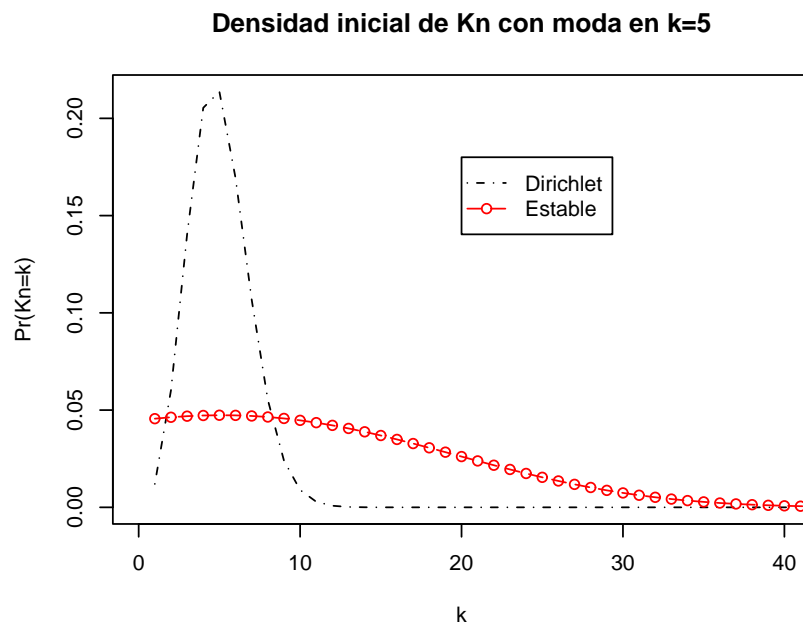


Figura 3.13: Distribuciones iniciales para  $K_n$  para los dos procesos normalizados con moda en  $k = 5$  para los datos de las galaxias.

Se ajustó el modelo jerárquico descrito en la sección 2.4, considerando un período de calentamiento de 2,000 muestras y generando 10,000 muestras para la obtención de la distribución predictiva final para cada uno de los dos procesos: proceso Dirichlet y Estable

normalizado. La figura 3.14, muestra que la distribución predictiva final para los procesos Dirichlet y Estable normalizado al inicio son iguales, pero en la parte donde existe una mayor densidad empiezan a tener diferencias. Además, se observa que en esta parte el proceso Dirichlet tiene un pequeña asimetría a la izquierda, en los valores de 21 a 23.

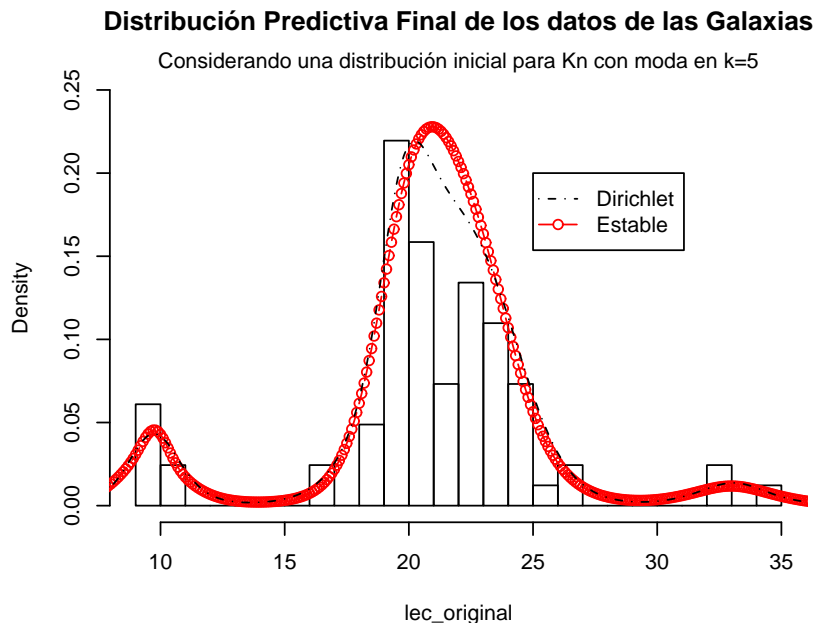


Figura 3.14: Distribución predictiva final para los dos procesos normalizados considerando una distribución inicial de  $K_n$  con moda en  $k=5$  para los datos de las galaxias.

En el caso de la distribución final para  $K_n$ , se observa en la figura 3.15 que el proceso Dirichlet le da una mayor densidad al valor de  $k = 6$ , mientras el proceso Estable normalizado se lo da al 8.

Se puede ver claramente en el cuadro 3.4 cómo cada proceso le da una mayor densidad a un valor diferente de  $K_n$ , y que la moda del proceso Dirichlet tiene mayor probabilidad en comparación con la moda resultante del proceso Estable normalizado.

Ahora, veamos qué pasa cuando se asigna una distribución inicial con moda en  $k = 25$ , para lo cual se consideraron los siguientes parámetros;

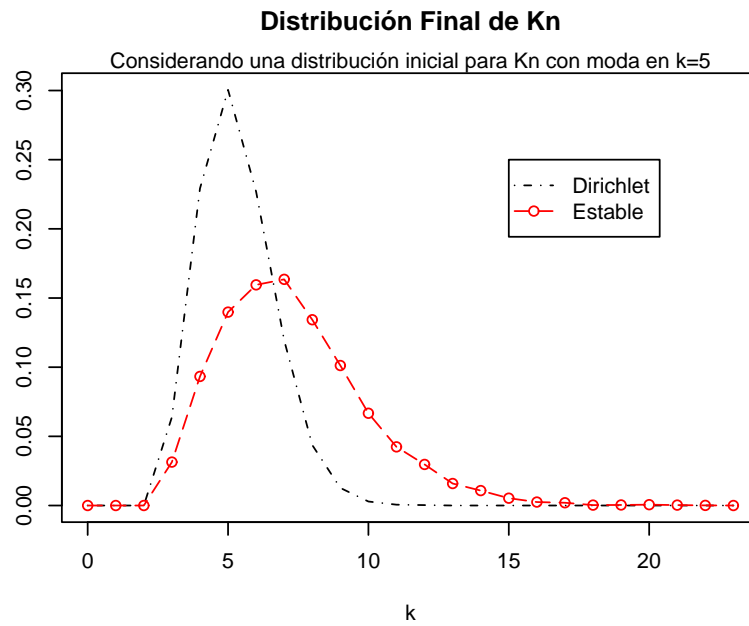


Figura 3.15: Distribución final de  $K_n$  considerando una distribución inicial con moda en  $k = 5$  para los datos de las galaxias.

Proceso	Parámetros
Dirichlet	$a = 11.8787$
Estable normalizado	$\gamma = 0.707$

La distribución inicial que es menos informativa es la del proceso Estable normalizado, figura (3.16).

Considerando un período de calentamiento de 2,000 iteraciones, se tomaron 10,000 muestras del modelo jerárquico para obtener la distribución predictiva final y la distribución final de  $K_n$  para cada uno de los dos procesos.

La figura 3.17 muestra la distribución predictiva final y se observa que las dos distribuciones tiene una asimetría hacia la izquierda.

En el caso de la distribución final para  $K_n$ , se observa en la figura 3.18 que el proceso

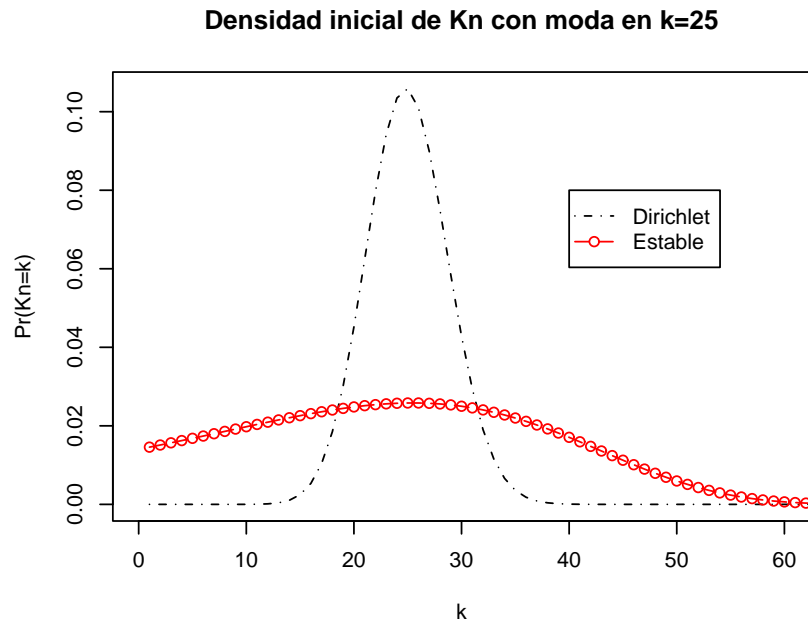


Figura 3.16: Distribuciones iniciales para  $K_n$  para los dos procesos normalizados con moda en  $k = 25$  para los datos de las galaxias.

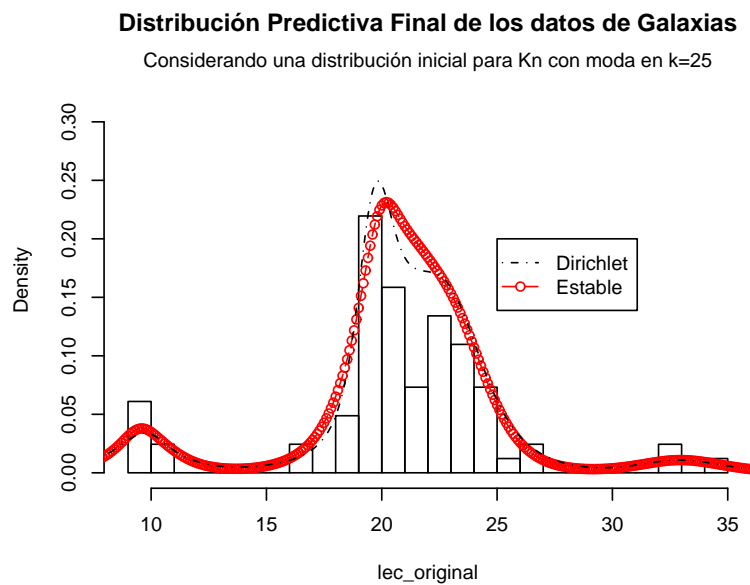


Figura 3.17: Distribución predictiva final para los dos procesos normalizados considerando una distribución inicial de  $K_n$  con moda en  $k = 25$  para los datos de las galaxias.



k	Distribución Inicial		Distribución Final	
	Dirichlet	Estable Normalizado	Dirichlet	Estable Normalizado
1	0.012050	0.045580	0.000000	0.000000
2	0.060163	0.046317	0.000000	0.000000
3	0.140294	0.046868	0.000000	0.000000
4	0.205417	0.047220	0.063900	0.031400
5	<b>0.213734</b>	<b>0.047362</b>	0.229200	0.093300
6	0.169333	0.047285	<b>0.300500</b>	0.139800
7	0.167825	0.046984	0.226800	0.159500
8	0.055284	0.046456	0.119400	<b>0.163500</b>
9	0.024044	0.045702	0.043700	0.134300

Cuadro 3.4: Distribución inicial y final de  $K_n$  al considerar una inicial con moda en  $k = 5$  para los datos de las galaxias.

Estable normalizado asigna una moda menor a 10.

En el cuadro 3.5 se puede ver claramente cómo cada proceso le da una mayor densidad a un valor distinto de  $K_n$ , pero siempre el proceso Estable normalizado está fuera del rango de 3 a 7 que proponen Escobar y West [1995], aunque sí está dentro del rango que se ha observado dentro de la literatura que mencionan Lijoi, Mena, y Prünster [2006a].

Con los datos de las galaxias, se observó que el proceso Estable normalizado asigna una mayor densidad a  $k = 8$  aún cuando se le asigna una distribución inicial de  $K_n$  con moda en  $k = 25$ . Este mismo proceso asigna un número de conglomerados cercano al intervalo entre 3 y 8.

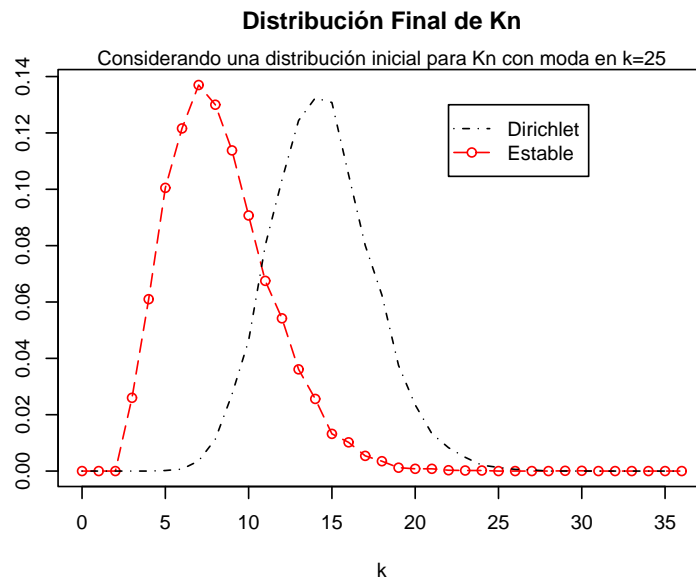


Figura 3.18: Distribución final de  $K_n$  considerando una distribución inicial con moda en  $k = 25$  para los datos de las galaxias.

El proceso Dirichlet se ve afectado grandemente por la distribución inicial de  $K_n$  que se asigne. A manera de ilustración, cuando se asigna una moda de  $k = 25$ , este proceso obtiene alrededor de 15 conglomerados.

k	Distribución Inicial		Distribución Final	
	Dirichlet	Estable Normalizado	Dirichlet	Estable Normalizado
4	0.000000	0.016242	0.000000	0.026000
5	0.000000	0.016819	0.000000	0.061000
6	0.000000	0.017403	0.000200	0.100500
7	0.000000	0.017992	0.000700	0.121600
8	0.000000	0.018585	0.003600	<b>0.137000</b>
9	0.000000	0.019179	0.011300	0.130000
10	0.000010	0.019771	0.027200	0.113800
11	0.000040	0.020358	0.046300	0.090700
12	0.000133	0.020938	0.080100	0.067500
13	0.000395	0.021506	0.103300	0.054200
14	0.001045	0.022059	0.124400	0.036100
15	0.002483	0.022593	<b>0.132400</b>	0.025600
16	0.005333	0.023103	0.130800	0.013200
17	0.010401	0.023586	0.105200	0.010200
18	0.018507	0.024035	0.080100	0.005400
19	0.030158	0.024446	0.062500	0.003500
20	0.045168	0.024814	0.037500	0.001200
21	0.062367	0.025133	0.023500	0.000800
22	0.079610	0.025399	0.013300	0.000800
23	0.094182	0.025605	0.008300	0.000300
24	0.103498	0.025747	0.004900	0.000200
25	<b>0.105860</b>	<b>0.025820</b>	0.002100	0.000200
26	0.100966	0.025818	0.001300	0.000000
27	0.089957	0.025738	0.000600	0.000000
28	0.074956	0.025576	0.000400	0.000000

Cuadro 3.5: Distribución inicial y final de  $K_n$  al considerar una inicial con moda en  $k = 25$  para los datos de las galaxias.

# Comentarios finales

La estadística Bayesiana no paramétrica resulta ser robusta en comparación a la estadística Bayesiana paramétrica porque no se tienen que hacer supuestos demasiados restrictivos sobre las distribuciones de los datos. Además, la estadística Bayesiana no paramétrica toma las distribuciones iniciales sobre un espacio de funciones de densidad y es más consistente con el teorema de de Finetti.

La presente tesis exploró alternativas al proceso Dirichlet llamadas distribuciones N-prior, las cuales consisten en la normalización de procesos de Lévy e incluyen al proceso Dirichlet. Además se expusieron las mezclas de procesos normalizados, observando que se pueden implementar los algoritmos de simulación basados en el esquema de la Urna de Pólya del proceso Dirichlet sustituyéndolo por un proceso normalizado.

En particular, en este trabajo se consideraron tres procesos normalizados: proceso Gamma (Dirichlet), proceso Estable y proceso Gaussiano Inverso, los cuales se utilizaron para ejemplificar las propiedades de las distribuciones N-prior, así como para el desarrollo de las aplicaciones presentadas. Las aplicaciones se realizaron con datos simulados así como con la base de datos de las velocidades de las galaxias que se ha utilizado dentro de la literatura de la estadística Bayesiana no paramétrica.

Adicionalmente se mostró la importancia de la distribución del número de valores distintos en una muestra,  $K_n$ , ya que ayuda para determinar los valores de los parámetros del

proceso normalizado, así como para encontrar el número de componentes en una mezcla. Actualmente se sigue investigando otros usos para la distribución de  $K_n$ .

Dentro de los tres procesos presentados en esta tesis, se observó que la distribución inicial de  $K_n$  para el proceso Estable normalizado y el proceso IG normalizado es casi plana y la distribución final de esta misma variable para estos procesos resultó tener una moda cercana o igual al número de componentes de la mezcla de donde provenían los datos.

Para el caso del proceso Dirichlet, la distribución final de  $K_n$  resultó, en las dos primeras aplicaciones presentadas aquí, ser afectada por la distribución inicial de esta variable ya que al dar una moda inicial diferente del número de componentes de la mezcla de donde provenían los datos, este proceso arrojaba una moda final para  $K_n$  contraria al número de distribuciones involucradas en cada una de las mezclas de las aplicaciones.

Los procesos presentados se pueden implementar de una manera sencilla, aunque desde mi punto de vista el proceso Estable normalizado arroja resultados cercanos a los valores de donde se obtuvieron las muestras y además converge con mayor velocidad en comparación con los otros dos procesos. Aunque cabe destacar que la distribución predictiva final de este proceso no en todas las aplicaciones se ajustó a los datos completamente.

Existen otras familias de distribuciones que se pueden usar dentro de la estadística Bayesiana no paramétrica; las distribuciones Gamma Generalizada, ver Lijoi, Mena, y Prünster [2005a], que incluye al proceso Estable normalizado y al IG normalizado; y las distribuciones Dirichlet Dependientes, para mayor referencia ver MacEachern [1999] y Griffin y Steel [2004].

# Apéndice

Los algoritmos se realizaron en R, considerando como base el modelo jerárquico y el algoritmo de Ishwaran y James [2001], y los pesos del proceso IG normalizado se obtuvieron en PARI porque tiene una mayor precisión y es capaz de manipular números más grandes y más pequeños en comparación con R.

Ambos paquetes computacionales se obtienen gratuitamente. R se obtiene de <http://www.r-project.org/> y PARI de <http://pari.math.u-bordeaux.fr/download.html>.

El algoritmo en R resulto ser lento porque se tarda entre 2 y 3 minutos en promedio para obtener 1,000 muestras.

## Algoritmo para la mezcla de normales.

Dada una muestra de tamaño  $n$  de  $Y$ , para generar la distribución predictiva final tenemos:

1. Generar una muestra de tamaño  $n$  de  $X_i^{(0)} \sim N(X_i | \bar{Y}, t^3)$
2. Obtener el vector  $X_{-i}^{(j-1)}$ .
3. Calcular el valores de  $k^{(j-1)}$ ,  $n_j^{(j-1)}$ 's y  $X_j^{*(j-1)}$ 's.
4. Calcular los pesos del proceso seleccionado,  $w^{(n-1)(j-1)}$  y  $w_j^{(n-1)(j-1)}$ .
5. Generar los pesos  $q_0^{*(j-1)}$  y  $q_j^{*(j-1)}$ , sabiendo que:

a)  $P(X_i \in \cdot) \propto N(Y_i|X_i, 1) N(X_i|\bar{Y}, t^3)$ , y de donde se obtiene que  $P(X_i \in \cdot) = N\left(X_i|\frac{t^3 Y_i + \bar{Y}}{t^3 + 1}, \frac{t^3 + 1}{t^3}\right)$ .

b)  $\int f(Y_i|X, \theta) G_0(dX) = \int N\left(X|\frac{t^3 Y_i + \bar{Y}}{t^3 + 1}, \frac{t^3 + 1}{t^3}\right) = N(Y_i|\bar{Y}, t^3 + 1)$ .

6.  $X_i^{(j)}|X_{-i}^{(j-1)}, 1, Y = \begin{cases} Z \sim N\left(Z|\frac{t^3 Y_i + \bar{Y}}{t^3 + 1}, \frac{t^3 + 1}{t^3}\right) & \text{con probabilidad } q_0^* \\ X_j^{*(j-1)} & \text{con probabilidad } q_j^* \end{cases}$ .

7. Regresar a paso 2.

### Algoritmo para los datos de las galaxias.

Dada una muestra de tamaño  $n$  de  $Y$ , para generar la distribución predictiva final tenemos:

1. Generar una muestra de tamaño  $n$  de  $V_i^{-1(0)} \sim Ga(V_i^{-1}|2, 1)$
2. Generar una muestra de tamaño  $n$  de  $m_i^{(0)} \sim N\left(m_i|\bar{Y}, \frac{V_i^{(0)}}{15}\right)$ .
3.  $X_i^{(0)} = \left(m_i^{(0)}, V_i^{-1(0)}\right)$
4. Obtener el vector  $X_{-i}^{(j-1)}$
5. Calcular el valores de  $k^{(j-1)}$ ,  $n_j^{(j-1)}$ 's y  $X_j^{*(j-1)}$ 's.
6. Calcular los pesos del proceso seleccionado,  $w^{(n-1)(j-1)}$  y  $w_j^{(n-1)(j-1)}$ .
7. Generar los pesos  $q_0^{*(j-1)}$  y  $q_j^{*(j-1)}$ , sabiendo que:

a)  $P(X_i \in \cdot) \propto N(Y_i|m_i, V_i) N(m_i|\mu, \frac{V_i}{\tau}) Ga(V_i^{-1}|2, 1)$ , y de donde se obtiene que

$$P(V_i \in \cdot) = Ga\left(V_i|\frac{1}{2} + 1, 1 + \frac{1}{2} \frac{(Y_i - \bar{Y})^2}{1 + 15}\right).$$

$$P(m_{,i} \in \cdot) = N\left(m_i|\frac{\bar{Y} + 15Y_i}{15}, \frac{15}{(1 + 15)V_i}\right)$$

b)  $\int f(Y_i|X, \theta) G_0(dX) = Stu(Y_i|5, \mu, \frac{5\tau}{\tau+1})$ .

$$8. X_i^{(j)} | X_{-i}^{(j-1)}, 1, Y = \begin{cases} Z \sim F \text{ con probabilidad } q_0^*. \\ X_j^{*(j-1)} \text{ con probabilidad } q_j^*. \end{cases},$$

donde  $Z = (m, V)$  y  $F = Ga\left(V | \frac{1}{2} + 1, 1 + \frac{1}{2} \frac{(Y_i - \bar{Y})^2}{1+15}\right) N\left(m | \frac{\bar{Y} + 15Y_i}{15}, \frac{15}{(1+15)V}\right)$ .

9. Regresar a paso 4.



## Bibliografía

- C. E. Antoniak. Mixtures of Dirichlet processes with applications to nonparametric problems. *Annals of Statistics*, 2:1152–1174, 1974.
- J. Bertoin y M. Yor. On subordinators, self-similar Markov processes and some factorizations of the exponential variable. *Electron. Comm. Probab.*, 6:95–106, 2001.
- D. Blackwell y J. B. MacQueen. Ferguson distributions via polya urn schemes. *Ann. Statist.*, 1:353–355, 1973.
- K. A. Doksum. Tailfree and neutral random probabilities and their posterior distributions. *Ann. Probab.*, 2:183–201, 1974.
- M. D. Escobar y M West. Bayesian density estimation and inference using mixtures. *J. Amer. Stat. Assoc.*, 90(430):577–588, 1995.
- W. J. Ewens. The sampling theory of selectively neutral alleles. *Theoretical Population Biology*, 3:87–112, 1972.
- T.S. Ferguson. A Bayesian analysis of some nonparametric problems. *Ann. Statist.*, 1:209–230, 1973.
- D. A. Freedman. On the asymptotic behavior of Bayes' estimates in the discrete case. *Ann. Math. Statist.*, 34:1194–1216, 1963.
- J. E. Griffin y M. F. J. Steel. Order-based dependent dirichlet processes. Research Report 430, Warwick Statistics., 2004.
- H. Ishwaran y L. F. James. Gibbs sampling methods for stick-breaking priors. *J. Amer. Stat. Assoc.*, 96:161–173, 2001.
- L. F. James, A. Lijoi, y I. Prünster. Conjugacy as a distinctive feature of the Dirichlet process. *SJS*, 33:105–120, 2006.

- P. J. Lenk. The logistic normal distribution for Bayesian nonparametric, predictive densities. *J. Amer. Statist. Assoc.*, 83(402):509–516, 1988.
- A. Lijoi, R. H. Mena, y I. Prünster. Bayesian nonparametric analysis for a generalized Dirichlet process prior. *Statistical Inference for Stochastic Processes*, 8:283–309, 2005a.
- A. Lijoi, R. H. Mena, y I. Prünster. Hierarchical mixture modeling with normalized Inverse-Gaussian priors. *J. Amer. Stat. Assoc.*, 100:1278–1291, 2005b.
- A. Lijoi, R. H. Mena, y I. Prünster. Bayesian clustering in nonparametric hierarchical mixture models. *Att: del XLII Riune Scientifica SIS (Torino Italy, June 14-16)*, 2006a.
- A. Lijoi, R. H. Mena, y I. Prünster. Controlling the reinforcement in Bayesian nonparametric mixture models. *Tech. Report*, 2006b.
- A. Y. Lo. On a class of Bayesian nonparametric estimates: I, density estimates. *Ann. Statist.*, 12:351–357, 1984.
- S. MacEachern. Dependent nonparametric processes. *Proceedings of the Section on Bayesian Statistics Science, American Statistical Association*, pp 50–55, 1999.
- J. Pitman. Exchangeable and partially exchangeable random partitions. *Probab. Theory Related Fields.*, 102:145–158, 1995.
- E. Regazzini, A. Lijoi, y I Prünster. Distributional results for means of normalized random measures with independent increments. *Ann. Statist.*, 31:560–585, 2003.
- K. Roeder. Density estimation with confidence sets simplified by superclusters and voids in the galaxies. *J. Amer. Stat. Assoc.*, 85:617–624, 1990.
- K. I. Sato. Self-similar processes with independent increments. *Probab. Theory Related Fields.*, 89:285–300, 1991.
- K. I. Sato. *Lévy processes and infinitely divisible distributions*. Cambridge University Press, 1999.