



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

PROGRAMA DE MAESTRÍA Y DOCTORADO EN CIENCIAS MATEMÁTICAS Y
DE LA ESPECIALIZACIÓN EN ESTADÍSTICA APLICADA

INSTITUTO DE INVESTIGACIONES EN MATEMÁTICAS APLICADAS Y EN SISTEMAS

CONSTRUCCIÓN DE MEDIDAS DE PROBABILIDAD ALEATORIAS VÍA
NORMALIZACIÓN DE MEDIDAS COMPLETAMENTE ALEATORIAS

T E S I S

QUE PARA OPTAR POR EL GRADO DE:

MAESTRO EN CIENCIAS MATEMÁTICAS

PRESENTA:

JOSÉ ANTONIO PERUSQUÍA CORTÉS

DIRECTOR DE TESIS:

DR. RAMSÉS HUMBERTO MENA CHÁVEZ

IIMAS

CIUDAD UNIVERSITARIA, CD. MX.

AGOSTO 2017



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.



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 está protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de vídeos, 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.

A mis abuelos †,

Agradecimientos

Han sido ya 9 años de mi ingreso a la Universidad Nacional Autónoma de México. A lo largo de este tiempo he conocido y terminado de conocer a muchas personas a las que quisiera agradecer; sin embargo, el espacio no es suficiente para nombrar a cada una de ellas, por lo cual de antemano les pido una disculpa si no aparecen de manera explícita.

En primer lugar quiero agradecer a mi familia por todo el apoyo y el cariño que me han brindado. Sin su presencia no habría sido posible este trabajo que representa la culminación de la que será, sin duda alguna, una de las épocas más felices de mi vida. Gracias a mi mamá y a mi hermano por su constante comprensión y soporte, sin los cuales no estaría donde estoy actualmente. El éxito cosechado es tan suyo como mío.

Quiero agradecer de igual manera a todos mis amigos, desde aquellos que han estado conmigo desde la preparatoria, pasando por las amistades hechas en la licenciatura y finalmente a todos aquellos que pude conocer en estos dos años de maestría. Gracias por su amistad incondicional y por todos los momentos llenos de risa, diversión y hasta de estrés; son un parte fundamental de mi vida y por lo cual no hay palabras suficientes para mostrarles todo mi agradecimiento.

También quiero extender este agradecimiento a todos los profesores que tuve a lo largo de estos 9 años y que han contribuido en mi desarrollo académico. De manera muy especial quiero agradecer a Ramsés por haber sido una fuente de inspiración y apoyo durante los últimos dos años. Gracias por permitirme trabajar contigo y ser parte fundamental en todo lo que he logrado y todo lo que está por venir en un futuro no muy lejano. De igual manera quiero extender este agradecimiento a Ruth. Gracias por ser uno de los pilares en mi desarrollo académico, profesional y personal; gracias por la confianza que siempre depositaste en mi.

Finalmente, quiero agradecer a la UNAM, la cual se ha convertido en mi segundo hogar. Es un honor pertenecer a la máxima casa de estudios de la nación. Gracias por estos 9 años llenos de dicha y conocimiento. Te llevaré conmigo a donde quiere que vaya y pondré en alto tu nombre en el mundo.

“Por mi raza hablará el espíritu,”

José A.

Introducción

La estadística bayesiana no paramétrica se podría decir que tuvo sus inicios a raíz del trabajo de Ferguson (1973) titulado *A Bayesian Analysis of some Nonparametric Problems*. Han sido más de 40 años de un interés creciente en esta área de las matemáticas y con auge significativo en los años recientes, gracias a los avances computacionales. Una rápida y sencilla caracterización, como bien describen Hjort, Holmes, Müller y Walker (2010), es el hecho de trabajar con espacios parametrales grandes, probablemente de dimensión infinita y la construcción de medidas de probabilidad en ellos.

Los métodos para construir estas medidas de probabilidad aleatoria son muchos y muy variados, uno de ellos es mediante la especificación de la ley que gobierna esta medida; sin embargo, este trabajo se centra en una construcción directa de la medida. Esta construcción se hará a partir de la normalización de medidas completamente aleatorias, lo que da lugar a lo que se conoce como una medida de probabilidad aleatoria.

El concepto de medidas completamente aleatorias fue introducido por Kingman (1967), en la década de los 60 y representa una herramienta que ha sido muy utilizada en la estadística bayesiana no paramétrica. Un ejemplo claro de ello es nuevamente el trabajo de Ferguson (1973). En él se hace introducción de un proceso clave en la estadística bayesiana no paramétrica: el proceso Dirichlet y del cual Ferguson da una construcción a partir de la normalización de una medida completamente aleatoria.

Este trabajo está dividido en 4 capítulos. El Capítulo 1 sirve como una introducción, tanto histórica como teórica hacia el problema de interés; para esto, primero se describe a grosso modo lo que es la estadística bayesiana paramétrica y no paramétrica y posteriormente se detallan los resultados que servirán como la base para los capítulos posteriores. En el Capítulo 2 se detalla la teoría concerniente a la construcción vía el proceso de normalización y resultados derivados a ella, como el cálculo de la distribución predictiva, la función de probabilidad para particiones intercambiables (EPPF por sus siglas en inglés), entre otros. En el Capítulo 3 nos enfocamos en medidas de probabilidad aleatorias obtenidas a partir de medidas completamente aleatorias que ya han sido exploradas con anterioridad y que son algunas de las más utilizadas, como el proceso Dirichlet, en los trabajos de Ferguson (1973), Lijoi y Prünster (2009); el proceso estable en el trabajo de Kingman (1975); o el proceso gama generalizado descrito por primera vez por Brix (1999). Finalmente, en el Capítulo 4 se explora la normalización de la medida completamente aleatoria de Bessel, introducida en el trabajo de Argiento, Bianchini y Guglielmi (2016).

Índice

Agradecimientos	IV
Introducción	VI
1. Preliminares	1
1.1. Estadística bayesiana no paramétrica	1
1.2. Intercambiabilidad	3
1.3. Procesos de Lévy	8
1.4. Medidas completamente aleatorias (CRM)	14
2. Construcción vía normalización	32
2.1. Medidas completamente aleatorias normalizadas	33
2.2. Distribución posterior	35
2.3. Distribución predictiva	42
2.4. Función de probabilidad para particiones intercambiables (EPPF)	46
2.5. Estimación de mezclas	50
3. Medidas de probabilidad aleatorias vía las CRM más utilizadas	54
3.1. Proceso Dirichlet	55
3.2. Proceso σ -estable	64
3.3. Proceso gama generalizado	72
3.4. Proceso Poisson-Dirichlet (σ, θ)	80
4. Medida completamente aleatoria de Bessel	86
Conclusiones	97
Apéndice A. Teorema de representación	99
Apéndice B.	104
Bibliografía	107

Capítulo 1

Preliminares

1.1. Estadística bayesiana no paramétrica

La estadística es la ciencia encargada del estudio de la incertidumbre. A grandes rasgos, se puede pensar que dentro de la estadística existen dos claras ideologías o paradigmas sobre cómo se debe hacer un análisis estadístico: la visión clásica o también llamada frecuentista y la visión bayesiana.

Para poder cuantificar la incertidumbre, se debe hacer inferencia sobre un parámetro, θ (no necesariamente univariado), dada una colección de datos, $x = (x_1, \dots, x_n)$. A grosso modo el enfoque clásico toma a θ como desconocido pero fijo y para hacer inferencia sobre él, se hace uso de la función de verosimilitud, lo cual provoca que los datos se consideren aleatorios, a pesar de haber sido ya observados. Por el contrario, bajo el paradigma bayesiano se toma a θ como aleatorio, por lo cual, al ser un objeto aleatorio tendrá una ley de probabilidad asociada (antes de observar los datos), que reflejará el conocimiento previo que se tenga sobre el fenómeno aleatorio. Para poder hacer la inferencia se debe actualizar este conocimiento a partir de la información proporcionada por los datos, lo cual se verá reflejado en la ley posterior de θ .

De esta manera, el modelo bayesiano consta de cuatro pasos claves:

1. Obtención de la distribución inicial de θ , $\pi(\theta)$.
2. Calcular la función de verosimilitud de los datos dado θ , $f(x|\theta)$.
3. Usando el teorema de Bayes, actualizar la información inicial y obtener la distribución posterior, i.e.,

$$f(\theta|x) = \frac{f(x|\theta)\pi(\theta)}{\int_{\Theta} f(x|\theta)\pi(\theta)d\theta}.$$

4. A partir de la distribución posterior hacer inferencia para θ . Y si se quieren hacer predicciones para futuras observaciones de los datos, se puede obtener la distribución predictiva, i.e.,

$$f(y|x) = \int_{\Theta} f(y|\theta, x)f(\theta|x)d\theta.$$

Uno de los argumentos que se da en contra de la visión clásica, es que presenta muchas inconsistencias, que pueden ser bien resueltas usando las herramientas del marco bayesiano. Sin embargo, la estadística bayesiana, a pesar de estar ligada al teorema de Bayes de 1763, tuvo su auge hasta mediados del siglo pasado. Este auge está ligado a los avances que hubo en la computación y los software matemáticos, que hicieron asequibles cálculos que salen con frecuencia al hacer un análisis estadístico bajo el paradigma bayesiano y que sólo son tratables de forma numérica.

La estadística bayesiana ha proporcionado muchas herramientas para la resolución de problemas estadísticos y es considerada por muchos como superior. Sin embargo, no fue suficiente para la resolución de problemas no paramétricos. Como bien lo menciona Ferguson (1973), el principal problema radica en la dificultad de encontrar distribuciones iniciales en el espacio parametral, que bajo el marco no paramétrico es el conjunto de medidas de probabilidad sobre algún espacio dotado de alguna topología, que en la mayoría de los casos será un objeto infinito dimensional. Esto último provoca que la complejidad matemática sea mucho mayor que en el caso paramétrico.

A pesar de la complejidad inherente a este enfoque, la estadística bayesiana no paramétrica se ha visto favorecida por los avances computacionales, por lo cual ya no es vista como únicamente teórica. Las aplicaciones son muchas y muy variadas y abarcan áreas como las finanzas, biología, ciencias de la computación, entre otras.

1.2. Intercambiabilidad

Habiendo mencionado en general algunas de las bondades de la estadística bayesiana, se tiene que hacer énfasis en los conceptos claves que la definen. Bajo el contexto clásico de la estadística se asume que los datos forman una muestra aleatoria, i.e., que las variables aleatorias que modelan estos datos son independientes e idénticamente distribuidas, por lo que no hay ninguna dependencia entre ellas. Sin embargo, se debe tener mucho cuidado con este supuesto, ya que asumir que las observaciones no tienen una dependencia física no implica que las variables aleatorias sean estocásticamente independientes. El aprendizaje estadístico demanda dependencia estocástica entre las variables aleatorias al ser réplicas del mismo fenómeno aleatorio. De esta manera, bajo el contexto bayesiano, asumiremos que sí existe una dependencia entre los datos, los cuales se asumen como intercambiables.

Definición 1.1. *Intercambiabilidad finita.*

Se dice que una colección finita de variables aleatorias $\{X_i\}_{i=1}^n$, definidas en el mismo espacio de probabilidad $(\Omega, \mathcal{F}, \mathbb{P})$, son intercambiables si para toda permutación $\sigma : \{1, \dots, n\} \mapsto \{1, \dots, n\}$. Se tiene

$$(X_1, \dots, X_n) \stackrel{d}{=} (X_{\sigma(1)}, \dots, X_{\sigma(n)}).$$

La intercambiabilidad establece que sólo lo que se ha observado es de interés, más no el orden con el que se observó. A partir de la definición se puede hacer notar que cualquier sucesión de variables independientes será intercambiable, más el reverso no tiene porque ser cierto.

El concepto de intercambiabilidad se extiende de una forma bastante natural cuando consideramos una sucesión infinita de variables aleatorias.

Definición 1.2. *Intercambiabilidad infinita.*

Se dice que una sucesión de variables aleatorias $\{X_i\}_{i=1}^{\infty}$, definidas en un mismo espacio de probabilidad $(\Omega, \mathcal{F}, \mathbb{P})$, es intercambiable si toda subcolección finita es finitamente intercambiable. Esto es, si para toda $n \in \mathbb{N}$ y toda sucesión de índices $\{i_1, \dots, i_n\}$ la colección de variables aleatorias $\{X_{i_k}\}_{k=1}^n$ es finitamente intercambiable.

La intercambiabilidad proporciona una noción de simetría en los datos y fue estudiada por primera vez de manera sistemática por el matemático italiano Bruno de Finetti en 1931. Del trabajo de de Finetti se desprende el teorema de representación para variables aleatorias binarias, enunciado a continuación.

Teorema 1.3. *Teorema de representación para variables binarias.*

Se dice que una secuencia de variables aleatorias binarias $\{X_i\}_{i=1}^{\infty}$ es intercambiable si y sólo si existe una distribución Q sobre el intervalo unitario tal que

$$\mathbb{P}(X_1 = x_1, \dots, X_n = x_n) = \int_0^1 \theta^{y_n} (1 - \theta)^{n - y_n} Q(d\theta).$$

Donde

$$Q(\theta) = \lim_{n \rightarrow \infty} \mathbb{P}\left(\frac{1}{n} \sum_{i=1}^n x_i \leq \theta\right), \quad Y_n = \sum_{i=1}^n X_i \quad y \quad y_n = \sum_{i=1}^n x_i.$$

Demostración. Por intercambiabilidad, para $0 \leq y_n \leq n$, se tiene que

$$\mathbb{P}(Y_n = y_n) = \binom{n}{y_n} \mathbb{P}(x_1, \dots, x_n) = \binom{n}{y_n} \mathbb{P}(x_{\sigma(1)}, \dots, x_{\sigma(n)}). \quad (1.1)$$

Con σ una permutación de $\{1, \dots, n\}$. Luego, de nueva cuenta por intercambiabilidad, si se toma N tal que $0 \leq y_n \leq n \leq N$, se sigue que:

$$\mathbb{P}(Y_n = y_n) = \sum \mathbb{P}(Y_n = y_n | Y_N = y_N) \mathbb{P}(Y_N = y_N).$$

Donde la suma se hace sobre el conjunto $\{y_n, \dots, N - (n - y_n)\}$. Se puede mostrar que la probabilidad condicional de $Y_n | Y_N = y_N$ sigue una distribución hipergeométrica, por lo que

$$\mathbb{P}(Y_n = y_n | Y_N = y_N) = \frac{\binom{y_N}{y_n} \binom{N - y_N}{n - y_n}}{\binom{N}{n}} = \frac{(y_N)_{y_n} (N - y_N)_{n - y_n}}{(N)_n},$$

donde $(x)_r := (x)(x - 1) \cdots (x - r + 1)$ denota el factorial decreciente. Reemplazando la expresión obtenida para la probabilidad condicional, se obtiene lo siguiente:

$$\mathbb{P}(Y_n = y_n) = \sum \frac{(y_N)_{y_n} (N - y_N)_{n - y_n}}{(N)_n} \mathbb{P}(Y_N = y_N).$$

Definiendo a $Q_N(\theta)$ como la función que tiene saltos de tamaño $\mathbb{P}(Y_N = y_N)$ en $\theta = \frac{y_N}{N}$ y reescribiendo la suma en términos de la integral de Lebesgue,

$$\mathbb{P}(Y_n = y_n) = \binom{n}{y_n} \int_0^1 \frac{(\theta N)_{y_n} ((1 - \theta)N)_{n - y_n}}{(N)_n} Q_N(d\theta). \quad (1.2)$$

El caso que interesa es cuando $N \rightarrow \infty$, para esto se utiliza el hecho de que para r fija se tiene que $\lim_{x \rightarrow \infty} (x)_r = x^r$. De esta forma

$$\frac{(\theta N)_{y_n} ((1 - \theta)N)_{n - y_n}}{(N)_n} \rightarrow \frac{(\theta N)^{y_n} ((1 - \theta)N)^{n - y_n}}{(N)^n} = \theta^{y_n} (1 - \theta)^{n - y_n}.$$

Finalmente, $Q_N(d\theta)$ es una función escalonada que cumple con las hipótesis del teorema de Helly, que garantiza la existencia de una subsucesión $\{Q_{N_i}(d\theta)\}$, tal que

$$\lim_{i \rightarrow \infty} Q_{N_i}(d\theta) \rightarrow Q(d\theta).$$

Con lo que, comparando la expresión (1.1) con la expresión (1.2) y tomando el límite cuando N tiende a infinito, se llega al resultado deseado. \square

Este resultado es uno de los pilares de la estadística bayesiana, ya que relaciona el concepto de intercambiabilidad con el concepto de independencia condicional. Esto último es más fácil de ver si se reescribe la integral como sigue,

$$\int_0^1 \theta^{\sum_{i=1}^n x_i} (1 - \theta)^{n - \sum_{i=1}^n x_i} Q(d\theta) = \int_0^1 \prod_{i=1}^n \theta^{x_i} (1 - \theta)^{1 - x_i} Q(d\theta).$$

Por lo que pensándolo en un sentido paramétrico, se está garantizando la existencia de un parámetro θ tal que,

$$\mathbb{P}(X_1 = x_1, \dots, X_n = x_n | \theta) = \prod_{i=1}^n \theta^{x_i} (1 - \theta)^{1 - x_i}.$$

El resultado de de Finetti fue extendido por Hewitt y Savage en 1955, dando lugar a lo que ahora se conoce como el teorema de representación. Para esto se debe considerar una colección de variables aleatorias $X^{(\infty)} = \{X_i\}_{i=1}^{\infty}$ definidas en un espacio de probabilidad $(\Omega, \mathcal{F}, \mathbb{P})$. Donde cada una de ellas toma valores en un espacio polaco \mathbb{X} dotado con su σ -álgebra de Borel \mathcal{X} . Finalmente, se debe de considerar a $\mathcal{P}_{\mathbb{X}}$ que denota al espacio de todas las medidas de probabilidad sobre $(\mathbb{X}, \mathcal{X})$.

Teorema 1.4. *Teorema de representación.*

La colección $X^{(\infty)}$ es intercambiable si y sólo si existe una medida \mathcal{Q} en $\mathcal{P}_{\mathbb{X}}$, tal que para toda $n \geq 1$ y $A_1, A_2, \dots, A_n \in \mathcal{X}$ se cumple

$$\mathbb{P}(X_1 \in A_1, \dots, X_n \in A_n) = \int_{\mathcal{P}_{\mathbb{X}}} \prod_{i=1}^n P(A_i) \mathcal{Q}(dP).$$

La demostración del teorema de representación puede resultar algo extensa, por lo cual se incluye en el Apéndice A. Este resultado puede ser expresado en una manera elegante y sencilla a través de un modelo jerárquico,

$$X_i|P \stackrel{iid}{\sim} P$$

$$P \sim Q.$$

El cual nos regresa a la idea de independencia condicional y de idéntica distribución de la secuencia de variables aleatorias. Los principales elementos que componen este teorema son P , que es una medida de probabilidad aleatoria y Q que se le conoce como la medida de de Finetti. La medida Q llega a tener la interpretación de ser la distribución inicial para P ; dependiendo de su soporte, se tendrá un problema paramétrico (si es finito-dimensional) o un problema no paramétrico (si es infinito-dimensional).

La estadística bayesiana no paramétrica está centrada en la construcción de distribuciones iniciales para P . Una forma de hacerlo es a través de la especificación de la ley que la rige, i.e., Q . La construcción también se puede hacer de forma directa. Existen varios métodos; sin embargo, como ya se mencionó antes este trabajo se centra en la construcción que está basada en la normalización de medidas completamente aleatorias obteniendo lo que se conoce en la literatura, como una NRMI, que denota por sus siglas en inglés una medida aleatoria normalizada con incrementos independientes.

Antes de proceder con el estudio de medidas completamente aleatorias, se puede hacer notar que cuando tomamos $\mathbb{X} = \mathbb{R}$, se estará haciendo énfasis en lo que se conoce como procesos aditivos o procesos con incrementos independientes y en algunas ocasiones se tendrá el caso en el que los incrementos sean estacionarios. Por lo que, a manera de introducción, en la siguiente sección se definen y caracterizan a los procesos de Lévy con respecto a lo que se conoce como variables aleatorias infinitamente divisibles y la representación de Lévy-Khintchine.

1.3. Procesos de Lévy

Los procesos de Lévy, que deben su nombre al matemático francés Paul Lévy, son una clase de procesos estocásticos caracterizados por tener incrementos independientes y estacionarios. Algunos de los ejemplos más conocidos son el proceso Poisson y el movimiento browniano con deriva, siendo este último el único proceso de Lévy con trayectorias continuas.

Los procesos de Lévy están muy ligados a la noción de variables aleatorias infinitamente divisibles. Esto último será de gran utilidad al momento de trabajar con la función característica y la transformada de Laplace de los procesos. El concepto de variables aleatorias infinitamente divisibles fue introducido por Bruno de Finetti (1929) para procesos con incrementos independientes y estacionarios.

Definición 1.5. *Variable aleatoria n -divisible.*

Se dice que la variable aleatoria X es n -divisible si para toda $n \geq 2$, con $n \in \mathbb{N}$, existen variables aleatorias independientes e idénticamente distribuidas Y_1, \dots, Y_n tales que

$$X \stackrel{d}{=} Y_1 + \dots + Y_n.$$

La propiedad de ser n -divisible puede tener muchas ventajas en el modelado de muchos fenómenos a estudiar; sin embargo, es un concepto que puede ser mejorado, en el sentido de que si una variable aleatoria X es $n + 1$ -divisible no implica necesariamente que X sea n -divisible. De esta forma surge el concepto de divisibilidad para toda n y en el que se hará mayor enfoque.

Definición 1.6. *Variable aleatoria infinitamente divisible.*

Se dice que una variable aleatoria X es infinitamente divisible si para toda $n \in \mathbb{N}$

$$X \stackrel{d}{=} X_{n,1} + \dots + X_{n,n}.$$

Donde las variables aleatorias $\{X_{n,i}\}_{i=1}^n$ son independientes y $X_{n,i} \stackrel{d}{=} X_n$ para toda i y alguna variable X_n .

La definición anterior puede ser expresada en términos de la función de distribución y de la función característica. De esta forma se dice que F es una distribución infinitamente divisible si para toda n , F es la n -ésima convolución de una distribución F_n consigo misma y ϕ una función característica es infinitamente divisible si para toda n , ϕ es la n -ésima potencia de alguna función característica ϕ_n . Esto es,

$$F = F_n^{*n} \quad \text{y} \quad \phi(t) = \phi_n(t)^n.$$

Se puede caracterizar a las variables aleatorias infinitamente divisibles a través de su exponente característico, que está definido como $\varphi(u) := -\log \mathbb{E}(e^{iuX})$ a través del siguiente teorema.

Teorema 1.7. *Fórmula de Lévy-Khintchine.*

Sea X una variable aleatoria con función de distribución $F_X(\cdot)$ y con exponente característico $\varphi(\cdot)$, donde

$$\int_{\mathbb{R}} e^{iux} F(dx) = e^{-\varphi(u)}.$$

Se dice que $F_X(\cdot)$ es infinitamente divisible si existen $a \in \mathbb{R}$, $\sigma \geq 0$ y ν una medida concentrada en $\mathbb{R} \setminus \{0\}$ tal que,

$$\varphi(u) = iau + \frac{1}{2}\sigma^2 u^2 + \int_{\mathbb{R}} (1 - e^{iux} + iux \mathbb{1}_{(|x|<1)}) \nu(dx).$$

Donde ν es tal que $\int_{\mathbb{R}} (1 \wedge x^2) \nu(dx) < \infty$ y se le conoce como la medida de Lévy. Y a la terna (a, σ, ν) se le conoce como la terna generadora.

La demostración de este teorema puede ser consultada en Sato (1990), donde además se pueden encontrar más resultados sobre el tema en cuestión. Ahora que ya se han caracterizado a las variables infinitamente divisible se continuará con los procesos de Lévy.

Definición 1.8. *Proceso de Lévy.*

Un proceso estocástico $\{X_t; t \geq 0\}$ definido en un espacio de probabilidad $(\Omega, \mathcal{F}, \mathbb{P})$ es un proceso de Lévy si cumple lo siguiente:

1. $X_0 = 0$ c.s.
2. Incrementos independientes, i.e., para $0 \leq s \leq t$, se tiene $X_t - X_s \perp \{X_u : u \leq s\}$.
3. Incrementos estacionarios, i.e., para $s, t > 0$, se tiene $X_{t+s} - X_t \stackrel{d}{=} X_s$.
4. Tiene trayectorias càdlàg.

Si $\{X_t; t \geq 0\}$ cumple todas las propiedades, excepto la de tener incrementos estacionarios, se dice que $\{X_t; t \geq 0\}$ es un proceso aditivo. Una clase especial de procesos de Lévy son los conocidos como subordinadores, que son aquellos procesos de Lévy con trayectorias no decrecientes. A partir de la definición se puede establecer la conexión con las variables infinitamente divisibles.

Proposición 1.9. *Sea $\{X_t; t \geq 0\}$ un proceso de Lévy, entonces para toda $t > 0$, X_t es una variable aleatoria perteneciente a la clase de variables infinitamente divisibles. Más aún, si se define para $u \in \mathbb{R}$ a*

$$\varphi_t(u) = -\log(\mathbb{E}(e^{iuX_t}))$$

entonces

$$\varphi_t(u) = t\varphi_1(u).$$

Demostración. Dada una $t > 0$, fija, se define a $\tau_k := \frac{kt}{n}$ para toda $n = 1, 2, \dots$, de esta manera

$$X_t = X_{\tau_1} + (X_{\tau_2} - X_{\tau_1}) + \dots + (X_{\tau_n} - X_{\tau_{n-1}}).$$

Utilizando las propiedades de incrementos independientes y estacionarios del proceso, se sigue que se tiene la suma de n variables independientes e idénticamente distribuidas, de acuerdo a la ley de X_{τ_1} . Por lo que efectivamente está en la clase de variables aleatorias infinitamente divisibles.

Para la segunda parte de la proposición basta notar que para enteros m, n se puede escribir a X_m de las siguientes maneras

$$X_m = X_1 + (X_2 - X_1) + \cdots + (X_m - X_{m-1})$$

$$X_m = X_{\frac{m}{n}} + \left(X_{\frac{2m}{n}} - X_{\frac{m}{n}}\right) + \cdots + \left(X_m - X_{\frac{(n-1)m}{n}}\right).$$

De donde se sigue que $m\varphi_1(u) = \varphi_m(u) = n\varphi_{\frac{m}{n}}(u)$. Por lo que la igualdad vale para números racionales. Si ahora se elige a t irracional, basta tomar una sucesión de racionales t_n tales que $t_n \downarrow t$ y usando la continuidad de $\exp(\cdot)$ y el teorema de convergencia dominada, se obtiene lo siguiente,

$$e^{-\varphi_t(u)} = \mathbb{E} \left[\lim_{n \rightarrow \infty} e^{iuX_{t_n}} \right] = \lim_{n \rightarrow \infty} \mathbb{E}[e^{iuX_{t_n}}] = \lim_{n \rightarrow \infty} e^{-\varphi_{t_n}(u)} = \lim_{n \rightarrow \infty} e^{-t_n\varphi_1(u)} = e^{-t\varphi_1(u)}$$

por lo que la igualdad se cumple para todo $t > 0$. □

Este resultado permite garantizar que la distribución de un proceso de Lévy está completamente determinada por X_1 , que es una variable infinitamente divisible. Siendo una especie de converso, el hecho de que dada una variable aleatoria infinitamente divisible X , entonces existe un proceso de Lévy $\{X_t; t \geq 0\}$ con $X_1 \stackrel{d}{=} X$. Esto último se puede demostrar construyendo las familias finito dimensionales y viendo que se satisfacen las condiciones del teorema de consistencia de Kolmogorov.

Utilizando los resultados anteriores y el teorema de representación de Lévy Khintchine, se puede garantizar que dado un exponente característico $\varphi(u)$, existe un espacio de probabilidad $(\Omega, \mathcal{F}, \mathbb{P})$ en el cuál está definido un proceso de Lévy $\{X_t; t \geq 0\}$ que tendrá ese exponente característico. De esta manera si recordamos que

$$\varphi(u) = iau + \frac{1}{2}\sigma^2 u^2 + \int_{\mathbb{R}} (1 - e^{iux} + iux \mathbb{1}_{(|x|<1)}) \nu(dx).$$

Se puede ver que el proceso está completamente caracterizado por la terna (a, σ, ν) . La medida de Lévy ν será de vital importancia ya que caracteriza los saltos y la frecuencia de los saltos; si esta medida es infinita, se dice que el proceso tiene actividad infinita.

En algunas ocasiones, los procesos a considerar pertenecen a los ya mencionados subordinadores. Se mencionó una característica esencial; sin embargo, podemos dar una definición más completa.

Definición 1.10. *Un proceso de Lévy es un subordinador si y sólo si $\nu(-\infty, 0) = 0$, $\int_{\mathbb{R}_+} (1 \wedge x) \nu(dx) < \infty$, $\sigma \equiv 0$ y $- \left(a + \int_0^1 x \nu(dx) \right) \geq 0$.*

Las condiciones anteriores permiten asegurar que el proceso no tendrá saltos negativos y que tendrá trayectorias no decrecientes, un análisis más extenso puede hacerse a través de la descomposición de Lévy-Itô, una referencia a lo anterior se puede encontrar en Kyprianou (2006).

Asumiendo las características anteriores, el exponente característico tendrá la siguiente representación para $u > 0$,

$$\varphi(u) = iau + \int_{\mathbb{R}_+} (1 - e^{-iux} + iux \mathbb{1}_{(|x|<1)}) \nu(dx).$$

En algunas ocasiones esta expresión se suele simplificar como sigue.

$$\varphi(u) = iua' + \int_{\mathbb{R}_+} (1 - e^{-iux})\nu(dx). \quad (1.3)$$

Donde

$$a' = a - \int_0^1 \nu(dx).$$

Más aún, en este caso como se está trabajando en \mathbb{R}_+ , se puede trabajar con la transformada de Laplace en lugar de la función característica, tomando $u = -iu$, de esta forma el exponente laplaciano está dado por

$$\varphi(u) = ua + \int_{\mathbb{R}_+} (1 - e^{-ux})\nu(dx) \quad (1.4)$$

Para ejemplificar todo lo anterior y siguiendo las últimas notaciones mencionadas, se presenta a continuación un proceso clave en el desarrollo de la estadística bayesiana no paramétrica y que será retomado más adelante.

Ejemplo 1.11. *Proceso gama.*

Sea $\{X_t; t \geq 0\}$ un proceso de Lévy con terna característica $(0, 0, \nu)$, donde ν está concentrada en $(0, \infty)$ y es tal que

$$\nu(dx) = \frac{e^{-x}}{x} dx.$$

De esta manera, es sencillo verificar que cumple lo necesario para ser un subordinador y además

$$\int_{\mathbb{R}_+} (1 \wedge x^2)\nu(dx) = \int_0^1 xe^{-x} dx + \int_1^\infty \frac{e^{-x}}{x} dx < \infty.$$

Y cuyo exponente característico está dado por

$$\begin{aligned} \varphi(u) &= \int_{\mathbb{R}_+} (1 - e^{iux})\nu(dx) \\ &= \log(1 - iu). \end{aligned}$$

Y de la misma forma su exponente laplaciano está dado por

$$\begin{aligned}\varphi(u) &= \int_{\mathbb{R}_+} (1 - e^{-ux})\nu(dx) \\ &= \log(1 + u).\end{aligned}$$

Más aún $\nu(\mathbb{R}_+) = \infty$, por lo que es una medida con actividad infinita.

El proceso gama fue utilizado por Ferguson (1973) para la construcción del proceso Dirichlet vía la normalización de este subordinador. De ahí su importancia para el desarrollo de la estadística bayesiana no paramétrica.

1.4. Medidas completamente aleatorias (CRM)

Una vez que se ha desarrollado un poco la teoría concerniente a los proceso de Lévy y que sirve como una introducción sencilla al problema de interés, ya que se restringe al caso donde $\mathbb{X} = \mathbb{R}$, podemos proceder al siguiente paso, que es la definición de medidas completamente aleatorias. Este concepto fue introducido por Kingman (1967). Sin embargo, primero se hará mención a los procesos conocidos como de Poisson en espacios generales. Mucha de la notación y teoría que se presenta a continuación, está basada en el libro de Kingman (1993) y en la tesis de maestría de Gil Leyva (2016).

Definición 1.12. *Proceso Poisson.*

Dado un espacio medible (S, \mathcal{S}) , se dice que un subconjunto aleatorio y contable de S , denotado como Π es un proceso Poisson en S si

1. *Para cualesquiera A_1, \dots, A_n subconjuntos medibles y disjuntos de S , las variables aleatorias $N(A_1), \dots, N(A_n)$ son independientes. Donde $N(A) = \#(\Pi \cap A)$.*
2. *$N(A)$ se distribuye como una variable Poisson($\mu(A)$), donde $0 \leq \mu(A) \leq \infty$.*

No es difícil verificar que $\mu(\cdot)$ es una medida en S y se le conoce como la medida media del proceso Poisson. Es importante notar que no cualquier medida puede ser una medida media, se tiene que garantizar que μ sea no atómica, i.e.,

$$\mu(\{x\}) = 0 \quad \forall x \in S.$$

A continuación se enuncian las principales propiedades de este tipo de procesos y cuyas demostraciones se omiten, pero que pueden ser consultadas en Kingman (1993). A partir de este momento y hasta indicar lo contrario estaremos trabajando en un espacio medible (S, \mathcal{S}) donde estará definido el proceso Poisson.

Teorema 1.13. *Teorema de superposición.*

Sean Π_1, Π_2, \dots una colección de procesos Poisson independientes en S , donde $\forall n$ Π_n tiene medida media μ_n . Entonces

$$\Pi = \bigcup_{n=1}^{\infty} \Pi_n$$

es un proceso Poisson con medida media

$$\mu = \sum_{n=1}^{\infty} \mu_n.$$

Teorema 1.14. *Teorema de restricción.*

Sea Π un proceso Poisson en S con medida media μ y sea S_1 subconjunto medible de S . Entonces

$$\Pi_1 = \Pi \cap S_1$$

es un proceso Poisson en S , con medida media

$$\mu_1(A) = \mu(A \cap S).$$

Equivalentemente, Π_1 es un proceso Poisson en S_1 con medida media $\mu|_{S_1}$.

Teorema 1.15. *Teorema del mapeo.*

Sean (S, \mathcal{S}) y (T, \mathcal{T}) dos espacios medibles, Π un proceso Poisson en S , con medida media μ , tal que μ es σ -finita y sea $f : S \mapsto T$ un mapeo medible tal que la medida inducida $\mu^*(\cdot) = \mu(f^{-1}(\cdot))$ es no atómica. Entonces $f(\Pi)$ es un proceso Poisson en T con medida media μ^* .

Una vez establecidas algunas de las propiedades, se puede proceder al teorema de Campbell.

Teorema 1.16. *Teorema de Campbell.*

Sea Π un proceso Poisson en S , con medida media μ y sea $f : S \mapsto \mathbb{R}$ medible. Entonces

$$\Sigma = \sum_{X \in \Pi} f(X)$$

es absolutamente convergente en probabilidad si y sólo si

$$\int_S (|f(x)| \wedge 1) \mu(dx) < \infty.$$

Más aún,

$$\mathbb{E}(e^{u\Sigma}) = \exp \left(\int_S (e^{uf(x)} - 1) \mu(dx) \right).$$

Para $u \in \mathbb{C}$ tal que la integral definida en la exponencial sea convergente.

Demostración. El primer paso consiste en tomar a f una función simple, i.e., existen A_1, \dots, A_n subconjuntos medibles y disjuntos (siempre podemos garantizar esto) y constantes a_1, \dots, a_n positivas distintas entre si tales que

$$f(x) = \sum_{i=1}^n a_i \mathbb{1}_{x \in A_i}.$$

De esta manera también se garantiza que las variables aleatorias $N(A_1), \dots, N(A_n)$ sean mutuamente independientes y con distribución Poisson($\mu(A_i)$) $\forall i \in \{1, \dots, n\}$. De esta forma,

$$\begin{aligned}
\mathbb{E}(e^{u\Sigma}) &= \mathbb{E} \left[\exp \left(u \sum_{X \in \Pi} \sum_{i=1}^n a_i \mathbb{1}_{X \in A_i} \right) \right] \\
&= \mathbb{E} \left[\exp \sum_{i=1}^n (u a_i N(A_i)) \right] \\
&= \prod_{i=1}^n \mathbb{E} [e^{u a_i N(A_i)}] \\
&= \prod_{i=1}^n \exp[\mu(A_i)(e^{u a_i} - 1)] \\
&= \exp \left[\sum_{i=1}^n \int_S (e^{u a_i} - 1) \mathbb{1}_{A_i} \mu(dx) \right] \\
&= \exp \left[\int_S (e^{u \sum_{i=1}^n a_i \mathbb{1}_{A_i}} - 1) \mu(dx) \right] \\
&= \exp \left[\int_S (e^{u f(x)} - 1) \mu(dx) \right].
\end{aligned}$$

Por lo que el teorema es válido para funciones simples. Sin embargo, como bien lo hace notar Kingman, hay varios puntos importantes a considerar.

1. Si f puede tomar valores tanto positivos como negativos, entonces es mejor que la parte real de u sea igual a 0 y obtener la función característica de Σ .
2. Si f sólo toma valores positivos es mejor que $u \geq 0$ para trabajar con $-u$ y así tener una especie de transformada de Laplace.

De esta manera si f es una función no negativa, se puede ser expresar como el límite de una sucesión creciente de funciones simples no negativas. Por lo que se toma a $u \geq 0$ y utilizando el teorema de convergencia monótona se obtiene que,

$$\begin{aligned}\mathbb{E}(e^{-u\Sigma}) &= \lim_{n \rightarrow \infty} \mathbb{E}(e^{-u \sum_{X \in \Pi} f_n(X)}) \\ &= \lim_{n \rightarrow \infty} \exp \left[- \int_S (1 - e^{-u f_n(x)}) \mu(dx) \right] \\ &= \exp \left[- \int_S (1 - e^{-u f(x)}) \mu(dx) \right].\end{aligned}$$

Más aún, la condición de integrabilidad para f garantiza que si $u \rightarrow 0$ entonces

$$\int_S (1 - e^{-u f(x)}) \mu(dx) \rightarrow 0.$$

Por lo que Σ es una variable finita casi seguramente. Y si ahora la integral diverge entonces $\mathbb{E}(e^{-u\Sigma}) = 0$, por lo que $\Sigma = \infty$ c.s. Finalmente, si f es cualquier función medible, se sabe que se puede expresar como $f = f^+ - f^-$, donde f^+ y f^- son funciones medibles no negativas. Así Σ converge si y sólo si

$$\Sigma_+ = \sum_{X \in \Pi} f^+(X) \quad \text{y} \quad \Sigma_- = \sum_{X \in \Pi} f^-(X)$$

convergen. Donde Σ_+ y Σ_- son procesos Poisson independientes ya que están restringidos a los conjuntos $\{x : f(x) > 0\}$ y $\{x : f(x) < 0\}$ respectivamente. De esta forma,

$$\begin{aligned}\mathbb{E}(e^{-u\Sigma}) &= \mathbb{E}(e^{-u\Sigma_+ + u\Sigma_-}) \\ &= \mathbb{E}(e^{-u\Sigma_+}) \mathbb{E}(e^{u\Sigma_-}) \\ &= \exp \left[- \int_{\{x: f(x) > 0\}} (1 - e^{-u f^+(x)}) \mu(dx) - \int_{\{x: f(x) < 0\}} (1 - e^{u f^-(x)}) \mu(dx) \right] \\ &= \exp \left[- \int_{\{x: f(x) > 0\}} (1 - e^{-u f(x)}) \mu(dx) - \int_{\{x: f(x) < 0\}} (1 - e^{-u f(x)}) \mu(dx) \right] \\ &= \exp \left[- \int_S (1 - e^{-u f(x)}) \mu(dx) \right].\end{aligned}$$

Con lo cual es válido para $u \geq 0$, por lo que debe valer para $u \in \mathbb{R}$ y mediante una extensión analítica se verifica que es válido para $u \in \mathbb{C}$. \square

Si en el teorema de Campbell se fija $f \geq 0$ y $u = 1$ se obtiene la funcional característica del proceso aleatorio

$$\mathbb{E}(e^{-\Sigma}) = \exp\left(-\int (1 - e^{-f(x)})\mu(dx)\right) \quad (1.5)$$

que caracteriza por completo al proceso Poisson. Antes de seguir con el análisis e importancia del teorema de Campbell se debe definir lo que es una medida aleatoria y una medida completamente aleatoria, como en el trabajo de Kingman (1967).

Definición 1.17. *Medida aleatoria y medida completamente aleatoria.*

Dado (S, \mathcal{S}) un espacio medible una medida aleatoria Φ es una aplicación $\Phi : \Omega \times \mathcal{S} \rightarrow [0, \infty)$, tal que

1. Para todo $\omega \in \Omega$, el mapeo $A \mapsto \Phi(\omega, A)$ es una medida en \mathcal{S} .
2. Para todo $A \in \mathcal{S}$, el mapeo $\omega \mapsto \Phi(\omega, A)$ es una variable aleatoria.

Más aún, se dice que Φ es una medida completamente aleatoria si para toda familia finita y disjunta A_1, \dots, A_n las variables aleatorias $\Phi(\omega, A_1), \dots, \Phi(\omega, A_n)$ son independientes.

Uno de los resultados más importantes para la teoría concerniente a las medidas completamente aleatorias es el teorema de representación de Kingman. Este teorema establece que una medida completamente aleatoria se puede representar como la suma de tres componentes, uno determinista, otro concentrado en un conjunto de átomos fijos y uno último concentrado en un conjunto de átomos aleatorios; por lo que, salvo una componente determinista, una medida completamente aleatoria es puramente atómica.

Antes de enunciar el teorema de representación de Kingman se debe continuar con el estudio de los procesos Poisson. Para esto, primero se puede ver que si se tiene a Π un proceso Poisson en S con medida media μ y se define a la función de conteo N como

$$N(A) = \#(\Pi \cap A).$$

Entonces N cumple lo siguiente,

1. $N(\emptyset) = \#(\Pi \cap \emptyset) = \#(\emptyset) = 0$.
2. Si se tiene una colección disjunta A_1, A_2, \dots , en S , entonces

$$N\left(\bigcup_{i=1}^{\infty} A_i\right) = \#\left(\Pi \cap \bigcup_{i=1}^{\infty} A_i\right) = \#\left(\bigcup_{i=1}^{\infty} (\Pi \cap A_i)\right) = \sum_{i=1}^{\infty} \#(\Pi \cap A_i) = \sum_{i=1}^{\infty} N(A_i).$$

Por lo que $N(\cdot)$ es una medida en S , de esta manera y dado que las demás condiciones las cumple trivialmente se dice que $N(\cdot)$ es una medida completamente aleatoria en S y se le conoce como la medida aleatoria de Poisson. La existencia de estas medidas se puede probar y para verlo uno puede referirse a Kingman (1993) o a Uribe (2013).

Lo que se busca es dar una representación en el espacio producto. La idea es que para cada punto $X \in \Pi$ se tiene asociada una variable aleatoria m_X , que toma valores en algún espacio M y donde se asume que la distribución de m_X sólo podrá depender de X , i.e., otro punto $Y \in \Pi$ no la alterará y más aún las variables aleatorias asociadas a diferentes puntos serán independientes.

De esta manera, la pareja $X^* = (X, m_X)$ se puede pensar como un punto aleatorio en $S \times M$ y el total de puntos denotado por $\Pi^* \{(X, m_X) : X \in \Pi\}$ forma un conjunto aleatorio que como se verá a continuación será un proceso Poisson en el espacio producto. Antes de enunciar el teorema que garantiza lo anterior se da la siguiente definición.

Definición 1.18. *Proceso Poisson marcado.*

Sea Π un proceso Poisson en S con medida media μ y una función de distribución $p(x, \cdot)$ en M que depende de $x \in S$ de tal forma que para $B \subseteq M$, se tiene que $p(\cdot, B)$ es una función medible en S . Se dice que una versión marcada de Π es un subconjunto aleatorio de $S \times M$ cuya proyección en S es Π y tal que la distribución condicional de Π^* dado Π hace que las variables aleatorias m_X sean independientes con distribución $p(X, \cdot)$.

Con esta definición ya se tienen los elementos necesarios para enunciar el teorema.

Teorema 1.19. *El subconjunto aleatorio Π^* definido como arriba, es un proceso Poisson en el espacio producto $S \times M$ con medida media μ^* dada por,*

$$\mu^*(A) = \iint_{(x,m) \in A} p(x, dm) \mu(dx).$$

Demostración. Para una función medible f en el espacio producto, sea

$$\Sigma^* = \sum_{X \in \Pi} f(X, m_X).$$

Por definición se sabe que dado Π , se tiene una suma de variables aleatorias independientes, de esta forma y utilizando la esperanza condicional.

$$\begin{aligned} \mathbb{E}(e^{-\Sigma^*} | \Pi) &= \mathbb{E} \left[\exp \left(- \sum_{X \in \Pi} f(X, m_X) \right) \middle| \Pi \right] \\ &= \prod_{X \in \Pi} \mathbb{E}[\exp(-f(X, m_X)) | \Pi] \\ &= \prod_{X \in \Pi} \int_M e^{-f(X,m)} p(X, dm). \end{aligned}$$

De esta manera,

$$\begin{aligned}
\mathbb{E}(e^{-\Sigma^*}) &= \mathbb{E}(\mathbb{E}(e^{-\Sigma^*}|\Pi)) \\
&= \mathbb{E}\left(\prod_{X \in \Pi} \int_M e^{-f(X,m)} p(X, dm)\right) \\
&= \mathbb{E}\left[\exp\left(-\sum_{X \in \Pi} -\log\left(\int_M e^{-f(X,m)} p(X, m)\right)\right)\right]
\end{aligned}$$

Aplicando el teorema de Campbell al proceso Poisson Π e intercambiando a f por g , que está definida como

$$g := -\log\left(\int_M e^{-f(X,m)} p(X, m)\right).$$

Se tiene que,

$$\begin{aligned}
\mathbb{E}(e^{-\Sigma^*}) &= \exp\left(-\int_S (1 - e^{-g(x)})\mu(dx)\right) \\
&= \exp\left[-\int_S \left(\int_M p(x, dm) - \int_M e^{-f(x,m)} p(x, dm)\right) \mu(dx)\right] \\
&= \exp\left[-\int_S \int_M (1 - e^{-f(x,m)}) p(x, dm) \mu(dx)\right] \\
&= \exp\left[-\int_{S \times M} (1 - e^{-f(x,m)}) \mu^*(dx, dm)\right].
\end{aligned}$$

Por lo que efectivamente Π^* es un proceso Poisson en el espacio producto y con medida media dada por μ^* . □

Una de las implicaciones inmediatas es que por el teorema del mapeo, m_X al ser una proyección es un proceso Poisson en M . De esta manera se puede aplicar el teorema de Campbell a $\sum m_X$ y de esta manera trabajar no sólo con funciones $f(X)$ que dependen de una forma determinista de X , sino de forma aleatoria e independiente de punto a punto.

De esta manera si las variables m_X son positivas, entonces su funcional característica estará dada por

$$\begin{aligned}\mathbb{E}(e^{-\Sigma m_X}) &= \exp \left[- \int_M (1 - e^{-m}) \mu_m(dm) \right] \\ &= \exp \left[- \int_S \int_M (1 - e^{-m}) \mu(dx) p(x, dm) \right].\end{aligned}$$

Ahora bien, regresando al tema de medidas completamente aleatorias, se sabe que para conjuntos disjuntos A_1, \dots, A_n se tiene que $\Phi(A_1), \dots, \Phi(A_n)$ es una colección de variables aleatorias independientes. Por lo cual su distribución conjunta está completamente especificada para cualquier familia de conjuntos (no necesariamente disjuntos), ya que siempre se puede considerar a la familia

$$\{B : B = A_1^* \cap \dots \cap A_n^*\}.$$

Donde $A_i^* = A_i$ ó $A_i^* = A_i^c$. Que es una familia disjunta y finita y que recupera a cada uno de los elementos originales A_1, \dots, A_n .

Una manera conveniente de caracterizar la distribución de $\Phi(A)$ es mediante la función

$$\lambda_u(A) = -\log[\mathbb{E}(e^{-u\Phi(A)})], \quad u > 0 \tag{1.6}$$

Para la cual se pueden agrupar sus propiedades y su clara relación con Φ en la siguiente proposición.

Proposición 1.20. *Definiendo a $\lambda_u(A)$ como en (1.4) se cumple lo siguiente.*

1. $\lambda_u(\cdot)$ es una medida en S y tal que $0 \leq \lambda_u(A) \leq \infty$.
2. λ_u y Φ son absolutamente continuas una con respecto a la otra.
3. λ_u es finita (infinita) si y sólo si Φ es finita (infinita) c.s.

Demostración. Claramente $\lambda_u(A) \in [0, \infty]$ por como está definida. Ahora bien, lo primero que se mostrará es que efectivamente es una medida. Para esto, claramente se tiene que es nula en el vacío ya que,

$$\lambda_u(\emptyset) = -\log[\mathbb{E}(e^{-u\Phi(\emptyset)})] = -\log[1] = 0.$$

Ahora, se puede ver que λ_u es contablemente aditiva. Para esto sea A_1, A_2, \dots , una colección mutuamente excluyente de tal forma que,

$$\begin{aligned} \lambda_u\left(\bigcup_{i=1}^{\infty} A_i\right) &= -\log\left[\mathbb{E}\left(e^{-u\Phi(\bigcup_{i=1}^{\infty} A_i)}\right)\right] \\ &= -\log\left[\mathbb{E}\left(e^{-u\sum_{i=1}^{\infty} \Phi(A_i)}\right)\right] \\ &= -\log\left[\mathbb{E}\left(\prod_{i=1}^{\infty} e^{-u\Phi(A_i)}\right)\right] \\ &= -\log\left[\prod_{i=1}^{\infty} \mathbb{E}(e^{-u\Phi(A_i)})\right] \\ &= \sum_{i=1}^{\infty} -\log[\mathbb{E}(e^{-u\Phi(A_i)})] \\ &= \sum_{i=1}^{\infty} \lambda_u(A_i). \end{aligned}$$

Por lo que efectivamente λ_u es una medida en S .

Ahora bien, si se considera a A tal que $\Phi(A) = 0$ se tiene que $\lambda_u(A) = 0$ (siguiendo los mismos pasos que para \emptyset). Si ahora se considera a A de tal forma que $\lambda_u(A) = 0$ se sigue que

$$0 = -\log[\mathbb{E}(e^{-u\Phi(A)})] \quad \Leftrightarrow \quad \mathbb{E}(e^{-u\Phi(A)}) = 1 \quad \Leftrightarrow \quad \Phi(A) = 0.$$

Esto último ya que $u > 0$. Por lo que λ_u y Φ son medidas absolutamente continuas una con respecto a la otra. Finalmente, si $\Phi(A)$ es infinita se sigue que $e^{-u\Phi(A)} = 0$ c.s. por lo que $\lambda_u(A) = \infty$, siendo cierto el converso también. El caso finito es idéntico. \square

Se pueden imponer restricciones a la medida λ_u , siendo una de ellas que sea σ finita, i.e., que exista una partición contable

$$S = \bigcup_{j=1}^{\infty} S_j$$

de S tal que $\Phi(S_j) < \infty$ con probabilidad positiva, cuando esto se cumpla se dice que Φ es Σ finita. Una implicación de esto, es que λ_u tendrá los mismos átomos que Φ , donde se dice que $\{x\}$ es un átomo de Φ si y sólo si

$$\mathbb{P}(\Phi(\{x\}) > 0) > 0.$$

Estos átomos estarán fijos y forman a lo más un subconjunto contable $X = \{x_1, x_2, \dots\}$ de S . De esta manera si se define a $S_0 := S \setminus X$, se puede demostrar que Φ_0 (la restricción de Φ a S_0) es una medida completamente aleatoria sin átomos fijos y que es independiente de las variables aleatorias $\{\Phi(\{x_i\})\}_{i=1}^{\infty}$.

Con el procedimiento mencionado arriba se llega a una primera descomposición para medidas completamente aleatorias Σ finitas. Definiendo a

$$\Phi_f(A) := \Phi(A \cap X), \quad \Phi_0(A) := \Phi(A \cap (S \setminus X)).$$

Que son las correspondientes restricciones, entonces se puede escribir a Φ como,

$$\Phi = \Phi_f + \Phi_0$$

donde Φ_0 es una medida completamente aleatoria sin átomos fijos y Φ_f es una medida completamente aleatoria que puede ser escrita como

$$\Phi_f = \sum_{x \in X} \phi(x) \delta_x.$$

Donde las variables aleatorias $\{\phi(x)\}_{x \in X}$ son independientes. De esta manera los puntos fijos se pueden remover, por lo que se puede asumir sin pérdida de generalidad que una medida Σ finita no tiene átomos fijos, i.e., $\Phi = \Phi_0$ y como consecuencia directa que la medida λ_u sea no atómica.

Asumiendo lo anterior y tomando A tal que $\lambda_u(A) = l < \infty$ un resultado debido al teorema de Lyapunov, garantiza la existencia de una partición $A_{n,j}$, con $j = 1, \dots, n$, de A tal que $\lambda_u(A_{n,j}) = \frac{l}{n}$. De esta manera se tiene que

$$\mathbb{E} \left[e^{-u\Phi(A_{n,j})} \right] = e^{-\frac{l}{n}}.$$

Ahora, bien la desigualdad de Markov, garantiza que para una variable aleatoria X no negativa y una constante $a > 0$

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}(X)}{a}.$$

De esta manera se obtiene lo siguiente,

$$\begin{aligned} \mathbb{P}(\Phi(A_{n,j}) \geq c) &= \mathbb{P}(e^{-u\Phi(A_{n,j})} \leq e^{-c}) \\ &= \mathbb{P}(1 - e^{-u\Phi(A_{n,j})} \geq 1 - e^{-c}) \\ &\leq \frac{1 - e^{-\frac{l}{n}}}{1 - e^{-c}}. \end{aligned}$$

Donde el último término converge uniformemente para toda j cuando n tiende a infinito.

Luego, como para toda n

$$\Phi(A) = \sum_{j=1}^n \Phi(A_{n,j}).$$

Se tiene que $\Phi(A)$ es infinitamente divisible, por lo que usando la representación de Lévy-Khintchine (para variables no negativas), se tiene que

$$\mathbb{E}(e^{-u\Phi(A)}) = \exp \left(-\beta(A)u - \int_{\mathbb{R}_+} (1 - e^{-us})\nu(A, ds) \right). \quad (1.7)$$

Donde $\beta \geq 0$ y $\nu(A, \cdot)$ es una medida en \mathbb{R}_+ . De esta manera λ_u tiene la siguiente representación.

$$\lambda_u(A) = \beta(A)u + \int_{\mathbb{R}_+} (1 - e^{-us})\nu(A, ds). \quad (1.8)$$

Al ser λ_u una medida, se tiene que $\beta(A)$ y $\nu(A, \cdot)$ están completamente determinadas por $\lambda_u(A)$ ya que se debe cumplir lo siguiente

$$\beta\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \beta(A_i),$$

$$\nu\left(\bigcup_{i=1}^{\infty} A_i, \cdot\right) = \sum_{i=1}^{\infty} \nu(A_i, \cdot).$$

Por lo que se tiene una representación para medidas completamente aleatorias Σ finitas y sin átomos fijos en términos de las ecuaciones (1.6) y (1.7). Con β una medida (no atómica) en S , $\nu(A, \cdot)$ medida en $\mathbb{R}_+ \cup \{\infty\}$ y $\nu(\cdot, B)$ medida (no atómica) en S . Sin embargo, no es todo lo que podemos decir sobre esta representación. A continuación se verá la relación que se tiene con los procesos Poisson.

El teorema de Campbell y la ecuación (1.6) están ligadas de tal forma que gracias a esta conexión se pueden construir medidas completamente aleatorias a partir del conocimiento de β y ν , definidas como arriba. Para esto se hará uso de la derivada de Radon-Nikodym, por lo que primero se enuncia este resultado.

Teorema 1.21. *Teorema de Radon-Nikodym*

Sea (S, \mathcal{S}) un espacio medible y α, γ dos medidas definidas en este espacio tal que α es σ -finita y γ es finita y absolutamente continua con respecto a α . Entonces existe $f \in \mathcal{L}^1$ tal que para toda $A \in \mathcal{S}$

$$\gamma(A) = \int_A f d\alpha.$$

Donde f es conocida como la derivada de Radon-Nikodym.

Si se supone que la medida definida en S como

$$\mu(A) = \nu(A, (0, \infty])$$

es σ -finita, se tiene que para $s > 0$ la medida

$$\mu_u(A) = \nu(A, (0, s])$$

es absolutamente continua con respecto a μ por lo que tiene una derivada de Radon-Nikodym que se denotará por $F(x, s)$, por lo que

$$\mu_u(A) = \int_A F(x, s) \mu(dx).$$

Donde $F(\cdot, s)$ está determinada salvo en un conjunto de medida cero para cada s . Para demostrar que $F(x, s)$ es la distribución de una variable aleatoria con soporte en $(0, \infty]$ para toda s , primero se debe tomar una versión para s en los racionales y después se extiende para toda s por continuidad por la derecha. Esto se puede consultar más a detalle en Kingman (1967).

Con lo anterior en mente se puede construir a la siguiente medida en el espacio producto $S^* = S \times (0, \infty]$

$$\mu^*(A^*) = \iint_{A^*} dF(x, s) \mu(dx).$$

Cuando $A^* = A \times B$, se tiene que

$$\mu^*(A \times B) = \int_A \int_B dF(x, s) \mu(dx) = \int_B d\nu(A, (0, s]) = \nu(A, B).$$

Se sabe que en general la medida ν no tiene porqué ser σ -finita; sin embargo, se sabe que cumple con una condición de integrabilidad a decir:

$$\int (1 - e^{-s}) \nu(S_j, ds) < \infty.$$

De tal manera que para todo $\epsilon > 0$ se tiene que $\nu(S_j, (\epsilon, \infty]) < \infty$ y así se puede definir a

$$\mu^n(A) := \nu\left(A, \left(\frac{1}{k+1}, \frac{1}{k}\right)\right)$$

que es una medida σ finita y a

$$\mu := \sum_{n=1}^{\infty} \mu^n.$$

Aplicando todo lo anterior a μ^n y después sumándolas para obtener a μ se tiene que en efecto hay una medida μ^* en S^* tal que $\mu^*(A \times B) = \nu(A, B)$.

De esta manera si Π^* es un proceso Poisson en S^* con medida media μ^* y se define a

$$\Psi(A) := \sum \{s : (x, s) \in \Pi^*, x \in A\},$$

se tiene que Ψ es una medida puramente atómica en S cuyos átomos corresponden a los del proceso Poisson Π^* . Por lo que si (x, s) es un punto en Π^* se tiene que Ψ tiene un átomo con peso s en x . Más aún, Ψ es una medida completamente aleatoria y su distribución se puede obtener a través del teorema de Campbell como,

$$\begin{aligned} \mathbb{E} [e^{-\Psi(A)}] &= \exp \left[- \int_A \int_{\mathbb{R}_+} (1 - e^{-us}) \mu^*(dx, ds) \right] \\ &= \exp \left[- \int_{\mathbb{R}_+} (1 - e^{-us}) \nu(A, ds) \right] \end{aligned} \quad (1.9)$$

De esta manera comparando las respectivas expresiones obtenidas para Φ (1.7) y para Ψ (1.9), se puede observar que

$$\Phi = \beta + \Psi.$$

Siendo esta finalmente la representación para medidas completamente aleatorias Σ finitas. Todo lo anterior puede ser resumido en el siguiente teorema.

Teorema 1.22. *Teorema de representación de Kingman.*

Sea Φ una medida completamente aleatoria definida en un espacio medible (S, \mathcal{S}) tal que Φ es Σ finita. Entonces, existen medidas β , Ψ y Φ_f tal que

$$\Phi = \Phi_f + \beta + \Psi.$$

Donde

1. β es una medida determinista y no atómica definida en S .
2. Ψ es una medida completamente aleatoria y puramente atómica. Y que puede ser escrita como

$$\Psi = \sum_{x \in \Pi^*} s_x \delta_x.$$

Esto es, Ψ tiene átomos correspondientes a los puntos de un proceso Poisson Π^* con medida media μ^* y con pesos aleatorios dados por las variables aleatorias $\{s_x\}_{x \in \Pi^*}$ de distribución $F(x, \cdot)$.

3. Φ_f es una medida completamente aleatoria con átomos fijos en el subconjunto contable $X = \{x_1, x_2, \dots\}$. Que puede ser escrita como,

$$\Phi_f = \sum_{x \in X} \phi(x) \delta_x.$$

Donde las variables aleatorias $\{\phi(x)\}_{x \in X}$ son independientes. Y además Φ_f es independiente de $\beta + \Psi$.

Para fines de la normalización se tomará a $\Phi = \Psi$, por lo cual se estará trabajando con medidas completamente aleatorias que cumplen con ser discretas casi seguramente. De esta manera pueden ser caracterizadas a partir de la ecuación (1.9),

$$\mathbb{E} [e^{-u\Phi(A)}] = \exp \left[- \int_{\mathbb{R}_+} (1 - e^{-us}) \nu(A, ds) \right].$$

También se sabe que al ser ésta su representación de Lévy-Khintchine, ν es denominada la medida de Lévy. Esta medida contiene toda la información referente sobre la distribución de los saltos y la localización. Esta medida se suele separar (para fines prácticos) como $\nu(dx, ds) = \alpha(dx)\rho_x(ds)$. Donde α es una medida en (S, \mathcal{S}) y ρ es un kernel de transición en $\mathcal{S} \times \mathcal{B}(\mathbb{R}^+)$ y donde a partir de ρ se pueden tener dos clases de medidas de Lévy.

1. Se dice que la medida de Lévy ν es homogénea si $\rho_x(ds) = \rho(ds)$. Que significa que los saltos serán independientes de los puntos donde se localiza el salto.
2. De lo contrario, se dice que ν es no-homogénea.

Para terminar el capítulo se presentan dos medidas de Lévy que serán estudiadas con mayor profundidad en el capítulo 3. La primera de ellas es la correspondiente con el proceso gama. Para esto sea ν como sigue

$$\nu(dx, ds) = \frac{e^{-s}}{s} ds \alpha(dx),$$

que es una medida homogénea.

La segunda medida de Lévy a considerar es la correspondiente con el proceso σ -estable.

Sea $\sigma \in (0, 1)$ y la siguiente intensidad de Lévy

$$\nu(dx, ds) = \frac{\sigma}{\Gamma(1 - \sigma) s^{1+\sigma}} ds \alpha(dx),$$

que también es una medida homogénea.

Capítulo 2

Construcción vía normalización

A partir del teorema de representación para variables aleatorias intercambiables se sabe que uno de los principales objetivos de la estadística bayesiana no paramétrica es la construcción de distribuciones iniciales para P , que es una medida de probabilidad aleatoria (RPM por sus siglas en inglés). Para esto ya se ha hecho una introducción al concepto de medidas completamente aleatorias y a sus propiedades, siendo una de las principales el que sean discretas casi seguramente.

La normalización de medidas completamente aleatorias es una herramienta que ha sido utilizada en una gran variedad de contextos alejados de la estadística bayesiana como problemas de almacenamiento y en ciertas áreas de la ciencia como ecología, genética, entre otras. A pesar de que una de las construcciones dadas por Ferguson (1973) del proceso Dirichlet viene de la normalización del proceso gama, esta técnica fue retomada hasta los trabajos de Regazzini, Lijoi y Prünster (2003) donde se presenta una normalización de un proceso con incrementos independientes para generar una medida de probabilidad aleatoria en \mathbb{R} . Lo que conllevó a un claro beneficio para la estadística bayesiana no paramétrica.

Como se hará una normalización de estas medidas se debe restringir a una clase especial de ellas. De esta manera, este capítulo se centra en la construcción y obtención de las principales características como: la distribución posterior, la distribución predictiva, la función de probabilidad para particiones intercambiables, entre otras. Para el desarrollo del capítulo se hace especial enfoque en la teoría desarrollada por Hjort, Holmes, Müller y Walker (2010) y en los artículos de Lijoi y Prünster (2009) y Lijoi, James y Prünster (2009).

2.1. Medidas completamente aleatorias normalizadas

A partir de este momento se debe considerar un espacio polaco $(\mathbb{X}, \mathcal{X})$ y se denotará por $\mathcal{M}_{\mathcal{X}}$ al espacio de medidas en $(\mathbb{X}, \mathcal{X})$ tales que para toda $\Phi \in \mathcal{M}_{\mathbb{X}}$ y cualquier conjunto acotado $A \in \mathcal{X}$ se tiene que $\Phi(A) < \infty$. Este espacio de medidas estará dotado con su σ -álgebra de Borel, $\mathcal{M}_{\mathbb{X}}$.

Recordando que si Φ es un mapeo medible de $(\Omega, \mathcal{F}, \mathbb{P})$ a $(\mathbb{M}_{\mathbb{X}}, \mathcal{M}_{\mathbb{X}})$ tal que para conjuntos ajenos $A_1, \dots, A_n \in \mathcal{X}$, las variables aleatorias $\Phi(A_1), \dots, \Phi(A_n)$ son independientes, entonces se dice que Φ es una medida completamente aleatoria; para la cual, se tiene un teorema de representación (1.22) que garantiza que se puede pensar a Φ como la suma de tres componentes, uno determinista, un componente aleatorio con átomos fijos (subconjunto contable) y un componente donde los saltos y sus localizaciones son aleatorios.

Pensando por un momento que el componente determinista es idénticamente cero, se puede escribir a Φ como

$$\Phi = \sum_{i=1}^M V_i \delta_{x_i} + \sum_{i=1}^{\infty} J_i \delta_{X_i} = \Phi_f + \Psi. \quad (2.1)$$

Donde $\{x_i\}_{i=1}^M$, con $M \in \mathbb{N} \cup \{\infty\}$, son los puntos fijos con masas aleatorias $\{V_i\}_{i=1}^M$ y $\{X_i\}_{i=1}^{\infty}$ son las localizaciones aleatorias con saltos aleatorios $\{J_i\}_{i=1}^{\infty}$. Y $\{V_i\}_{i=1}^M$ son variables mutuamente independientes e independientes de Ψ .

Más aún, se sabe que Ψ está completamente determinada por su representación de Lévy-Khintchine, por lo que si f es una función medible de $(\mathbb{X}, \mathcal{X})$ a $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ tal que

$$\int |f| d\Psi < \infty$$

entonces, se tiene que la funcional de Laplace estará dada por

$$\mathbb{E} \left[\exp \left(- \int_{\mathbb{X}} f(x) \Psi(dx) \right) \right] = \exp \left[- \int_{\mathbb{R}_+ \times \mathbb{X}} (1 - e^{-sf(x)}) \nu(ds, dx) \right]. \quad (2.2)$$

Donde ν cumple con la siguiente condición de integrabilidad

$$\int_B \int_{\mathbb{R}_+} (s \wedge 1) \nu(ds, dx) < \infty.$$

Y se le conoce como la medida de Lévy y que además contiene toda la información referente a los saltos y a los sitios donde ocurren. De esta manera se puede dar una primera definición de lo que se denotará como una NRMI.

Definición 2.1. *NRMI*

Sea Φ una CRM en \mathbb{X} tal que $0 < \Phi(\mathbb{X}) < \infty$ c.s. Entonces la medida de probabilidad aleatoria \mathcal{P} definida como:

$$\mathcal{P} = \frac{\Phi}{\Phi(\mathbb{X})}.$$

Es una medida completamente aleatoria normalizada con incrementos independientes (NRMI).

A partir de la definición es claro el por qué $\Phi(\mathbb{X})$ debe ser finita y positiva. Ahora bien, sabiendo que Φ está completamente caracterizada por la intensidad de Lévy asociada ν , las condiciones impuestas para Φ pueden ser reexpresadas en términos de ν . Recordando la descomposición de $\nu(ds, dx) = \rho_x(ds)\alpha(dx)$, se tiene lo siguiente:

1. Que Φ sea finita es equivalente a pedir que el exponente laplaciano sea finito para todo $u \geq 0$, i.e.,

$$\varphi(u) = \int_{\mathbb{R}_+ \times \mathbb{X}} (1 - e^{-us}) \rho_x(ds) \alpha(dx) < \infty. \quad (2.3)$$

Cuando ν es homogénea esta condición se reduce a pedir que $\theta := \alpha(\mathbb{X}) < \infty$. Cuando ocurra esto, la medida α puede ser representada como $\alpha(dx) = \theta P_0$ donde P_0 es una medida de probabilidad en \mathbb{X} .

2. Que Φ sea positiva es equivalente a pedir que exista un conjunto A tal que $\alpha(A) > 0$ y que

$$\int_A \rho_x(\mathbb{R}_+) \alpha(dx) = \infty.$$

Cuando ν es homogénea es equivalente a pedir que $\rho(\mathbb{R}_+) = \infty$ i.e., que sea una medida de actividad infinita.

Una vez definida la medida de probabilidad aleatoria el siguiente paso es la descripción del análisis posterior que ya se ha mencionado en el primer capítulo. De esta manera, en la siguiente sección se verá la construcción de la distribución posterior que, salvo un caso, representa desafíos importantes.

2.2. Distribución posterior

Un concepto que permite que el análisis posterior sea más sencillo en la estadística bayesiana es el de familias conjugadas. Lo que en el caso paramétrico se define como:

Definición 2.2. *Familias conjugadas.*

Se dice que una clase \mathcal{C} de distribuciones iniciales para θ forma una familia conjugada con respecto a $f(x|\theta)$ si la distribución posterior pertenece a \mathcal{C} .

De esta manera y dada la mayor complejidad matemática inherente a los problemas no paramétricos, se podría buscar esta característica al momento de normalizar para facilitar los cálculos; sin embargo, se ha demostrado que las NRMI's no son conjugadas, salvo para el proceso Dirichlet. Por lo que para poder obtener la distribución posterior se debe introducir una variable latente gracias a la cual se obtiene que la distribución posterior también sea una NRMI.

Como es usual, primero se debe considerar una sucesión $\{X_n\}_{n=1}^{\infty}$ de variables aleatorias intercambiables definidas en un espacio de probabilidad $(\Omega, \mathcal{F}, \mathbb{P})$ con valores en el espacio polaco \mathbb{X} , de tal manera que recordando el modelo jerárquico se tiene que

$$\mathbb{P}(X_1 \in A_1, \dots, X_n \in A_n | \mathcal{P}) = \prod_{i=1}^n \mathcal{P}(A_i). \quad (2.4)$$

Dado que se están considerando medidas completamente aleatorias puramente atómicas, el vector aleatorio $\mathbf{X}^{(n)} = (X_1, \dots, X_n)$ admite la siguiente representación $(\mathbf{Y}^{(n)}, \pi)$ en donde $\mathbf{Y}^{(n)} = (Y_1, \dots, Y_{n(\pi)})$ denota las distintas observaciones y π es una partición del conjunto $[n] := \{1, \dots, n\}$ de tamaño $n(\pi)$, que permite identificar cuáles observaciones fueron iguales. El número de elementos en el i -ésimo bloque está dado por n_i para $i = 1, \dots, n(\pi)$ de tal forma que $\sum_{i=1}^{n(\pi)} n_i = n$.

Ahora bien, si se considera a W_n una variable aleatoria gama de parámetros $(n, 1)$ independiente de $\Phi(\mathbb{X}) = T$, se define a la variable aleatoria

$$U_n := \frac{W_n}{T}.$$

Cuya densidad está dada por

$$f_{U_n}(u) = \int_{\mathbb{R}_+} f_{W_n}(uv) f_T(v) v dv = \frac{u^{n-1}}{\Gamma(n)} \int_{\mathbb{R}_+} v^n f_T(v) e^{-uv} dv.$$

Es importante mencionar que la distribución de T se asume absolutamente continua con respecto a la medida de Lebesgue, por lo cual admite una densidad y se puede llegar al resultado anterior.

Para el análisis posterior que se necesita, la distribución de $U_n|\mathbf{X}^{(n)}$ será de gran importancia. En lo que sigue, para una pareja de objetos aleatorios X e Y definidos en $(\Omega, \mathcal{F}, \mathbb{P})$ se denotará por $X^{(Y)}$ al objeto aleatorio cuya distribución coincide con la de distribución condicional de X dado Y . Con esto en mente se puede enunciar el teorema que ejercerá un papel primordial en el análisis que sigue.

Teorema 2.3. *Sea \mathcal{P} una NRMI con intensidad de Lévy, $\nu(ds, x) = \rho_x(ds)\alpha(dx)$. Entonces*

$$\Phi^{(U_n, \mathbf{X}^{(n)})} \stackrel{d}{=} \Phi^{(U_n)} + \sum_{i=1}^{n(\pi)} J_i^{(U_n, \mathbf{X}^{(n)})} \delta_{Y_i}.$$

Donde

1. $\Phi^{(U_n)}$ es una medida completamente aleatoria con intensidad

$$\nu^{(U_n)}(ds, dx) = e^{-U_n s} \rho_x(ds)\alpha(dx).$$

2. Y_i con $i = 1, \dots, n(\pi)$ son los puntos fijos de discontinuidad y $J_i^{(U_n, \mathbf{X}^{(n)})}$ son sus correspondientes saltos que son mutuamente independientes e independientes de $\Phi^{(U_n)}$ y cuya densidad $f_i(s)$ es tal que

$$f_i(s) \propto s^{n_i} e^{-U_n s} \rho_{Y_i}(ds).$$

Demostración. Hay dos formas de demostrar este teorema, una trabajando con el proceso Poisson asociado a la medida completamente aleatoria y otra trabajando con la funcional de Laplace. Las dos pruebas se pueden encontrar en el trabajo de Lijoi, James y Prünster (2009). En este trabajo se hace énfasis en la prueba a partir del proceso Poisson y que está guiada en resultados descritos por James (2005) y que por completez se pueden consultar en el Apéndice B.

Recordando que se está considerando a Φ de tal forma que es discreta casi seguramente, se tiene que $\Phi = \Psi = \sum_{x \in \Pi^*} J_i \delta_{X_i}$, donde Ψ tiene una construcción basada en un proceso Poisson, Π , definido en $\mathbb{R}_+ \times \mathbb{X}$ y de intensidad ν , por lo que para todo $B \in \mathcal{X}$ se tiene

$$\Phi(B) = \int_{\mathbb{R}_+ \times B} s \Pi(ds, dx).$$

Denotando por \mathbb{P}_ν a la distribución del proceso Poisson asociado, Π de intensidad ν y a $\mathbb{E}_\nu(\cdot)$ como la esperanza respecto a \mathbb{P}_ν . Y recordando que la medida de probabilidad aleatoria está definida como $\mathcal{P}(dx) := \theta^{-1} \Phi(dx)$, se define a

$$\tilde{\Pi}_n := \Pi' + \sum_{i=1}^{n(\pi)} \delta_{J_i, Y_i} \quad \text{y} \quad \tilde{\Phi}_n(dx) := \Phi'(dx) + \sum_{i=1}^{n(\pi)} J_i \delta_{Y_i}(dx).$$

Donde Π' y Φ' son de la misma forma que Π y Φ respectivamente. Así,

$$\tilde{\Phi}_n(\mathbb{X}) = \Phi'(\mathbb{X}) + \sum_{i=1}^{n(\pi)} J_i = T' + \sum_{i=1}^{n(\pi)} J_i.$$

Siguiendo la línea del artículo de James, se debe definir a

$$q(Y_j, \tilde{\Phi}_n) := \tilde{\Phi}_n(\mathbb{X})^{-1} = \frac{1}{T' + \sum_{i=1}^{n(\pi)} J_i}.$$

Que no depende de los puntos de discontinuidad Y_i ni del índice j , por lo que

$$\prod_{j=1}^{n(\pi)} [q(Y_j, \tilde{\Phi}_n)]^{n_j} = \frac{1}{(T' + \sum_{i=1}^{n(\pi)} J_i)^n}.$$

También, si se denota por $\mathbf{J} = (J_1, \dots, J_{n(\pi)})$ al vector de saltos, se debe definir a

$$\phi_n(\mathbf{J}, \mathbf{X}^{(n)}) := \int_{\mathbb{M}_x} \frac{\mathbb{P}_\nu(d\Pi)}{(T' + \sum_{i=1}^{n(\pi)} J_i)^n} = \int_{\mathbb{R}_+} \frac{f_T(t)}{(t + \sum_{i=1}^{n(\pi)} J_i)^n} dt.$$

El Teorema B.2. inciso 1., que corresponde al teorema 3.2 del trabajo de James (2005), permite asegurar que la distribución posterior de $\Phi|\mathbf{X}^{(n)}$ es equivalente a la distribución de $\tilde{\Phi}_n|\mathbf{X}^{(n)}$ y está determinada por la distribución posterior de $\Pi|\mathbf{X}^{(n)}$ que a su vez es equivalente a la distribución de $\tilde{\Pi}_n|\mathbf{X}^{(n)}$.

Usando de nueva cuenta el inciso 1. del Teorema B.2, se puede obtener la distribución de (Π', \mathbf{J}) dado $\mathbf{X}^{(n)}$, como sigue,

$$\begin{aligned}
f(\Pi', \mathbf{J}|\mathbf{X}^{(n)}) &= f(\Pi'|\mathbf{J}, \mathbf{X}^{(n)})f(\mathbf{J}|\mathbf{X}^{(n)}) \\
&\propto \phi_n^{-1}(\mathbf{J}, \mathbf{X}^{(n)}) \left(\prod_{j=1}^{n(\pi)} [q(Y_j, \tilde{\Phi}_n)]^{n_j} \right) \mathbb{P}_\nu(d\Pi') \phi_n(\mathbf{J}, \mathbf{X}^{(n)}) \prod_{j=1}^{n(\pi)} J_j^{n_j} \rho_{Y_j}(dJ_j) \\
&= \frac{1}{(T' + \sum_{i=1}^{n(\pi)} J_i)^n} \mathbb{P}_\nu(d\Pi') \prod_{j=1}^{n(\pi)} J_j^{n_j} \rho_{Y_j}(dJ_j). \tag{2.5}
\end{aligned}$$

Por lo que la distribución de (Π', \mathbf{J}) dado $\mathbf{X}^{(n)}$, evaluada en un punto $(\pi', s_1, \dots, s_{n(\pi)})$, es proporcional a

$$\frac{1}{(T' + \sum_{i=1}^{n(\pi)} s_i)^n} \mathbb{P}_\nu(d\pi') \prod_{j=1}^{n(\pi)} s_j^{n_j} \rho_{Y_j}(ds_j) \tag{2.6}$$

Que es una expresión que determina la distribución posterior de $\Phi|\mathbf{X}^{(n)}$. Una vez teniendo lo anterior sólo hace falta ver lo que ocurre al introducir la variable latente U_n , para esto, primero se debe hacer notar que al considerar a Z una variable aleatoria con distribución gama de parámetros (α, β) ,

$$\int_{\mathbb{R}_+} \frac{z^{\alpha-1} \beta^\alpha e^{-z\beta}}{\Gamma(\alpha)} dz = 1 \quad \Rightarrow \quad \frac{1}{\beta^\alpha} = \int_{\mathbb{R}_+} \frac{z^{\alpha-1} e^{-z\beta}}{\Gamma(\alpha)} dz.$$

De esta manera,

$$\frac{1}{(T' + \sum_{i=1}^{n(\pi)} J_i)^n} = \frac{1}{\Gamma(n)} \int_{\mathbb{R}_+} u^{n-1} \exp \left[-u \left(T' + \sum_{i=1}^{n(\pi)} J_i \right) \right] du$$

Usando la identidad anterior, conocida como identidad gama, la expresión (2.6) puede ser aumentada para poder encontrar la distribución conjunta del vector $(\Pi', \mathbf{J}, U_n, \mathbf{X}^{(n)})$, que es proporcional a

$$u^{n-1} e^{-uT'} \mathbb{P}_\nu(d\pi') \prod_{j=1}^{n(\pi)} s_j^{n_j} e^{-us_j} \rho_{Y_j}(ds_j) \alpha(dY_j). \quad (2.7)$$

Utilizando la Proposición B.1., se obtiene la siguiente igualdad para las medidas,

$$e^{-uT'} \mathbb{P}_\nu(d\pi') = e^{-\varphi(u)} \mathbb{P}_{\nu_u}(d\pi').$$

Donde $\nu_u = e^{-us} \rho_x(ds) \alpha(dx)$ y $\mathbb{E}_\nu(e^{-uT'}) = e^{-\varphi(u)}$, por lo que (2.7) se puede reescribir como

$$u^{n-1} e^{-\varphi(u)} \mathbb{P}_{\nu_u}(d\pi') \prod_{j=1}^{n(\pi)} s_j^{n_j} e^{-us_j} \rho_{Y_j}(ds_j) \alpha(dY_j) \quad (2.8)$$

La expresión (2.8) permite derivar una expresión para la distribución de $\Pi', \mathbf{J} | U_n, \mathbf{X}^{(n)}$, que es proporcional a

$$\mathbb{P}_{\nu_u}(d\pi') \prod_{j=1}^{n(\pi)} s_j^{n_j} e^{-us_j} \rho_{Y_j}(ds_j).$$

Donde $\Pi' | U_n, \mathbf{X}^{(n)}$ es un proceso Poisson con intensidad dada por $\nu^{(U_n)} = e^{-U_n s} \rho_x(ds) \alpha(dx)$, que tiene asociada una medida completamente aleatoria asociada, que se denotó como $\Phi^{(U_n)}$. Y los $J_i^{(U_n, \mathbf{X}^{(n)})}$ son los puntos de discontinuidad, independientes de Π' y cuya densidad es proporcional a

$$s_i^{n_i} e^{-U_n s_i} \rho_{Y_i}(ds_i).$$

□

El teorema anterior muestra que condicional a una variable latente, la distribución posterior de la medida completamente aleatoria Φ sigue siendo una medida completamente aleatoria con puntos fijos de discontinuidad correspondientes a las $n(\pi)$ distintas observaciones. Otro resultado importante derivado del Teorema 2.3 es la siguiente proposición.

Proposición 2.4. *Sea \mathcal{P} una NRMI. Entonces la distribución condicional de U_n dado $\mathbf{X}^{(n)}$ admite una función de densidad $f_{U_n|\mathbf{X}^{(n)}}(u)$ tal que,*

$$f_{U_n|\mathbf{X}^{(n)}}(u) \propto u^{n-1} \prod_{i=1}^{n(\pi)} \tau_{n_i}(u|Y_i) e^{-\varphi(u)}.$$

Donde

$$\tau_{n_i}(u|Y_i) = \int_{\mathbb{R}_+} s^{n_i} e^{-us} \rho_{Y_i}(ds), \quad i = 1, \dots, n(\pi),$$

y $\varphi(u)$ es el exponente laplaciano como en (2.3).

Demostración. Para mostrar este resultado basta recordar la expresión (2.8), la cual da salvo una constante de proporcionalidad la distribución del vector $(\Pi', \mathbf{J}, U_n, \mathbf{X}^{(n)})$ y que es igual a

$$u^{n-1} e^{-\varphi(u)} \mathbb{P}_{\nu_u}(d\pi') \prod_{j=1}^{n(\pi)} s_j^{n_j} e^{-us_j} \rho_{Y_j}(ds_j) \alpha(dY_j).$$

Basta integrar con respecto a π y a las s_i para obtener una expresión para la distribución conjunta de $(U_n, \mathbf{X}^{(n)})$. No es un cálculo complicado; sin embargo, lo que podemos notar del proceso es que al integrar sobre las s_i se obtiene la expresión dada por $\tau_{n_i}(u|Y_i)$, de esta manera la distribución conjunta es proporcional a

$$u^{n-1} e^{-\varphi(u)} \prod_{i=1}^{n(\pi)} \left[\int_{\mathbb{R}_+} s^{n_i} e^{-us} \rho_{Y_i}(ds) \right] \alpha(dY_i)$$

Y de donde se sigue que la distribución de $U_n|\mathbf{X}^{(n)}$ es proporcional a

$$u^{n-1} e^{-\varphi(u)} \prod_{i=1}^{n(\pi)} \tau_{n_i}(u|Y_i)$$

□

La proposición anterior será de gran utilidad al momento de calcular la distribución predictiva. Para finalizar la sección se presenta el teorema del cual se desprende la distribución posterior de \mathcal{P} .

Teorema 2.5. *Sea \mathcal{P} una NRMI con intensidad de Lévy $\nu(ds, dx) = \rho_x(ds)\alpha(dx)$, entonces la distribución posterior de \mathcal{P} dada U_n es nuevamente una NRMI, con puntos fijos de discontinuidad y es tal que*

$$\mathcal{P} | (\mathbf{X}^{(n)}, U_n) \stackrel{d}{=} w \frac{\Phi^{(U_n)}}{\Phi^{(U_n)}(\mathbb{X})} + (1 - w) \frac{\sum_{i=1}^{n(\pi)} J_i^{(U_n, \mathbf{X}^{(n)})} \delta_{Y_i}}{\sum_{r=1}^{n(\pi)} J_r^{(U_n, \mathbf{X}^{(n)})}}.$$

Donde

$$w = \frac{\Phi^{(U_n)}(\mathbb{X})}{\Phi^{(U_n)}(\mathbb{X}) + \sum_{i=1}^{n(\pi)} J_i^{(U_n, \mathbf{X}^{(n)})}}.$$

De esta manera se tiene que la distribución posterior (condicional a la variable latente) es una mezcla de NRMI's con puntos fijos de discontinuidad que corresponden a las distintas observaciones. La introducción de una variable latente causa que sea más elaborado el proceso; sin embargo, permite tener una representación de la distribución posterior de la que se pueden obtener algoritmos para muestrear trayectorias de una manera ciertamente no tan complicada.

2.3. Distribución predictiva

Cuando un problema estadístico se resuelve basado en la metodología bayesiana se sabe que además de la distribución posterior, la distribución predictiva tiene un rol de suma importancia, ya que permite estudiar el comportamiento de futuras observaciones, lo cual es vital en estudios de muestreo de especies por mencionar uno.

Para obtener dicha distribución predictiva, ya se tiene la herramienta necesaria que fue desarrollada en la sección anterior y del cual se puede desprender la siguiente proposición.

Proposición 2.6. Dada \mathcal{P} una NRMi con intensidad $\nu(ds, dx) = \rho_x(ds)\alpha(dx)$, entonces la distribución predictiva de X_{n+1} dado $\mathbf{X}^{(n)}$ está dada por

$$\mathbb{P}(X_{n+1} \in dx | \mathbf{X}^{(n)}) = \omega^{(n)} P_0(dx) + \frac{1}{n} \sum_{i=1}^{n(\pi)} \omega_i^{(n)} \delta_{Y_i}(dx). \quad (2.9)$$

Donde para $i = 1, \dots, n(\pi)$

$$\omega^{(n)} = \frac{\theta}{n} \int_{\mathbb{R}_+} u \tau_1(u|x) f_{U_n | \mathbf{X}^{(n)}}(u) du, \quad \omega_i^{(n)} = \int_{\mathbb{R}_+} u \frac{\tau_{n_i+1}(u|Y_i)}{\tau_{n_i}(u|Y_i)} f_{U_n | \mathbf{X}^{(n)}}(u) du$$

Demostración. La demostración está basada en el hecho de que el lado izquierdo de la ecuación (2.9) puede ser escrito como

$$\mathbb{P}(X_{n+1} \in dx | \mathbf{X}^{(n)}) = \mathbb{E}(\mathcal{P}(dx) | \mathbf{X}^{(n)}) = \int_{\mathbb{R}_+} \mathbb{E}(\mathcal{P}(dx) | U_n = u, \mathbf{X}^{(n)}) f_{U_n | \mathbf{X}^{(n)}}(u) du.$$

Utilizando el teorema 2.5, se tiene que

$$\begin{aligned} \mathbb{E}(\mathcal{P}(dx) | U_n = u, \mathbf{X}^{(n)}) &= \mathbb{E} \left[\omega \frac{\Phi^{(u)}(dx)}{T^{(u)}} \right] + \mathbb{E} \left[(1 - \omega) \frac{\sum_{i=1}^{n(\pi)} J_i^{(u, \mathbf{X}^{(n)})} \delta_{Y_i}(dx)}{\sum_{r=1}^{n(\pi)} J_r^{(u, \mathbf{X}^{(n)})}} \right] \\ &= \mathbb{E} \left[\frac{\Phi^u(dx)}{T^u + \sum_{i=1}^{n(\pi)} J_i^{(u, \mathbf{X}^{(n)})}} \right] + \mathbb{E} \left[\frac{\sum_{i=1}^{n(\pi)} J_i^{(u, \mathbf{X}^{(n)})} \delta_{Y_i}(dx)}{T^u + \sum_{i=1}^{n(\pi)} J_i^{(u, \mathbf{X}^{(n)})}} \right] \\ &= I_1(u, x, \mathbf{X}^{(n)}) + I_2(u, x, \mathbf{X}^{(n)}). \end{aligned}$$

De esta manera,

$$\mathbb{P}(X_{n+1} \in dx | \mathbf{X}^{(n)}) = \int_{\mathbb{R}_+} I_1(u, x, \mathbf{X}^{(n)}) f_{U_n | \mathbf{X}^{(n)}}(u) du + \int_{\mathbb{R}_+} I_2(u, x, \mathbf{X}^{(n)}) f_{U_n | \mathbf{X}^{(n)}}(u) du. \quad (2.10)$$

Lo que se mostrará es que el primer sumando de la expresión (2.10) será igual a $\omega^{(n)}\alpha(dx)$, para esto primero se debe ver lo siguiente,

$$\begin{aligned} I_1(u, x, \mathbf{X}^{(n)}) &= \int_{\mathbb{R}_+ \times \mathbb{X}} \frac{\Phi^u(dx)}{T^u + \sum_{i=1}^{n(\pi)} J_i^{(u, \mathbf{X}^{(n)})}} \mathcal{P}_{\Phi^u}(d\Phi^u) \\ &= \int_{\mathbb{R}_+} \mathbb{E} \left[e^{-s \sum_{i=1}^{n(\pi)} J_i^{(u, \mathbf{X}^{(n)})}} \right] \mathbb{E}[\Phi^u(dx) e^{-sT^u}] ds \end{aligned}$$

Donde,

$$\begin{aligned} \mathbb{E} \left[e^{-v \sum_{i=1}^{n(\pi)} J_i^{(u, \mathbf{X}^{(n)})}} \right] &= \prod_{i=1}^{n(\pi)} \mathbb{E} \left[e^{-v J_i^{(u, \mathbf{X}^{(n)})}} \right] \\ &= \prod_{i=1}^{n(\pi)} \int_{\mathbb{R}_+} e^{-vs} \frac{s^{n_i} e^{-us} \rho_{Y_i}(ds)}{\int_{\mathbb{R}_+} e^{-vs} s^{n_i} e^{-us} \rho_{Y_i}(ds)} \\ &= \prod_{i=1}^{n(\pi)} \int_{\mathbb{R}_+} \frac{s^{n_i} e^{-(u+v)s} \rho_{Y_i}(ds)}{\tau_{n_i(u|Y_i)}} \\ &= \prod_{i=1}^{n(\pi)} \frac{\tau_{n_i(u+v|Y_i)}}{\tau_{n_i(u|Y_i)}}. \end{aligned}$$

Y

$$\begin{aligned} \mathbb{E}[\Phi^u(dx) e^{-sT^u}] &= \int_{\mathbb{R}_+} \mathbb{E}(e^{-v\Phi^u}) s e^{-(u+v)s} \rho_x(ds) \alpha(dx) \\ &= \int_{\mathbb{R}_+} e^{-\log \mathbb{E}[e^{-v\Phi^u}]} \alpha(dx) s e^{-(u+v)s} \rho_x(ds) \\ &= e^{-\varphi^u(v)} \alpha(dx) \tau_1(u+v|x). \end{aligned}$$

Juntando estas expresiones y utilizando que

$$\begin{aligned} \varphi^u(v) + \varphi(u) &= \int_{\mathbb{R}_+ \times \mathbb{X}} (1 - e^{-sv}) e^{-su} \rho_x(ds) \alpha(dx) + \int_{\mathbb{R}_+ \times \mathbb{X}} (1 - e^{-su}) \rho_x(ds) \alpha(dx) \\ &= \int_{\mathbb{R}_+ \times \mathbb{X}} (1 - e^{-s(v+u)}) \rho_x(ds) \alpha(dx) \\ &= \varphi(v+u), \end{aligned}$$

se tiene que

$$I_1(u, x, \mathbf{X}^{(n)}) = \int_{\mathbb{R}_+} e^{-\varphi^u(v)} \alpha(dx) \tau_1(u+v|x) \prod_{i=1}^{n(\pi)} \frac{\tau_{n_i}(u+v|Y_i)}{\tau_{n_i}(u|Y_i)} dv.$$

Y el primer sumando de (2.10) es igual a

$$\begin{aligned} & \int_{\mathbb{R}_+} \int_{\mathbb{R}_+} e^{-\varphi^u(v)} \alpha(dx) \tau_1(u+v|x) \prod_{i=1}^{n(\pi)} \frac{\tau_{n_i}(u+v|Y_i)}{\tau_{n_i}(u|Y_i)} f_{U_n|\mathbf{X}^{(n)}}(u) dv du \\ &= \alpha(dx) \int_{\mathbb{R}_+} \int_{\mathbb{R}_+} e^{-\varphi(v+u)} \tau_1(u+v|x) \prod_{i=1}^{n(\pi)} \frac{\tau_{n_i}(u+v|Y_i)}{\tau_{n_i}(u|Y_i)} u^{n-1} \prod_{i=1}^{n(\pi)} \tau_{n_i}(u|Y_i) dv du \\ &= \alpha(dx) \int_{\mathbb{R}_+} \int_{\mathbb{R}_+} e^{-\varphi(v+u)} \tau_1(u+v|x) \prod_{i=1}^{n(\pi)} \tau_{n_i}(u+v|Y_i) u^{n-1} dv du \\ & \quad (\text{haciendo } w = v + u \text{ y } z = u) \\ &= \alpha(dx) \int_{\mathbb{R}_+} \prod_{i=1}^{n(\pi)} \tau_{n_i}(w|Y_i) \tau_1(w|x) e^{-\varphi(w)} \int_0^w z^{n-1} dz dw \\ &= \frac{\alpha(dx)}{n} \int_{\mathbb{R}_+} w^n \tau_1(w|x) \prod_{i=1}^{n(\pi)} \tau_{n_i}(w|Y_i) e^{-\varphi(w)} dw \\ &= \frac{\alpha(dx)}{n} \int_{\mathbb{R}_+} w \tau_1(w|x) \left[w^{n-1} \prod_{i=1}^{n(\pi)} \tau_{n_i}(w|Y_i) e^{-\varphi(w)} \right] dw \\ &= \frac{\theta \cdot \alpha(dx)}{\theta \cdot n} \int_{\mathbb{R}_+} w \tau_1(w|x) f_{U_n|\mathbf{X}^{(n)}}(w) dw \\ &= P_0(dx) w^{(n)}. \end{aligned}$$

Los pesos $w_i^{(n)}$ se obtiene de forma análoga. □

Se puede observar que la distribución predictiva es una combinación lineal de P_0 y de una versión ponderada de la distribución empírica, lo cual es de cierta manera intuitivo ya que se tiene que las NRMI son discretas casi seguramente, por lo cual hay probabilidad positiva de empates. Esta representación será de gran utilidad al momento de implementar esquemas de muestreo tipo Blackwell-MacQueen.

2.4. Función de probabilidad para particiones intercambiables (EPPF)

La teoría de las particiones aleatorias intercambiables tiene su origen en los estudios hechos por Kingman (1982) y más adelante desarrollada, entre otros, por Pitman (2006). Antes de estudiar la relación entre particiones aleatorias y las NRMI, se presenta una recapitulación sobre conceptos de combinatoria y particiones.

Dado un conjunto (finito) A , se dice que una partición de A en k bloques es una colección de conjuntos no vacíos y disjuntos $\{A_1, \dots, A_k\}$ tales que $\cup_{i=1}^k A_i = A$. Se denotará por $\mathcal{P}_{[n]}^k$ al conjunto de posibles particiones en k bloques de $[n] = \{1, \dots, n\}$ y $\mathcal{P}_{[n]} := \cup_{k=1}^n \mathcal{P}_{[n]}^k$ el conjunto de todas las particiones de $[n]$. Dada una partición en k bloques de $[n]$, se puede considerar al vector de tamaños de cada bloque, i.e., $(|A_1|, \dots, |A_k|)$ que forman una composición de n , que no es más que una secuencia de enteros no negativos tales que la suma es igual a n . Al conjunto de todas las posibles composiciones de n se le denotará por \mathcal{C}_n .

Una partición aleatoria de $[n]$, denotada a partir de este momento como Π_n , se dice intercambiable si su distribución es invariante ante permutaciones de $[n]$, i.e., para cada partición $\{A_1, \dots, A_k\}$ de $[n]$ se tiene que

$$\mathbb{P}(\Pi_n = \{A_1, \dots, A_k\}) = p(|A_1|, \dots, |A_k|)$$

para alguna función simétrica p de las composiciones (n_1, \dots, n_k) de $[n]$. Esta función p se le conoce como función de probabilidad para particiones intercambiables o EPPF por sus siglas en inglés y será denotada por $\Pi_k^{(n)}(n_1, \dots, n_k)$. De esta manera se puede dar la siguiente definición involucrando todo lo anterior.

Definición 2.7. *Sea $\{X_i\}_{i \geq 1}$ una secuencia intercambiable. Entonces, considerando a $\Pi_k^{(n)}$ como arriba, la colección $\{\Pi_k^{(n)} : 1 \leq k \leq n, n \geq 1\}$ define una EPPF.*

Una característica esencial de las EPPF es que satisface la regla de adición, i.e.,

$$\Pi_k^{(n)}(n_1, \dots, n_k) = \Pi_{k+1}^{(n+1)}(n_1, \dots, n_k, 1) + \sum_{j=1}^k \Pi_k^{(n)}(n_1, \dots, n_{j-1}, n_j + 1, n_{j+1}, \dots, n_k).$$

De esta manera es sencillo ver cómo el estudio de las NRMI está ciertamente ligado con el estudio de las particiones intercambiables y su función de probabilidad asociada, EPPF. Se ha visto que dada una muestra $\mathbf{X}^{(n)}$, ésta admite una representación $(\mathbf{Y}^{(n)}, \pi)$, donde $\mathbf{Y}^{(n)}$ denota las distintas observaciones y π es una partición de $[n] = \{1, \dots, n\}$. Por lo que puede haber empates y esto se obtiene por el hecho de que Φ es discreta casi seguramente. Con lo anterior en mente se puede definir una partición de $[n]$ mediante la siguiente asignación:

$$i \sim j \Leftrightarrow X_i = X_j \quad \forall i, j \in \{1, \dots, n\}, i \neq j.$$

Otra de las ventajas de contar con la EPPF es que permite obtener el sistema de distribuciones predictivas inducidas por \mathcal{Q} , ya que teniendo una muestra de tamaño n y $n(\pi) = k$ distintas observaciones Y_1, \dots, Y_k se cumple que

$$\begin{aligned} \mathbb{P}(X_{n+1} \neq Y_j \mid \forall j \in \{1, \dots, k\} | \mathbf{X}^{(n)}) &= \frac{\Pi_{k+1}^{(n+1)}(n_1, \dots, n_k, 1)}{\Pi_k^{(n)}(n_1, \dots, n_k)} \\ \mathbb{P}(X_{n+1} = Y_j | \mathbf{X}^{(n)}) &= \frac{\Pi_k^{(n)}(n_1, \dots, n_{j-1}, n_j + 1, n_{j+1}, \dots, n_k)}{\Pi_k^{(n)}(n_1, \dots, n_k)}. \end{aligned}$$

Donde no es difícil de mostrar, a partir de la ecuación (2.8) que

$$\Pi_k^{(n)}(n_1, \dots, n_k) = \frac{1}{\Gamma(n)} \int_{\mathbb{R}_+} u^{n-1} e^{-\varphi(u)} \left[\prod_{i=1}^k \int_{\mathbb{X}} \tau_{n_i}(u|x) \alpha(dx) \right] du. \quad (2.11)$$

Más aún, con la expresión de la EPPF dada por (2.11) se recupera la distribución predictiva obtenida en la sección anterior, ecuación (2.9). Es importante señalar que sólo para ciertos casos se tienen expresiones cerradas, siendo la NRMI basada en la normalización del proceso gama generalizado uno de estos casos. El tener una expresión cerrada nos permite implementar algoritmos de muestreo de Blackwell-MacQueen. Sin embargo, para cuando sea necesario aún se puede muestrear de la distribución predictiva mediante el condicionamiento de la variable latente U_n .

La idea detrás de condicionar con la variable latente U_n es poder expresar a la EPPF como

$$\Pi_k^{(n)} = V_{n,k} \prod_{i=1}^k W_{n_i}.$$

Donde $V_{n,k}$ es una cantidad positiva y que no depende de la composición (n_1, \dots, n_k) y W_{n_i} es una cantidad positiva que sólo depende de n_i . Si es posible encontrar una EPPF con esta descomposición se dice que la partición aleatoria es una partición de Gibbs. De esta manera y de nueva cuenta a través de la ecuación (2.8) se puede verificar que condicionando a U_n

$$\Pi^{(n)}(\pi|u) = \frac{e^{-\varphi(u)} \prod_{i=1}^{n(\pi)} \kappa_{n_i}(u)}{\int_{\mathbb{R}_+} t^n e^{-ut} f_T(t) dt}, \quad (2.12)$$

donde,

$$\kappa_{n_i}(u) = \int_{\mathbb{X}} \tau_{n_i}(u|x) \alpha(dx).$$

Por lo que la partición π condicionada a U_n es una partición de Gibbs. Y de esta manera, se tiene que la distribución predictiva condicionada a la variable latente es

$$\mathbb{P}(X_{n+1} \in dx | \mathbf{X}^{(n)}, U_n = u) \propto \kappa_1(u) \tau_1(u|x) P_0(dx) + \sum_{j=1}^k \frac{\tau_{n_j+1}(u|Y_j)}{\tau_{n_j}(u|Y_j)} \delta_{Y_j}(dx). \quad (2.13)$$

Con la representación anterior se puede implementar un esquema de muestreo como el de Blackwell-MacQueen para obtener una muestra. Para esto se debe definir a

$$m(dx|u) \propto \tau_1(u|x)\alpha(dx)$$

$$m(dx_i|x_1, \dots, x_{i-1}, u) = \mathbb{P}(X_i \in dx_i|X_1, \dots, X_{i-1}, U_{i-1} = u) \quad i \geq 2.$$

Y U_0 una variable aleatoria positiva con densidad $q_0(u) \propto \theta e^{-\varphi(u)} \int_{\mathbb{X}} \tau_1(u|x)P_0(dx)$. De esta manera el algoritmo queda descrito como:

1. Simular U_0 de q_0 .
2. Simular X_1 de $m(dx|U_0)$.
3. En el paso i
 - a. Simular U_{i-1} de $f_{U_{i-1}|X_1, \dots, X_{i-1}}(u)$.
 - b. Generar ϵ_i de $f_i(\epsilon) \propto \tau_1(U_{i-1}|\epsilon)P_0(d\epsilon)$.
 - c. Simular X_i de $m(dx|X_1, \dots, X_{i-1}, U_{i-1})$, con lo cual está ocurriendo lo siguiente:

$$X_i = \begin{cases} \epsilon_i & \propto \kappa_1(U_{i-1}) \\ Y_{j,i-1} & \propto \tau_{n_{j,i-1}+1}(U_{i-1}|Y_{j,i-1})/\tau_{n_{j,i-1}}(U_{i-1}|Y_{j,i-1}). \end{cases}$$

Donde $Y_{j,i-1}$ es el j -ésimo valor distinto habiendo muestreado $i - 1$ valores y con $n_{j,i-1}$ el número de empates.

Para este punto del trabajo se ha visto que las NRMI son discretas casi seguramente, por lo cual hay probabilidad positiva de empates, de esta manera se sabe que al tener n observaciones se tendrán $n(\pi)$ distintos grupos de observaciones. Con lo cual surge de manera muy intuitiva el estudio de conglomerados en la estructura de los datos y estimación de densidades en modelos de mezclas. Lo cual es posible ya que ya se cuenta con un esquema de muestreo.

2.5. Estimación de mezclas

Este método fue propuesto por primera vez en 1984 por Lo. La idea principal consiste en la introducción de una secuencia intercambiable de variables $\{\beta_i\}_{i \geq 1}$, gobernadas por una medida de probabilidad completamente aleatoria discreta \mathcal{P} y con valores en un espacio polaco \mathbb{B} , que está dotado de su σ -álgebra de Borel y un kernel k . Donde k es una aplicación medible del espacio producto $\mathbb{X} \times \mathbb{B}$ a \mathbb{R}_+ y es tal que para λ una medida σ -finita en \mathbb{X} se tiene que la aplicación

$$C \mapsto \int_C k(x, \beta) \lambda(dx)$$

define una medida de probabilidad en \mathbb{X} para toda $\beta \in \mathbb{B}$. En la práctica un kernel sumamente utilizado es el gaussiano con vector de parámetros $\beta = (\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}_+$. De esta manera se puede construir un modelo jerárquico,

$$\begin{aligned} X_i | \beta_i, \mathcal{P} &\stackrel{\text{iid}}{\sim} k(X_i, \beta_i) \\ \beta_i | \mathcal{P} &\stackrel{\text{iid}}{\sim} \mathcal{P} \\ \mathcal{P} &\sim \mathcal{Q}. \end{aligned}$$

Si se define a una función de densidad aleatoria dada por

$$x \mapsto f_{\mathcal{P}}(x) = \int_{\mathbb{B}} k(x, \beta) \mathcal{P}(d\beta).$$

El modelo jerárquico puede ser escrito como

$$\begin{aligned} X_i | \mathcal{P} &\stackrel{\text{iid}}{\sim} f_{\mathcal{P}} \\ \mathcal{P} &\sim \mathcal{Q}. \end{aligned}$$

Al ser \mathcal{P} discreta, propicia que haya empates en las variables latentes $\{\beta_i\}_{i \geq 1}$, lo que se verá reflejado en una partición de las observaciones. De esta manera si se denota a $\beta_1^*, \dots, \beta_k^*$ como las distintas observaciones en una muestra de tamaño n y se define a $C_j = \{i : \beta_i = \beta_j^*\}$, entonces cualesquiera dos observaciones r y s estarán en el mismo grupo C_j si y sólo si $\beta_r = \beta_s$. El número de distintas observaciones permite identificar el número de conglomerados que se denotará como K_n .

El tratar de esta manera el problema de mezclas de densidades tiene como objetivo hacer inferencia sobre \mathcal{P} y sobre K_n condicionado a la muestra. Sin embargo, obtener expresiones cerradas no es posible y un claro ejemplo de esto último se puede observar al momento de dar un posible estimador de $f_{\mathcal{P}}$, dado por

$$\begin{aligned} \mathbb{E}(f_{\mathcal{P}}|\mathbf{X}^{(n)}) &= \int_{\mathbb{B}} k(x, \beta) \mathbb{E}(\mathcal{P}(d\beta)|\mathbf{X}^{(n)}) \\ &= \int_{\mathbb{B}} k(x, \beta) \int_{\mathbb{B}^n} \mathbb{E}(\mathcal{P}(d\beta)|\mathbf{X}^{(n)}, \beta^{(n)}) \mathcal{Q}(d\beta_1^*, \dots, d\beta_{n(\pi)}^*|\mathbf{X}^{(n)}) \\ &= \sum_{k=1}^n \int_{\mathbb{B}^{k+1}} k(x, \beta) \sum_{\pi \in \mathcal{P}_{[n]}^k} \mathbb{E}(\mathcal{P}(d\beta)|\beta_1^*, \dots, \beta_{n(\pi)}^*) \mathcal{Q}(d\beta_1^*, \dots, d\beta_{n(\pi)}^*|\mathbf{X}^{(n)}). \end{aligned}$$

La principal dificultad al evaluar esta expresión es el tener que hacer la suma sobre todas las particiones, ya que la distribución predictiva $\mathbb{E}(\mathcal{P}(d\beta)|\beta_1^*, \dots, \beta_{n(\pi)}^*)$ se puede tener de forma cerrada para algunas NRMI. Aunado a lo anterior es necesario usar métodos de Monte Carlo vía Cadenas de Markov (MCMC) para generar una muestra de variables latentes provenientes de la distribución posterior, $\mathcal{Q}(d\beta_1^*, \dots, d\beta_{n(\pi)}^*|\mathbf{X}^{(n)})$.

Para dar solución a las dificultades que se presentan con esta metodología se han descrito varios algoritmos, siendo uno de los principales el obtenido por Escobar (1988,1994) y por Escobar y West (1995). Este algoritmo fue descrito para cuando se utiliza el proceso Dirichlet, pero que se puede extender para cualquier medida \mathcal{P} tal que la EPPF asociada o equivalentemente el sistema de distribuciones predictivas se conozca de forma explícita.

Procediendo de manera similar a como se ha venido trabajando se tiene que

$$\mathcal{Q}(d\beta_1^*, \dots, d\beta_{n(\pi)}^* | \mathbf{X}^{(n)}) \propto \Pi_k^{(n)}(n_1, \dots, n_k) \prod_{j=1}^k P_0(d\beta_j^*) \prod_{i \in \mathcal{C}_j} k(X_i, \beta_j^*).$$

Para muestrear de esta distribución, se utilizan las condicionales de las variables latentes, i.e.,

$$\mathbb{P}(\beta_i \in d\beta_i | \beta_{-i}, \mathbf{X}^{(n)}) = q_{i,0}^* P_{0,i}(d\beta_i) + \sum_{j=1}^{k_{i,n-1}} q_{i,j}^* \delta_{\beta_{i,j}^*}(d\beta_i).$$

Donde $\beta_{-i} = (\beta_1, \dots, \beta_{i-1}, \beta_{i+1}, \dots, \beta_n)$, $k_{i,n-1}$ es el número de valores distintos $\beta_{i,j}^*$ en β_{-i} , $n_{i,j}$ es la frecuencia de $\beta_{i,j}^*$ en β_{-i} y

$$\begin{aligned} P_{0,i}(d\beta_i) &\propto k(X_i, \beta_i) P_0(d\beta_i) \\ q_{i,0}^* &\propto \Pi_{k_{i,n-1}+1}^{(n)}(n_{i,1}, \dots, n_{i,k_{i,n-1}}, 1) \int_{\mathbb{B}} k(X_i, \beta) P_0(d\beta) \\ q_{i,j}^* &\propto \Pi_{k_{i,n-1}}^{(n)}(n_{i,1}, \dots, n_{i,j} + 1, \dots, n_{i,k_{i,n-1}}) k(X_i, \beta_{i,j}^*). \end{aligned}$$

A partir de los pesos anteriores es fácil darse cuenta el por qué se necesita tener expresiones cerradas para la EPPF, situación que se verá en el capítulo siguiente cuando se introduzcan los ejemplos más conocidos y tratados en la literatura. De esta manera se puede implementar el siguiente algoritmo de MCMC.

1. Simular valores iniciales $\beta_1^{(0)}, \dots, \beta_n^{(0)}$ de P_0 .
2. En el paso i simular $\beta_1^{(i)}, \dots, \beta_n^{(i)}$ como,

$$\begin{aligned} \beta_1^{(i)} &\sim \mathbb{P}(d\beta_1 | \beta_2^{(i-1)}, \dots, \beta_n^{(i-1)}, \mathbf{X}^{(n)}) \\ \beta_2^{(i)} &\sim \mathbb{P}(d\beta_2 | \beta_1^{(i)}, \beta_3^{(i-1)}, \dots, \beta_n^{(i-1)}, \mathbf{X}^{(n)}) \\ &\vdots \\ \beta_n^{(i)} &\sim \mathbb{P}(d\beta_n | \beta_1^{(i)}, \dots, \beta_{n-1}^{(i)}, \mathbf{X}^{(n)}) \end{aligned}$$

En la iteración j se tiene $k^{(j)}$ el número de valores distintos, $\beta_{1,j}^*, \dots, \beta_{k^{(j)},j}^*$.

Después un cierto número apropiado de iteraciones, digamos N , se pueden dar los siguientes estimadores para $f_{\mathcal{P}}$ y para la distribución posterior del número de conglomerados.

$$\tilde{f}_{\mathcal{P}}(x) \approx \frac{1}{N} \sum_{i=1}^N \int_{\mathbb{B}} k(x, \beta) \mathbb{E}(\mathcal{P}(d\beta) | \beta_{1,i}^*, \dots, \beta_{k^{(i)},i}^*)$$

$$\mathbb{P}(K_n = k | \mathbf{X}^{(n)}) \approx \frac{1}{N} \sum_{i=1}^N \mathbb{1}_{k^{(i)}=k}.$$

Una posible complicación en este esquema de Gibbs Sampler ocurre cuando los pesos $q_{i,j}^*$ son mucho más grandes que $q_{i,0}^*$, problema que ha sido estudiado y tratado en los trabajos de MacEachern (1994), Busch y MacEachern (1996) e Ishwaran y James (2001) y al cual dan como una posible solución la introducción de un paso adicional; el cual consiste en una vez teniendo las $k^{(j)}$ distintas observaciones, proceder con un remuestreo de los valores de dichas observaciones, i.e., dado $k^{(j)}$ y el vector $\beta^{(j)} = (\beta_{1,j}^*, \dots, \beta_{k^{(j)},j}^*)$ se debe muestrear de la distribución

$$\mathbb{P}(\beta_{i,j}^* \in d\beta_1, \dots, \beta_{k^{(j)},j}^* \in d\beta_{k^{(j)}} | \mathbf{X}^{(n)}, \beta^{(j)}) \propto \prod_{i=1}^{k^{(j)}} \prod_{l \in C_{i,j}} k(X_l, \beta_i) P_0(d\beta_i)$$

Capítulo 3

Medidas de probabilidad aleatorias vía las CRM más utilizadas

A lo largo de este capítulo se explorarán las medidas de probabilidad aleatorias generadas por los procesos de normalización más conocidos, para los cuales se tienen expresiones cerradas para la función de probabilidad para particiones aleatorias (EPPF), lo que permite a su vez obtener la distribución predictiva e implementar los algoritmos vistos en la sección anterior.

A partir de este momento se considerará a $\mathbb{X} = \mathbb{R}$, el cual es claramente un espacio polaco y que estará dotado de su σ -álgebra de Borel denotada por $\mathcal{B}(\mathbb{R})$, por lo cual el problema se interpreta como la construcción de medidas de probabilidad en la recta real y que permite establecer con mayor claridad todas las características posteriores desarrolladas en la sección anterior.

El capítulo inicia con un proceso que ya ha sido mencionado a lo largo de este trabajo y que representó el inicio de la estadística bayesiana no paramétrica. El proceso Dirichlet es una NRMI que cuenta con la propiedad de conjugación, lo cual propicia que los cálculos sean más fáciles de tratar a comparación de otras NRMI. Sin embargo, no siempre modela con total fidelidad el problema a estudiar por lo cual se han tenido que utilizar otros procesos como el σ -estable, el gama generalizado, entre otros.

Estos procesos con la ayuda de la variable latente U_n descrita en el capítulo anterior pueden ser manejados y son algunos de los cuales se pueden obtener expresiones cerradas para la EPPF y por lo tanto para los que se pueden implementar algoritmos similares a los del proceso Dirichlet.

3.1. Proceso Dirichlet

Antes de introducir la construcción vía normalización del proceso Dirichlet, se hará una breve introducción a la distribución Dirichlet. Esta distribución es conocida en la rama de la estadística bayesiana paramétrica, por ser conjugada para la distribución multinomial.

Para su construcción se consideran a Z_1, \dots, Z_n variables aleatorias independiente, tales que $Z_i \sim \text{Gama}(\alpha_i, 1)$ con $\alpha_i > 0$ (otra posible construcción permite que haya variables degeneradas, para lo cual se definen a los parámetros $\alpha_i \geq 0$ y que para algún índice j se tenga que $\alpha_j > 0$) y se define a

$$Y_j := \frac{Z_j}{\sum_{i=1}^n Z_i} \quad j = 1, 2, \dots, n.$$

Entonces se tiene que el vector aleatorio (Y_1, \dots, Y_n) tiene distribución Dirichlet con vector de parámetros $(\alpha_1, \dots, \alpha_n)$ y donde $\sum_{i=1}^n Y_i = 1$. Dado que se está tomando a $\alpha_i > 0$ para toda $i \in \{1, \dots, n\}$, se sigue que el vector es absolutamente continuo con respecto a la medida de Lebesgue, por lo cual acepta una densidad dada por

$$f(y_1, \dots, y_{n-1}) = \frac{\Gamma(\sum_{i=1}^n \alpha_i)}{\prod_{i=1}^n \Gamma(\alpha_i)} \left(\prod_{i=1}^{n-1} y_i^{\alpha_i-1} \right) \left(1 - \sum_{i=1}^{n-1} y_i \right)^{\alpha_n-1} \mathbb{1}_{\Delta}(y_1, \dots, y_{n-1}).$$

Donde Δ denota el simplejo de dimensión $n - 1$, i.e.,

$$\Delta = \left\{ (y_1, \dots, y_{n-1}) \mid y_i \geq 0, \sum_{i=1}^{n-1} y_i \leq 1 \right\}.$$

Si un vector (Y_1, \dots, Y_n) tiene distribución Dirichlet, que se denotará como $\mathcal{D}(\alpha_1, \dots, \alpha_n)$, los momentos de orden 1 y 2 son

$$\begin{aligned}\mathbb{E}(Y_i) &= \frac{\alpha_i}{\alpha} \\ \mathbb{E}(Y_i^2) &= \frac{\alpha_i(\alpha_i + 1)}{\alpha(\alpha + 1)} \\ \mathbb{E}(Y_i Y_j) &= \frac{\alpha_i \alpha_j}{\alpha(\alpha + 1)} \quad i \neq j.\end{aligned}$$

Con estas expresiones para los momentos, se pueden obtener la varianza y la covarianza y donde $\alpha = \sum_{i=1}^n \alpha_i$. La idea de introducir la distribución antes del proceso es que su construcción además de la proveniente mediante la normalización de un subordinador, puede hacerse a través de las distribuciones infinito dimensionales. Ferguson (1973) demuestra que al tomar una medida finita y no nula α sobre un espacio polaco $(\mathbb{X}, \mathcal{X})$, entonces una medida de probabilidad aleatoria \mathcal{P} con valores en el espacio de todas las medidas sobre \mathbb{X} tiene una distribución Proceso Dirichlet, (\mathcal{D}_α) , si para toda partición medible (A_1, \dots, A_n) se tiene que

$$(\mathcal{P}(A_1), \dots, \mathcal{P}(A_n)) \sim \mathcal{D}(\alpha(A_1), \dots, \alpha(A_n)).$$

Una construcción que se puede checar más a fondo en el trabajo de Ferguson. El otro método es a través de la normalización del proceso gama y que también es tratado en el artículo de Ferguson y que puede encontrarse en gran parte de la literatura concerniente con la estadística bayesiana no paramétrica. Para esto se debe considerar a la siguiente medida de Lévy,

$$\nu(ds, dx) = \rho(ds)\alpha(dx) = \frac{e^{-s}}{s} ds \alpha(dx).$$

Se puede observar que es una medida homogénea y donde α será una medida no nula y finita y que por ende se sabe que tiene la siguiente representación $\alpha(dx) = \theta P_0(dx)$. La medida α funcionará como un parámetro. Para poder garantizar que sea una medida apropiada debe cumplir ciertas características.

Primero y como ya se vio en el Capítulo 1, se tiene que $\rho(\mathbb{R}_+) = \infty$ lo cual garantiza que ν es una medida de actividad infinita. También debe cumplir con una condición de integrabilidad dada por la siguiente ecuación

$$\int_B \int_{\mathbb{R}_+} (s \wedge 1) \nu(ds, dx) < \infty \quad (3.1)$$

Como ν es homogénea y α será elegida de tal forma que sea finita, la siguiente condición se cumple si y sólo si $\int_{\mathbb{R}_+} (s \wedge 1) \rho(ds) < \infty$, pero esto es cierto ya que

$$\int_{\mathbb{R}_+} (s \wedge 1) \rho(ds) = \int_0^1 e^{-s} ds + \int_1^\infty \frac{e^{-s}}{s} ds < \infty$$

La última condición tiene que ver con el exponente laplaciano, $\varphi(u)$, que se le pide que sea finito, para esto primero sea f una función medible de $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ a $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ tal que $\int |f| d\Phi < \infty$ y calculando la funcional de Laplace,

$$\int_{\mathbb{R} \times \mathbb{R}_+} (1 - e^{-sf(x)}) \nu(ds, dx) = \int_{\mathbb{R}} \left[\int_{\mathbb{R}_+} (1 - e^{-sf(x)}) \frac{e^{-s}}{s} ds \right] \alpha(dx).$$

Donde

$$\begin{aligned} \int_{\mathbb{R}_+} (1 - e^{-sf(x)}) \frac{e^{-s}}{s} ds &= \int_{\mathbb{R}_+} \left[\sum_{j=1}^{\infty} \frac{(-1)^{j+1} (sf(x))^j}{j!} \right] \frac{e^{-s}}{s} ds \\ &= \sum_{j=1}^{\infty} \frac{(-1)^{j+1} f(x)^j}{j!} \int_{\mathbb{R}_+} s^{j-1} e^{-s} ds \\ &= \sum_{j=1}^{\infty} \frac{(-1)^{j+1} f(x)^j}{j!} \Gamma(j) \\ &= \sum_{j=1}^{\infty} \frac{(-1)^{j+1} f(x)^j}{j} \\ &= \log(1 + f(x)). \end{aligned}$$

De esta manera se tiene que el exponente laplaciano es

$$\varphi(u) = \int_{\mathbb{R}} \log(1+u)\alpha(dx) = \theta \log(1+u) < \infty$$

Más aún, de la expresión (2.2) se sigue que

$$\mathbb{E} \left[\exp \left(- \int_{\mathbb{R}} f(x)\Phi(dx) \right) \right] = \exp \left[- \int_{\mathbb{R}} \log(1+f(x))\alpha(dx) \right]. \quad (3.2)$$

Que está bien definida para funciones f , tal que $\int \log(1+f)\alpha < \infty$. Si se considera a $f(x) = \lambda \mathbb{1}_A$ para alguna $A \in \mathcal{B}(\mathbb{R})$ y $\lambda > 0$ se tiene que

$$\begin{aligned} \mathbb{E}(e^{-\lambda\Phi(A)}) &= \exp \left[- \int_{\mathbb{R}} \log(1+f(x))\alpha(dx) \right] \\ &= \exp \left[-\theta \int_{\mathbb{R}} \log(1+\lambda \mathbb{1}_A)P_0(dx) \right] \\ &= \exp[-\theta P_0(A) \log(1+\lambda)] \\ &= (1+\lambda)^{-\theta P_0(A)}. \end{aligned}$$

Por lo que $\Phi(A) \sim \text{gama}(\theta P_0(A), 1)$. Y de esta manera dada una partición medible A_1, \dots, A_n de \mathbb{R} se tiene que la colección de variables aleatorias $\Phi(A_1), \dots, \Phi(A_n)$ son independientes y cada una con distribución gama de parámetros $(\theta P_0(A_i), 1)$, por lo que

$$\mathcal{P}(A_i) = \frac{\Phi(A_i)}{\Phi(\mathbb{R})} = \frac{\Phi(A_i)}{\sum_{j=1}^n \Phi(A_j)}.$$

Que es la construcción dada para la distribución Dirichlet, por lo tanto, para toda partición medible se tiene que

$$(\mathcal{P}(A_1), \dots, \mathcal{P}(A_n)) \sim \mathcal{D}(\theta P_0(A_1), \dots, \theta P_0(A_n)).$$

Que es justo lo que se estaba buscando. Ahora bien, como ya se mencionó anteriormente, si $\mathcal{P} \sim \mathcal{D}_{\theta P_0}$ entonces se va tener un modelo conjugado por lo cual la distribución posterior también caerá dentro de la misma familia pero con una actualización en el parámetro. Para ver esto, sólo basta recordar el Teorema 2.3. por el cual se tiene que $\Phi^{(U_n)}$ tendrá una medida de Lévy

$$\nu^{(U_n)}(ds, dx) = e^{-U_n s} \rho(ds) \theta P_0(dx) = \frac{e^{-(1+U_n)s}}{s} ds \theta P_0(dx).$$

Y los saltos $J_i^{(U_n, \mathbf{X}^{(n)})}$ tienen densidad proporcional a

$$s^{n_i} e^{-s U_n} \rho_{Y_i}(ds) = s^{n_i-1} e^{-(1+U_n)s} ds.$$

Por lo cual $J_i^{(U_n, \mathbf{X}^{(n)})} \sim \text{gama}(n_i, 1 + U_n)$. Con lo anterior en mente se puede obtener la transformada de Laplace para la medida posterior, la cual tendrá la siguiente expresión

$$\begin{aligned} & \mathbb{E} \left[\exp \left(- \int_{\mathbb{R}} f(x) \Phi(dx) \right) \middle| U_n, \mathbf{X}^{(n)} \right] \\ &= \mathbb{E} \left[\exp \left(- \int_{\mathbb{R}} f(x) \left\{ \Phi^{(U_n)}(dx) + \sum_{i=1}^{n(\pi)} J_i^{(U_n, \mathbf{X}^{(n)})} \delta_{Y_i}(dx) \right\} \right) \right] \\ &= \mathbb{E} \left[\exp \left(- \int_{\mathbb{R}} f(x) \Phi^{(U_n)}(dx) \right) \right] \mathbb{E} \left[\exp \left(- \int_{\mathbb{R}} f(x) \left(\sum_{i=1}^{n(\pi)} J_i^{(U_n, \mathbf{X}^{(n)})} \delta_{Y_i}(dx) \right) \right) \right] \end{aligned} \quad (3.3)$$

El primer término de la ecuación (3.3) se resuelve de forma análoga a lo que se hizo arriba para calcular la transformada de Laplace de Φ , por lo cual se sigue de manera inmediata que

$$\mathbb{E} \left[\exp \left(- \int_{\mathbb{R}} f(x) \Phi^{(U_n)}(dx) \right) \right] = \exp \left[-(\theta + n) \int_{\mathbb{R}} \log \left(1 + \frac{f(x)}{1 + U_n} \right) \frac{\theta P_0(dx)}{\theta + n} \right].$$

Para el segundo término se utilizará el hecho de que se conoce de manera exacta la distribución de los saltos, por lo que

$$\begin{aligned}
\mathbb{E} \left[\exp \left(- \int_{\mathbb{R}} f(x) \left(\sum_{i=1}^{n(\pi)} J_i^{(U_n, \mathbf{X}^{(n)})} \delta_{Y_i}(dx) \right) \right) \right] &= \mathbb{E} \left[\prod_{i=1}^{n(\pi)} \exp \left(- J_i^{(U_n, \mathbf{X}^{(n)})} f(Y_i) \right) \right] \\
&= \prod_{i=1}^{n(\pi)} \left(1 + \frac{f(Y_i)}{1 + U_n} \right)^{n_i} \\
&= \exp \left(- \sum_{i=1}^{n(\pi)} n_i \log \left(1 + \frac{f(Y_i)}{1 + U_n} \right) \right)
\end{aligned}$$

Utilizando el hecho de que $n_i = \sum_{j=1}^n \delta_{X_j}(\{Y_i\})$, se tiene

$$\begin{aligned}
\exp \left(- \sum_{i=1}^{n(\pi)} n_i \log \left(1 + \frac{f(Y_i)}{1 + U_n} \right) \right) &= \exp \left(- \sum_{j=1}^n \int_{\mathbb{R}} \log \left(1 + \frac{f(x)}{1 + U_n} \right) \delta_{X_j}(dx) \right) \\
&= \exp \left(-(\theta + n) \int_{\mathbb{R}} \log \left(1 + \frac{f(x)}{1 + U_n} \right) \frac{\sum_{j=1}^n \delta_{X_j}(dx)}{(\theta + n)} \right).
\end{aligned}$$

Juntando las últimas expresiones se tiene que

$$\mathbb{E} \left[\exp \left(- \int_{\mathbb{R}} f(x) \Phi(dx) \right) \middle| U_n, \mathbf{X}^{(n)} \right] = \exp \left(-(\theta + n) \int_{\mathbb{R}} \log \left(1 + \frac{f(x)}{1 + U_n} \right) P_{0,n}(dx) \right).$$

Donde $P_{0,n} = \frac{\theta}{\theta+n} P_0 + \frac{\sum_{j=1}^n \delta_{X_j}(dx)}{(\theta+n)}$, por lo que $\Phi^{(U_n, \mathbf{X}^{(n)})}$ es una CRM gama con intensidad dada por

$$\nu_n(ds, dx) = \frac{e^{-(1+U_n)s}}{s} ds (\theta + n) P_{0,n}(dx).$$

Más aún, se puede observar que al tomar a $f(x) = \lambda \mathbb{1}_B$ para $B \in \mathcal{B}(\mathbb{R})$ se tiene que

$$\mathbb{E} \left[\exp(-\lambda \Phi(B)) \middle| U_n, \mathbf{X}^{(n)} \right] = \left(1 + \frac{f(x)}{1 + U_n} \right)^{-(\theta+n)P_{0,n}(B)}.$$

De esta manera $\Phi^{(U_n, \mathbf{X}^{(n)})}(B) \sim \text{gama}((\theta + n)P_{0,n}(B), 1 + U_n)$, por lo que se puede fijar a $1 + U_n = 1$ y así, al momento de normalizar como se hizo para Φ se obtiene un proceso Dirichlet de parámetro $((\theta + n)P_{0,n})$.

Ahora que ya se tiene caracterizada la distribución posterior, se puede proceder con la distribución predictiva, para la cual se necesitan los siguientes elementos.

$$\begin{aligned} e^{-\varphi(u)} &= \exp \left[- \int_{\mathbb{R}} \int_{\mathbb{R}_+} (1 - e^{-us}) \frac{e^{-s}}{s} ds \theta P_0(dx) \right] \\ &= \exp \left[- \int_{\mathbb{R}} \log(1 + u) \theta P_0(dx) \right] \\ &= (1 + u)^{-\theta}. \end{aligned}$$

También se requiere una expresión para $\tau_{n_i}(u|Y_i)$ que estará dada por,

$$\begin{aligned} \tau_{n_i}(u|Y_i) &= \int_{\mathbb{R}_+} s^{n_i} e^{-us} \frac{e^{-s}}{s} ds \\ &= \int_{\mathbb{R}_+} s^{n_i-1} e^{-(1+u)s} ds \\ &= \frac{\Gamma(n_i)}{(1 + u)^{n_i}}. \end{aligned}$$

Finalmente, la distribución posterior de la variable latente dada la muestra,

$$\begin{aligned} f_{U_n|\mathbf{X}^{(n)}}(u) &\propto u^{n-1} \prod_{i=1}^{n(\pi)} \tau_{n_i}(u|Y_i) e^{-\varphi(u)} \\ &= \frac{u^{n-1}}{(1 + u)^\theta} \prod_{i=1}^{n(\pi)} \frac{\Gamma(n_i)}{(1 + u)^{n_i}} \\ &\propto \frac{u^{n-1}}{(1 + u)^{n+\theta}}. \end{aligned}$$

Y donde

$$\int_{\mathbb{R}_+} \frac{u^{n-1}}{(1 + u)^{n+\theta}} du = \frac{\Gamma(n)\Gamma(\theta)}{\Gamma(n + \theta)}.$$

Por lo que

$$f_{U_n|\mathbf{X}^{(n)}}(u) = \frac{\Gamma(n+\theta)}{\Gamma(n)\Gamma(\theta)} \cdot \frac{u^{n-1}}{(1+u)^{n+\theta}}.$$

Con las expresiones anteriores se puede obtener los pesos $\omega^{(n)}$ y $\omega_i^{(n)}$, que permiten dar una expresión cerrada para la distribución predictiva.

$$\begin{aligned} \omega^{(n)} &= \frac{\theta}{n} \int_{\mathbb{R}_+} u \tau_1(u|x) f_{U_n|\mathbf{X}^{(n)}}(u) du \\ &= \frac{\theta}{n} \int_{\mathbb{R}_+} \frac{u}{1+u} \frac{\Gamma(n+\theta)}{\Gamma(n)\Gamma(\theta)} \cdot \frac{u^{n-1}}{(1+u)^{n+\theta}} \\ &= \frac{\theta}{n} \cdot \frac{\Gamma(n+\theta)}{\Gamma(n)\Gamma(\theta)} \int_{\mathbb{R}_+} \frac{u^n}{(1+u)^{n+\theta+1}} du \\ &= \frac{\theta}{n} \cdot \frac{\Gamma(n+\theta)}{\Gamma(n)\Gamma(\theta)} \cdot \frac{\Gamma(n+1)\Gamma(\theta)}{\Gamma(n+\theta+1)} \\ &= \frac{\theta}{n+\theta}. \end{aligned}$$

De forma análoga se obtiene que

$$\omega_i^{(n)} = \int_{\mathbb{R}_+} u \frac{\tau_{n_i+1}(u|Y_i)}{\tau_{n_i}(u|Y_i)} f_{U_n|\mathbf{X}^{(n)}}(u) du = \frac{n \cdot n_i}{n+\theta}.$$

Por lo que la distribución predictiva tiene la siguiente forma

$$\begin{aligned} \mathbb{P}(X_{n+1} \in dx | \mathbf{X}^{(n)}) &= \omega^{(n)} P_0(dx) + \frac{1}{n} \sum_{i=1}^{n(\pi)} \omega_i^{(n)} \delta_{Y_i}(dx) \\ &= \frac{\theta}{n+\theta} P_0(dx) + \frac{1}{n+\theta} \sum_{i=1}^{n(\pi)} n_i \delta_{Y_i}(dx) \end{aligned}$$

Lo cual dice lo siguiente sobre la observación $n+1$,

$$\begin{aligned} \mathbb{P}(X_{n+1} \neq Y_j \quad \forall j \in \{1, \dots, n(\pi)\} | \mathbf{X}^{(n)}) &= \frac{\theta}{n+\theta} \\ \mathbb{P}(X_{n+1} = Y_j \text{ p.a. } j \in \{1, \dots, n(\pi)\} | \mathbf{X}^{(n)}) &= \frac{n_j}{n+\theta}. \end{aligned}$$

Por lo cual se recuperan las siguientes expresiones sobre la EPPF asociada, ya que se cumple lo siguiente

$$\frac{\Pi_{k+1}^{(n)}(n_1, \dots, n_k, 1)}{\Pi_k^{(n)}(n_1, \dots, n_k)} = \frac{\theta}{n + \theta}$$

$$\frac{\Pi_k^{(n)}(n_1, \dots, n_{j-1}, n_j + 1, n_{j+1}, \dots, n_k)}{\Pi_k^{(n)}(n_1, \dots, n_k)} = \frac{n_j}{n + \theta}.$$

Para el proceso Dirichlet no es difícil encontrar una expresión para la EPPF, ya que está definida como

$$\begin{aligned} \Pi_k^{(n)}(n_1, \dots, n_k) &= \frac{1}{\Gamma(n)} \int_{\mathbb{R}_+} u^{n-1} e^{-\varphi(u)} \left(\prod_{i=1}^k \int_{\mathbb{R}} \tau_{n_i}(u|x) \theta P_0(dx) \right) du \\ &= \frac{1}{\Gamma(n)} \int_{\mathbb{R}_+} u^{n-1} (1+u)^{-\theta} \left(\prod_{i=1}^k \int_{\mathbb{R}} \frac{\Gamma(n_i)}{(1+u)^{n_i}} \theta P_0(dx) \right) du \\ &= \frac{\theta^k}{\Gamma(n)} \prod_{i=1}^k \Gamma(n_i) \int_{\mathbb{R}_+} \frac{u^{n-1}}{(1+u)^{n+\theta}} du \\ &= \frac{\theta^k \Gamma(\theta)}{\Gamma(n+\theta)} \prod_{i=1}^k \Gamma(n_i) \\ &= \frac{\theta^k}{(\theta)_n} \prod_{i=1}^k \Gamma(n_i). \end{aligned}$$

Donde $(\theta)_n$ es el factorial decreciente definido en el Capítulo 1 y que también es conocido como el símbolo de Pochhammer.

Regresando a la distribución predictiva, uno puede darse cuenta de que para el proceso Dirichlet el muestreo de una muestra no requiere de simular la variable latente y se establece el esquema de muestreo de Blackwell-MacQueen de la siguiente forma.

1. Simular X_1 de P_0 .

2. Al paso i .

a) Simular ϵ_i de P_0 .

$$\text{b) } X_i = \begin{cases} \epsilon_i & \text{con probabilidad } \frac{\theta}{\theta+i-1} \\ X_1 & \text{con probabilidad } \frac{1}{\theta+i-1} \\ \vdots & \\ X_{i-1} & \text{con probabilidad } \frac{1}{\theta+i-1}. \end{cases}$$

Algoritmo que será de utilidad al momento de trabajar con la estimación de mezclas.

3.2. Proceso σ -estable

Esta medida de probabilidad aleatoria fue introducida por primera vez por Kingman (1975), para su construcción se debe considerar un parámetro $\sigma \in (0, 1)$ y la medida completamente aleatoria, Φ , tendrá la siguiente intensidad de Lévy

$$\nu(ds, dx) = \frac{\sigma}{\Gamma(1-\sigma)} s^{-1-\sigma} ds \alpha(dx).$$

Que al igual que para el proceso Dirichlet es una medida homogénea y donde si α es una medida no negativa y finita se puede realizar el mismo proceso de normalización y definir a la medida de probabilidad aleatoria como

$$\mathcal{P}(dx) = \frac{\Phi(dx)}{\Phi(\mathbb{R})}.$$

De la expresión anterior se puede observar que

$$\rho(ds) = \frac{\sigma}{\Gamma(1-\sigma)} s^{-1-\sigma} ds \quad \Rightarrow \quad \rho(\mathbb{R}_+) = \infty.$$

Por lo que se tiene una medida con actividad infinita y que además cumple con la condición de integrabilidad ya que,

$$\begin{aligned}
\int_{\mathbb{R}_+} (s \wedge 1) \frac{\sigma}{\Gamma(1-\sigma)} s^{-1-\sigma} ds &= \frac{\sigma}{\Gamma(1-\sigma)} \left(\int_0^1 s^{-\sigma} ds + \int_1^\infty s^{-1-\sigma} ds \right) \\
&= \frac{\sigma}{\Gamma(1-\sigma)} \left(\frac{s^{1-\sigma}}{1-\sigma} \Big|_0^1 - \frac{s^{-\sigma}}{\sigma} \Big|_1^\infty \right) \\
&= \frac{\sigma}{\Gamma(1-\sigma)} \left(\frac{1}{1-\sigma} + \frac{1}{\sigma} \right) < \infty.
\end{aligned}$$

Para obtener el exponente laplaciano primero se puede ver lo que sucede cuando tomamos una función, f , medible y se obtiene la funcional de Laplace,

$$\int_{\mathbb{R} \times \mathbb{R}_+} (1 - e^{-sf(x)}) \nu(ds, dx) = \int_{\mathbb{R}} \left[\int_{\mathbb{R}_+} (1 - e^{-sf(x)}) \frac{\sigma}{\Gamma(1-\sigma)} s^{-1-\sigma} ds \right] \alpha(dx).$$

Donde mediante integración por partes se tiene que

$$\int_{\mathbb{R}_+} (1 - e^{-sf(x)}) s^{-1-\sigma} ds = -(1 - e^{-sf(x)}) \frac{s^{-\sigma}}{\sigma} \Big|_0^\infty + \int_{\mathbb{R}_+} \frac{s^{-\sigma} f(x) e^{-sf(x)}}{\sigma} ds.$$

Donde el primer término es igual a 0 y para resolver el segundo término se hace lo siguiente,

$$\int_{\mathbb{R}_+} \frac{s^{-\sigma} f(x) e^{-sf(x)}}{\sigma} ds = \frac{f(x)^\sigma}{\sigma} \int_{\mathbb{R}_+} f(x) (sf(x))^{-\sigma} e^{-sf(x)} ds.$$

Realizando el cambio de variable $w = sf(x)$ se llega a que esta última expresión es igual

a

$$\frac{f(x)^\sigma}{\sigma} \int_{\mathbb{R}_+} w^{-\alpha} e^{-w} dw = \frac{\Gamma(1-\sigma) f(x)^\sigma}{\sigma}.$$

Por lo que,

$$\int_{\mathbb{R} \times \mathbb{R}_+} (1 - e^{-sf(x)}) \nu(ds, dx) = \int_{\mathbb{R}} f(x)^\sigma \alpha(dx) = \theta \int_{\mathbb{R}} f(x)^\sigma P_0(dx)$$

Que está bien definido para cualquier función f tal que $\int f^\sigma \alpha < \infty$. De esta manera, se sigue de forma inmediata que el exponente laplaciano será igual a

$$\varphi(u) = \theta u^\sigma < \infty.$$

Más aún, tomando a $f = \lambda \mathbb{1}_B$ para $B \in \mathcal{B}(\mathbb{R})$. Se obtiene que el exponente laplaciano de $\Phi(B)$ está dado por

$$\begin{aligned} \mathbb{E}(e^{-\lambda \Phi(B)}) &= \exp \left[-\theta \int_{\mathbb{R}} \lambda^\sigma \mathbb{1}_B P_0(dx) \right] \\ &= \exp [-\theta \lambda^\sigma P_0(B)]. \end{aligned}$$

Por lo cual $\Phi(B)$ tiene la misma transformada de Laplace que la de una variable aleatoria positiva estable. De esta manera se dice que la correspondiente medida de probabilidad aleatoria se distribuye como un proceso σ -estable, que se suele denotar en la literatura como $\mathcal{P} \sim N - St(\sigma, \theta P_0)$.

El siguiente paso es la caracterización de la distribución posterior, que como ya se ha dicho anteriormente y a diferencia de lo que pasa con el proceso Dirichlet, no se tendrá una familia conjugada. Para esto, primero se debe obtener la medida de Lévy de $\Phi^{(U_n)}$ y la distribución de $J_i^{(U_n, \mathbf{X}^{(n)})}$. De esta manera se tiene que

$$\nu^{(U_n)}(ds, dx) = e^{-U_n s} \frac{\sigma}{\Gamma(1 - \sigma)} s^{-1 - \sigma} ds \theta P_0(dx).$$

Y la densidad de los saltos $J_i^{(U_n, \mathbf{X}^{(n)})}$ será proporcional a

$$s^{n_i} e^{-U_n s} \frac{\sigma}{\Gamma(1 - \sigma)} s^{-1 - \sigma} ds \propto s^{n_i - \sigma - 1} e^{-U_n s}.$$

Se puede notar que es el kernel de la distribución gama con parámetros $(n_i - \sigma, U_n)$. Con lo anterior en mente se puede obtener la funcional de Laplace de $\Phi^{(U_n, \mathbf{X}^{(n)})}$ y que estará dada por

$$\begin{aligned} & \mathbb{E} \left[\exp \left(- \int_{\mathbb{R}} f(x) \Phi(dx) \middle| U_n, \mathbf{X}^{(n)} \right) \right] \\ = & \mathbb{E} \left[\exp \left(- \int_{\mathbb{R}} f(x) \Phi^{(U_n)}(dx) \right) \right] \mathbb{E} \left[\exp \left(- \int_{\mathbb{R}} f(x) \left(\sum_{i=1}^{n(\pi)} J_i^{(U_n, \mathbf{X}^{(n)})} \delta_{Y_i}(dx) \right) \right) \right] \end{aligned} \quad (3.4)$$

Donde el primer término del lado derecho de la ecuación (3.4) se desarrolla como sigue

$$\begin{aligned} & \mathbb{E} \left[\exp \left(- \int_{\mathbb{R}} f(x) \Phi^{(U_n)}(dx) \right) \right] \\ = & \exp \left(- \int_{\mathbb{R}} \int_{\mathbb{R}_+} (1 - e^{-sf(x)}) e^{-U_n s} \frac{\sigma}{\Gamma(1 - \sigma)} s^{-1 - \sigma} ds \theta P_0(dx) \right) \\ = & \exp \left(- \theta \int_{\mathbb{R}} \frac{\sigma}{\Gamma(1 - \sigma)} \int_{\mathbb{R}_+} (e^{-U_n s} - e^{-(f(x) + U_n)s}) s^{-1 - \sigma} ds P_0(dx) \right). \end{aligned} \quad (3.5)$$

Donde mediante integración por partes y de manera similar a como se procedió arriba se tiene

$$\begin{aligned} & \int_{\mathbb{R}_+} (e^{-U_n s} - e^{-(f(x) + U_n)s}) s^{-1 - \sigma} ds \\ = & -(e^{-U_n s} - e^{-(f(x) + U_n)s}) \frac{s^{-\sigma}}{\sigma} \Big|_0^\infty + \int_{\mathbb{R}_+} \frac{s^{-\sigma}}{\sigma} (-U_n e^{-U_n s} + (f(x) + U_n) e^{-(f(x) + U_n)s}) ds. \end{aligned}$$

El primer término es 0 y la segunda integral se puede dividir y mediante cambios de variable dados por $w = -U_n s$ y $w' = (f(x) + U_n)s$ se tiene

$$\int_{\mathbb{R}_+} \frac{s^{-\sigma}}{\sigma} (-U_n e^{-U_n s} + (f(x) + U_n) e^{-(f(x) + U_n)s}) ds = ((f(x) + U_n)^\sigma - U_n^\sigma) \frac{\Gamma(1 - \sigma)}{\sigma}.$$

Por lo que, sustituyendo esta expresión en la ecuación (3.5), se tiene que

$$\mathbb{E} \left[\exp \left(- \int_{\mathbb{R}} f(x) \Phi^{(U_n)}(dx) \right) \right] = \exp \left(-\theta \int_{\mathbb{R}} ((f(x) + U_n)^\sigma - U_n^\sigma) P_0(dx) \right).$$

Y el segundo término del lado izquierdo de la ecuación (3.4) y aprovechando que se conoce con exactitud la distribución de $J_i^{(U_n, \mathbf{X}^{(n)})}$ se desarrolla como sigue,

$$\begin{aligned} \mathbb{E} \left[\exp \left(- \int_{\mathbb{R}} f(x) \left(\sum_{i=1}^{n(\pi)} J_i^{(U_n, \mathbf{X}^{(n)})} \delta_{Y_i}(dx) \right) \right) \right] &= \prod_{i=1}^{n(\pi)} \mathbb{E} \left[\exp \left(-J_i^{(U_n, \mathbf{X}^{(n)})} f(Y_i) \right) \right] \\ &= \prod_{i=1}^{n(\pi)} \left(1 + \frac{f(Y_i)}{U_n} \right)^{-(n_i - \sigma)} \\ &= \prod_{i=1}^{n(\pi)} \frac{U_n^{n_i - \sigma}}{(U_n + f(Y_i))^{n_i - \sigma}}. \end{aligned}$$

De esta manera, la funcional de Laplace de la medida $\Phi^{(U_n, \mathbf{X}^{(n)})}$ está dada por

$$\exp \left(-\theta \int_{\mathbb{R}} ((f(x) + U_n)^\sigma - U_n^\sigma) P_0(dx) \right) \prod_{i=1}^{n(\pi)} \frac{U_n^{n_i - \sigma}}{(U_n + f(Y_i))^{n_i - \sigma}}.$$

Una vez obtenida la caracterización posterior para la medida completamente aleatoria, el siguiente paso en el estudio del proceso σ -estable es la obtención de la distribución predictiva. Para esto se necesita recordar que $e^{-\varphi(u)} = e^{-\theta u^\sigma}$ y que se deben obtener las siguientes cantidades.

$$\begin{aligned} \tau_{n_i}(u|Y_i) &= \int_{\mathbb{R}_+} s^{n_i} e^{-us} \frac{\sigma}{\Gamma(1 - \sigma)} s^{-1 - \sigma} ds \\ &= \frac{\sigma}{\Gamma(1 - \sigma)} \int_{\mathbb{R}_+} s^{n_i - \sigma - 1} e^{-us} ds \\ &= \frac{\sigma \Gamma(n_i - \sigma)}{\Gamma(1 - \sigma) u^{n_i - \sigma}}. \end{aligned}$$

Y la distribución posterior de la variable latente dada la muestra, suponiendo que $n(\pi) = k$, se tiene que

$$\begin{aligned} f_{(U_n|\mathbf{X}^{(n)})}(u) &\propto u^{n-1} e^{-\theta u^\sigma} \prod_{i=1}^k \frac{\sigma \Gamma(n_i - \sigma)}{\Gamma(1 - \sigma) u^{n_i - \sigma}} \\ &\propto u^{k\sigma - 1} e^{-\theta u^\sigma}. \end{aligned}$$

Para encontrar la constante de normalización basta hacer el cambio de variable

$$u^\sigma = w \quad \Rightarrow \quad w^{\frac{1}{\sigma}} = u \quad \Rightarrow \quad du = \frac{1}{\sigma} w^{\frac{1}{\sigma} - 1} dw.$$

Por lo que,

$$\begin{aligned} \int_{\mathbb{R}_+} u^{k\sigma - 1} e^{-\theta u^\sigma} du &= \int_{\mathbb{R}_+} (w^{\frac{1}{\sigma}})^{k\sigma - 1} e^{-\theta w} \frac{1}{\sigma} w^{\frac{1}{\sigma} - 1} dw \\ &= \frac{1}{\sigma} \int_{\mathbb{R}_+} w^{k-1} e^{-\theta w} dw \\ &= \frac{1}{\sigma \theta^k} \int_{\mathbb{R}_+} \theta (\theta w)^{k-1} e^{-\theta w} dw \\ (\text{haciendo } z = \theta w) &= \frac{1}{\sigma \theta^k} \int_{\mathbb{R}_+} z^{k-1} e^{-\theta z} dz \\ &= \frac{\Gamma(k)}{\sigma \theta^k} \end{aligned}$$

Que proporciona una expresión cerrada para la densidad posterior de la variable latente y que está dada por

$$f_{(U_n|\mathbf{X}^{(n)})}(u) = \frac{\sigma \theta^k}{\Gamma(k)} u^{k\sigma - 1} e^{-\theta u^\sigma}.$$

Con las expresiones anteriores se tiene todo lo necesario para la obtención de los pesos $\omega^{(n)}$ y $\omega_i^{(n)}$ que permiten caracterizar a la distribución predictiva.

$$\begin{aligned}
\omega^{(n)} &= \frac{\theta}{n} \int_{\mathbb{R}} u \tau_1(u|x) f_{(U_n|\mathbf{X}^{(n)})}(u) du \\
&= \frac{\theta}{n} \int_{\mathbb{R}_+} u \sigma u^{\sigma-1} \frac{\sigma \theta^k}{\Gamma(k)} u^{k\sigma-1} e^{-\theta u^\sigma} du \\
&= \frac{\sigma^2 \theta^{k+1}}{n \Gamma(k)} \int_{\mathbb{R}_+} u^{\sigma(k+1)-1} e^{-\theta u^\sigma} du \\
&= \frac{\sigma^2 \theta^{k+1}}{n \Gamma(k)} \cdot \frac{\Gamma(k+1)}{\sigma \theta^{k+1}} \\
&= \frac{k\sigma}{n}.
\end{aligned}$$

Y con pesos para las distintas observaciones,

$$\begin{aligned}
\omega_i^{(n)} &= \int_{\mathbb{R}_+} u \frac{\tau_{n_i+1}(u|Y_i)}{\tau_{n_i}(u|Y_i)} f_{(U_n|\mathbf{X}^{(n)})}(u) du \\
&= \int_{\mathbb{R}_+} u \cdot \frac{\sigma \Gamma(n_i+1-\sigma)}{\Gamma(1-\sigma) u^{n_i+1-\sigma}} \cdot \frac{\Gamma(1-\sigma) u^{n_i-\sigma}}{\sigma \Gamma(n_i-\sigma)} \frac{\sigma \theta^k}{\Gamma(k)} u^{k\sigma-1} e^{-\theta u^\sigma} du \\
&= \frac{(n_i-\sigma) \sigma \theta^k}{\Gamma(k)} \int_{\mathbb{R}_+} u^{k\sigma-1} e^{-\theta u^\sigma} du \\
&= \frac{(n_i-\sigma) \sigma \theta^k}{\Gamma(k)} \cdot \frac{\Gamma(k)}{\sigma \theta^k} \\
&= (n_i - \sigma).
\end{aligned}$$

Teniendo estas expresiones se puede dar una expresión cerrada y fácil de analizar, como en el caso Dirichlet, de la distribución predictiva, que estará dada por

$$\mathbb{P}(X_{n+1} \in dx | \mathbf{X}^{(n)}) = \frac{k\sigma}{n} P_0(dx) + \frac{1}{n} \sum_{i=1}^k (n_i - \sigma) \delta_{Y_i}(dx).$$

Lo cual dice lo siguiente sobre la observación $n + 1$ dada la muestra,

$$\begin{aligned}\mathbb{P}(X_{n+1} \neq Y_j \quad \forall j \in \{1, \dots, n(\pi)\} | \mathbf{X}^{(n)}) &= \frac{k\sigma}{n} \\ \mathbb{P}(X_{n+1} = Y_j \text{ p.a. } j \in \{1, \dots, n(\pi)\} | \mathbf{X}^{(n)}) &= \frac{n_j - \sigma}{n}.\end{aligned}$$

Que además permite recuperar cierta información relacionada con la EPPF, ya que se cumple lo siguiente,

$$\begin{aligned}\frac{\Pi_{k+1}^{(n)}(n_1, \dots, n_k, 1)}{\Pi_k^{(n)}(n_1, \dots, n_k)} &= \frac{k\sigma}{n} \\ \frac{\Pi_k^{(n)}(n_1, \dots, n_{j-1}, n_j + 1, n_{j+1}, \dots, n_k)}{\Pi_k^{(n)}(n_1, \dots, n_k)} &= \frac{n_j - \sigma}{n}.\end{aligned}$$

Y donde la obtención de la EPPF resulta factible y que está dada por

$$\begin{aligned}\Pi_k^{(n)}(n_1, \dots, n_k) &= \frac{1}{\Gamma(n)} \int_{\mathbb{R}_+} u^{n-1} e^{-\varphi(u)} \left(\prod_{i=1}^k \tau_{n_i}(u|x) \theta P_0(dx) \right) du \\ &= \frac{1}{\Gamma(n)} \int_{\mathbb{R}_+} u^{n-1} e^{-\theta u^\sigma} \left(\prod_{i=1}^k \int_{\mathbb{R}} \frac{\sigma \Gamma(n_i - \sigma)}{\Gamma(1 - \sigma) u^{n_i - \sigma}} \theta P_0(dx) \right) du \\ &= \frac{\theta^k \sigma^k}{\Gamma(n) \Gamma(1 - \sigma)} \prod_{i=1}^k \Gamma(n_i - \sigma) \int_{\mathbb{R}_+} u^{\sigma k - 1} e^{-\theta u^\sigma} du \\ &= \frac{\theta^k \sigma^k}{\Gamma(n)} \prod_{i=1}^k \frac{\Gamma(n_i - \sigma)}{\Gamma(1 - \sigma)} \cdot \frac{\Gamma(k)}{\sigma \theta^k} \\ &= \frac{\sigma^{k-1} \Gamma(k)}{\Gamma(n)} \prod_{i=1}^k (1 - \sigma)_{n_i - 1}.\end{aligned}$$

Se puede observar que θ está siendo irrelevante para la EPPF y para la distribución predictiva, por lo cual se suele tomar a $\theta = 1$. Otro detalle a destacar es que al tener expresiones cerradas para la distribución predictiva, el esquema de muestreo de Blackwell-MacQueen se simplifica y es ciertamente intuitivo y cuyo algoritmo es el siguiente.

1. Simular X_1 de P_0 .
2. Al momento $i + 1$, se denota por $Y_{1,i}, \dots, Y_{k_i,i}$ los distintos valores observados en la colección X_1, \dots, X_i y con respectivas frecuencias $n_{1,i}, \dots, n_{k_i,i}$.
 - a. Simular ϵ_{i+1} de P_0 .

$$\text{b. } X_{i+1} = \begin{cases} \epsilon_{i+1} & \text{con probabilidad } \frac{k\sigma}{n} \\ Y_{1,i} & \text{con probabilidad } \frac{n_{1,i}-\sigma}{n} \\ \vdots & \\ Y_{k_i,i} & \text{con probabilidad } \frac{n_{k_i,i}-\sigma}{n}. \end{cases}$$

3.3. Proceso gama generalizado

Hasta este momento se ha visto que el proceso Dirichlet y el proceso gama son dos procesos que tienen estructuras similares, como ejemplo de esto se tiene que la distribución de los saltos condicionados a la variable latente y a la muestra se distribuyen gama en ambos casos y se han podido obtener expresiones cerradas y ciertamente sencillas para la distribución predictiva. Esto no es casualidad y como se verá a continuación ambos procesos pertenecen a una clase más amplia, que es la del proceso gama generalizado.

El proceso gama generalizado fue introducido por Brix (1999) y para el cual se considera una medida de Lévy homogénea dada por

$$\nu(ds, dx) = \frac{1}{\Gamma(1-\sigma)} s^{-1-\sigma} e^{-\gamma s} ds \alpha(dx).$$

Donde como de costumbre se pedirá que α sea una medida no nula y finita y donde $\sigma \in (0, 1)$ y $\gamma > 0$. De este proceso se desprenden casos particulares de interés como:

1. Si $\gamma \downarrow 0$ se recupera el proceso σ -estable, con una ligera modificación en la constante que acompaña a la $\rho(ds)$, que es un término igual a σ , todo esto con el objeto de poder recuperar el proceso Dirichlet.
2. Si $\gamma = 1$ y $\sigma \downarrow 0$ se recupera el proceso Dirichlet.
3. Si $\sigma = \frac{1}{2}$ se tiene el proceso inverso gaussiano, que ha sido estudiado por Lijoi, Mena, Prünster (2005) en el trabajo titulado *Hierarchical mixture modelling with normalized inverse Gaussian priors*. En el cual también se da una caracterización para las distribuciones finito dimensionales.

Verificando las mismas condiciones que para los casos anteriores se tiene que

$$\int_{\mathbb{R}_+} s^{-1-\sigma} e^{-\gamma s} ds = -\frac{e^{-\gamma s} s^{-\sigma}}{\sigma} \Big|_0^\infty - \int_{\mathbb{R}_+} s^{-\sigma} \gamma e^{-\gamma s} ds = \infty.$$

Ya que la segunda integral es finita y el primer término es igual a ∞ , por lo cual se tiene que $\rho(\mathbb{R}_+) = \infty$ y así se cuenta una medida de actividad infinita. Más aún, se sigue que

$$\begin{aligned} \int_{\mathbb{R}_+} (s \wedge 1) \frac{1}{\Gamma(1-\sigma)} s^{-1-\sigma} e^{-\gamma s} ds &= \int_0^1 s^{-\sigma} e^{-\gamma s} ds + \int_1^\infty s^{-1-\sigma} e^{-\gamma s} ds \\ &< \gamma^{\sigma-1} \int_{\mathbb{R}_+} \gamma(\gamma s)^{-\sigma} e^{-\gamma s} ds + \int_1^\infty s^{-1-\sigma} e^{-\gamma s} ds \\ &= \gamma^{\sigma-1} \Gamma(1-\sigma) - \frac{e^{-\gamma s} s^{-\sigma}}{\sigma} \Big|_1^\infty - \int_1^\infty s^{-\sigma} \gamma e^{-\gamma s} ds. \end{aligned}$$

Siendo todos los términos finitos por lo cual se cumple la condición de integrabilidad, i.e.,

$$\int_{\mathbb{R} \times \mathbb{R}_+} (s \wedge 1) \frac{1}{\Gamma(1-\sigma)} s^{-1-\sigma} e^{-\gamma s} ds \alpha(dx) < \infty.$$

Para obtener la funcional de Laplace se debe calcular la siguiente integral, que se puede identificar como una de las integrales que se resolvieron para obtener la funcional de Laplace de $\Phi^{(U_n, \mathbf{X}^{(n)})}$ en el proceso σ -estable

$$\begin{aligned} \int_{\mathbb{R}_+} (1 - e^{-sf(x)})s^{-1-\sigma}e^{-\gamma s} ds &= \int_{\mathbb{R}_+} (e^{-\gamma s} - e^{-(\gamma+f(x))s})s^{-1-\sigma} ds \\ &= \frac{\Gamma(1-\sigma)}{\sigma}((f(x) + \gamma)^\sigma - \gamma^\sigma). \end{aligned}$$

Por lo que la funcional de Laplace en esta ocasión está dada por

$$\begin{aligned} \mathbb{E} \left[\exp \left(- \int_{\mathbb{R}} f(x) \Phi(dx) \right) \right] &= \int_{\mathbb{R}} \int_{\mathbb{R}_+} (1 - e^{-sf(x)}) \frac{1}{\Gamma(1-\sigma)} s^{-1-\sigma} e^{-\gamma s} ds \alpha(dx) \\ &= \frac{\theta}{\sigma} \int_{\mathbb{R}_+} ((f(x) + \gamma)^\sigma - \gamma^\sigma) P_0(dx). \end{aligned}$$

Que está bien definida para todas aquellas funciones medibles, f , tales que $\int (f + \gamma)^\sigma \alpha < \infty$. De esta última expresión se sigue de manera inmediata que el exponente laplaciano estará dado por

$$\varphi(u) = \frac{\theta}{\sigma} ((u + \gamma)^\sigma - \gamma^\sigma) < \infty.$$

Y más aún, al tomar a $f = \lambda \mathbb{1}_B$ se puede caracterizar la distribución de la variable aleatoria $\Phi(B)$, cuya transformada de Laplace queda como

$$\mathbb{E}(e^{-\lambda \Phi(B)}) = \exp \left[- \frac{\theta}{\sigma} ((\lambda + \gamma)^\sigma - \gamma^\sigma) P_0(B) \right].$$

Claramente al tomar $\gamma \downarrow 0$ se recupera la transformada de Laplace para la distribución σ -estable y mediante el uso de la regla de L'Hôpital se recupera la transformada de Laplace para la distribución gama cuando $\sigma \downarrow 0$.

Para encontrar la distribución posterior de la medida completamente aleatoria, se debe recordar que se necesitan los siguientes elementos.

$$\nu^{(U_n)}(ds, dx) = e^{-U_n s} \frac{1}{\Gamma(1-\sigma)} s^{-1-\sigma} e^{-\gamma s} ds \theta P_0(dx).$$

Y la densidad de $J_i^{(U_n, \mathbf{X}^{(n)})}$,

$$\begin{aligned} f_i(s) &\propto s^{n_i} e^{-U_n s} \frac{1}{\Gamma(1-\sigma)} s^{-1-\sigma} e^{-\gamma s} \\ &\propto s^{n_i - \sigma - 1} e^{-(\gamma + U_n)s}. \end{aligned}$$

Por lo que $J_i^{(U_n, \mathbf{X}^{(n)})} \sim \text{gama}(n_i - \sigma, \gamma + U_n)$, lo cual era algo de esperarse en este momento. De esta manera ya se tienen los elementos necesarios para obtener la funcional de Laplace de $\Phi^{(U_n, \mathbf{X}^{(n)})}$.

$$\begin{aligned} &\mathbb{E} \left[\exp \left(- \int_{\mathbb{R}} f(x) \Phi(dx) \middle| U_n, \mathbf{X}^{(n)} \right) \right] \\ = &\mathbb{E} \left[\exp \left(- \int_{\mathbb{R}} f(x) \Phi^{(U_n)}(dx) \right) \right] \mathbb{E} \left[\exp \left(- \int_{\mathbb{R}} f(x) \left(\sum_{i=1}^{n(\pi)} J_i^{(U_n, \mathbf{X}^{(n)})} \delta_{Y_i}(dx) \right) \right) \right] \end{aligned} \quad (3.6)$$

Donde el primer término del lado derecho de la ecuación (3.6) se desarrolla como sigue,

$$\begin{aligned} &\mathbb{E} \left[\exp \left(- \int_{\mathbb{R}} f(x) \Phi^{(U_n)}(dx) \right) \right] \\ = &\exp \left(- \int_{\mathbb{R}} \int_{\mathbb{R}_+} (1 - e^{-sf(x)}) \frac{e^{-U_n s}}{\Gamma(1-\sigma)} s^{-1-\sigma} e^{-\gamma s} ds \theta P_0(dx) \right) \\ = &\exp \left(- \theta \int_{\mathbb{R}} \frac{1}{\Gamma(1-\sigma)} \int_{\mathbb{R}_+} (e^{-(U_n + \gamma)s} - e^{-(U_n + \gamma + f(x))s}) s^{-1-\sigma} P_0(dx) \right) \\ = &\exp \left(- \frac{\theta}{\sigma} \int_{\mathbb{R}} ((f(x) + U_n + \gamma)^\sigma - (U_n + \gamma)^\sigma) P_0(dx) \right). \end{aligned}$$

Y el término correspondiente a los saltos que está dado por,

$$\begin{aligned}
\mathbb{E} \left[\exp \left(- \int_{\mathbb{R}} f(x) \left(\sum_{i=1}^{n(\pi)} J_i^{(U_n, \mathbf{X}^{(n)})} \delta_{Y_i}(dx) \right) \right) \right] &= \prod_{i=1}^{n(\pi)} \mathbb{E} \left[\exp \left(- J_i^{(U_n, \mathbf{X}^{(n)})} f(Y_i) \right) \right] \\
&= \prod_{i=1}^{n(\pi)} \left(1 + \frac{f(Y_i)}{\gamma + U_n} \right)^{-(n_i - \sigma)} \\
&= \prod_{i=1}^{n(\pi)} \frac{(\gamma + U_n)^{n_i - \sigma}}{(\gamma + U_n + f(Y_i))^{n_i - \sigma}}.
\end{aligned}$$

Por lo cual, la funcional de Laplace es igual a

$$\exp \left(- \frac{\theta}{\sigma} \int_{\mathbb{R}} ((f(x) + U_n + \gamma)^\sigma - (U_n + \gamma)^\sigma) P_0(dx) \right) \prod_{i=1}^{n(\pi)} \frac{(\gamma + U_n)^{n_i - \sigma}}{(\gamma + U_n + f(Y_i))^{n_i - \sigma}}.$$

Con lo cual ya se tiene caracterizada a $\Phi^{(U_n, \mathbf{X}^{(n)})}$ y por ende la medida de probabilidad aleatoria asociada. Para la obtención de los pesos de la distribución predictiva se necesitan la siguiente expresión.

$$\begin{aligned}
\tau_{n_i}(u|Y_i) &= \int_{\mathbb{R}_+} s^{n_i} e^{-us} \frac{1}{\Gamma(1 - \sigma)} s^{-1 - \sigma} e^{-\gamma s} ds \\
&= \frac{1}{\Gamma(1 - \sigma)} \int_{\mathbb{R}_+} s^{n_i - \sigma - 1} e^{-\gamma s} ds \\
&= \frac{\Gamma(n_i - \sigma)}{\Gamma(1 - \sigma)(u + \gamma)^{n_i - \sigma}}.
\end{aligned}$$

Y la densidad condicional de la variable latente dada la muestra, que para el proceso gama generalizado tendrá la siguiente forma, (suponiendo $n(\pi) = k$),

$$\begin{aligned}
f_{(U_n|\mathbf{X}^{(n)})}(u) &\propto u^{n-1} \left(\prod_{i=1}^k \frac{\Gamma(n_i - \sigma)}{\Gamma(1 - \sigma)} \right) (u + \gamma)^{\sigma k - n} e^{-\frac{\theta}{\sigma}((u+\gamma)^\sigma - \gamma^\sigma)} \\
&\propto u^{n-1} (u + \gamma)^{\sigma k - n} e^{-\frac{\theta}{\sigma}((u+\gamma)^\sigma - \gamma^\sigma)}.
\end{aligned}$$

En esta caso la constante de normalización no tendrá una expresión tan sencilla como para los procesos anteriores; sin embargo, tiene una expresión cerrada, lo cual permitirá a su vez encontrar expresiones cerradas para los pesos y por ende para la distribución predictiva, lo cual es importante al momento de establecer algoritmos de muestreo del tipo Blackwell-MacQueen.

Procediendo con el cálculo de la constante de normalización, lo primero que se hace es considerar una reparametrización, elegida como $\beta = \frac{\theta\gamma}{\sigma}$, por lo que,

$$\int_{\mathbb{R}_+} u^{n-1}(u+\gamma)^{\sigma k-n} e^{-\frac{\theta}{\sigma}(u+\gamma)^\sigma - \gamma^\sigma} du = \int_{\mathbb{R}_+} u^{n-1}(u+\gamma)^{\sigma k-n} e^{-\frac{\theta}{\sigma}(u+\gamma)^\sigma} e^\beta \quad (3.7)$$

El siguiente paso es considerar el cambio de variable

$$w = (u + \gamma)^\sigma \quad \Rightarrow \quad u = w^{\frac{1}{\sigma}} - \gamma \quad \Rightarrow \quad du = \frac{1}{\sigma} w^{\frac{1}{\sigma}-1} dw.$$

De esta manera la integral en la ecuación (3.7) se convierte en

$$\begin{aligned} & \frac{e^\beta}{\sigma} \int_{\gamma^\sigma}^{\infty} (y^{\frac{1}{\sigma}} - \gamma)^{n-1} (y^{\frac{1}{\sigma}})^{\sigma k-n+1-\sigma} e^{-\frac{\theta}{\sigma}y} dy \\ &= \frac{e^\beta}{\sigma} \int_{\gamma^\sigma}^{\infty} (y^{\frac{1}{\sigma}} - \gamma)^{n-1} y^{k-\frac{n}{\sigma}+\frac{1}{\sigma}-1} e^{-\frac{\theta}{\sigma}y} dy \\ &= \frac{e^\beta}{\sigma} \int_{\gamma^\sigma}^{\infty} \sum_{i=0}^{n-1} \binom{n-1}{i} (y^{\frac{1}{\sigma}})^{n-1-i} (-\gamma)^i (y^{\frac{1}{\sigma}} - \gamma)^{n-1} y^{k-\frac{n}{\sigma}+\frac{1}{\sigma}-1} e^{-\frac{\theta}{\sigma}y} dy \\ &= \frac{e^\beta}{\sigma} \sum_{i=0}^{n-1} \binom{n-1}{i} (-\gamma)^i \int_{\gamma^\sigma}^{\infty} y^{k-\frac{i}{\sigma}-1} e^{-\frac{\theta}{\sigma}y} dy \\ &= \frac{e^\beta}{\sigma} \sum_{i=0}^{n-1} \binom{n-1}{i} (-\gamma)^i \left(\frac{\sigma}{\theta}\right)^{k-\frac{i}{\sigma}} \int_{\gamma^\sigma}^{\infty} \frac{\theta}{\sigma} \left(\frac{\theta y}{\sigma}\right)^{k-\frac{i}{\sigma}-1} e^{-\frac{\theta}{\sigma}y} dz \\ &= \frac{e^\beta}{\sigma} \sum_{i=0}^{n-1} \binom{n-1}{i} (-1)^i \beta^{\frac{i}{\sigma}} \left(\frac{\sigma}{\theta}\right)^{\frac{i}{\sigma}} \left(\frac{\sigma}{\theta}\right)^{k-\frac{i}{\sigma}} \int_{\beta}^{\infty} z^{k-\frac{i}{\sigma}-1} e^{-z} dz \\ &= \frac{e^\beta}{\sigma} \left(\frac{\sigma}{\theta}\right)^k \sum_{i=0}^{n-1} \binom{n-1}{i} (-1)^i \beta^{\frac{i}{\sigma}} \Gamma\left(k - \frac{i}{\sigma}; \beta\right). \end{aligned}$$

Donde $\Gamma(x; s)$ es la función gama incompleta, i.e.,

$$\Gamma(x; s) = \int_s^{\infty} z^{x-1} e^{-z} dz.$$

Denotando por c la constante de normalización, se pueden obtener los pesos $\omega^{(n)}$ y $\omega_i^{(n)}$, que estarán dados por

$$\begin{aligned} \omega^{(n)} &= \frac{\theta}{nc} \int_{\mathbb{R}_+} u(u+\gamma)^{\sigma-1} u^{n-1} (u+\gamma)^{\sigma k-n} e^{-\frac{\theta}{\sigma}((u+\gamma)^\sigma - \gamma^\sigma)} du \\ &= \frac{\theta}{nc} \int_{\mathbb{R}_+} u^n (u+\gamma)^{\sigma(k+1)-n-1} e^{-\frac{\theta}{\sigma}((u+\gamma)^\sigma - \gamma^\sigma)}. \end{aligned}$$

Que es una integral que se resuelve de forma análoga a la constante de normalización, por lo cual se obtiene de forma inmediata que

$$\begin{aligned} \omega^{(n)} &= \frac{\theta \frac{e^\beta}{\sigma} \left(\frac{\sigma}{\theta}\right)^{k+1} \sum_{i=0}^n \binom{n}{i} (-1)^i \beta^{\frac{i}{\sigma}} \Gamma\left(k+1 - \frac{i}{\sigma}; \beta\right)}{n \frac{e^\beta}{\sigma} \left(\frac{\sigma}{\theta}\right)^k \sum_{i=0}^{n-1} \binom{n-1}{i} (-1)^i \beta^{\frac{i}{\sigma}} \Gamma\left(k - \frac{i}{\sigma}; \beta\right)} \\ &= \frac{\sigma \sum_{i=0}^n \binom{n}{i} (-1)^i \beta^{\frac{i}{\sigma}} \Gamma\left(k+1 - \frac{i}{\sigma}; \beta\right)}{n \sum_{i=0}^{n-1} \binom{n-1}{i} (-1)^i \beta^{\frac{i}{\sigma}} \Gamma\left(k - \frac{i}{\sigma}; \beta\right)} \end{aligned}$$

Y de forma similar se obtiene una expresión para $\omega_i^{(n)}$ que será igual a

$$\begin{aligned} &\frac{1}{c} \int_{\mathbb{R}_+} u \left(\frac{\Gamma(n_i + 1 - \sigma)}{(u+\gamma)^{n_i - \sigma + 1}} \right) \left(\frac{(u+\gamma)^{n_i - \sigma}}{\Gamma(n_i - \sigma)} \right) u^{n-1} (u+\gamma)^{\sigma k-n} e^{-\frac{\theta}{\sigma}((u+\gamma)^\sigma - \gamma^\sigma)} du \\ &= \frac{(n_i - \sigma)}{c} \int_{\mathbb{R}_+} u^n (u+\gamma)^{\sigma k-n-1} e^{-\frac{\theta}{\sigma}((u+\gamma)^\sigma - \gamma^\sigma)} du. \end{aligned}$$

Donde la última integral se resuelve de forma similar a lo que ya se hizo anteriormente, por lo cual

$$\omega_i^{(n)} = (n_i - \sigma) \frac{\sum_{i=0}^n \binom{n}{i} (-1)^i \beta^{\frac{i}{\sigma}} \Gamma\left(k - \frac{i}{\sigma}; \beta\right)}{\sum_{i=0}^{n-1} \binom{n-1}{i} (-1)^i \beta^{\frac{i}{\sigma}} \Gamma\left(k - \frac{i}{\sigma}; \beta\right)}.$$

Que como ya se mencionó antes, son expresiones ciertamente un poco más complicadas que las obtenidas con el proceso Dirichlet y el proceso σ -estable y que conllevan la evaluación de varias integrales incompletas, que si bien pueden ser resueltas involucrarán una rutina más compleja. Sin embargo, son expresiones cerradas y eso permite poder obtenerlas de forma exacta.

Finalmente, la EPPF también puede ser obtenida de forma cerrada y estará dada por,

$$\begin{aligned}
\Pi_k^{(n)}(n_1, \dots, n_k) &= \frac{1}{\Gamma(n)} \int_{\mathbb{R}_+} u^{n-1} e^{-\varphi(u)} \left(\prod_{i=1}^k \int_{\mathbb{R}} \frac{\Gamma(n_i - \sigma)(u + \gamma)^{\sigma - n_i}}{\Gamma(1 - \sigma)} \theta P_0(dx) \right) du \\
&= \frac{\theta^k}{\Gamma(n)} \prod_{i=1}^k (1 - \sigma)_{n_i - 1} \int_{\mathbb{R}_+} u^{n-1} (u + \gamma)^{\sigma k - n} e^{-\frac{\theta}{\sigma}((u + \gamma)^\sigma - \gamma^\sigma)} du \\
&= \frac{\theta^k}{\Gamma(n)} \prod_{i=1}^k (1 - \sigma)_{n_i - 1} \sum_{i=0}^{n-1} \binom{n-1}{i} (-1)^i \beta^{\frac{i}{\sigma}} \Gamma\left(k - \frac{i}{\sigma}; \beta\right) \frac{e^\beta}{\sigma} \left(\frac{\sigma}{\theta}\right)^k \\
&= \frac{\sigma^{k-1} e^\beta}{\Gamma(n)} \prod_{i=1}^k (1 - \sigma)_{n_i - 1} \sum_{i=0}^{n-1} \binom{n-1}{i} (-1)^i \beta^{\frac{i}{\sigma}} \Gamma\left(k - \frac{i}{\sigma}; \beta\right).
\end{aligned}$$

Ahora que ya se tiene una expresión para la EPPF asociada al proceso gama generalizado es importante hacer notar que puede ser escrita como

$$\Pi_k^{(n)} = V_{n,k} \prod_{i=1}^k W_{n_i}.$$

Por lo cual esta familia genera una partición aleatoria de Gibbs y que para el proceso Dirichlet se tiene que

$$V_{n,k} = \frac{\theta^k}{(\theta)_n}, \quad W_{n_i} = \Gamma(n_i).$$

Y para el proceso σ -estable,

$$V_{n,k} = \frac{\sigma^{k-1} \Gamma(k)}{\Gamma(n)}, \quad W_{n_i} = (1 - \sigma)_{n_i - 1}.$$

3.4. Proceso Poisson-Dirichlet (σ, θ)

En las secciones anteriores se han estudiado a fondo 2 de los procesos más utilizados en la estadística bayesiana no paramétrica para la construcción de medidas aleatorias de probabilidad, por lo que la normalización de medidas completamente aleatorias ha probado ser de gran utilidad. Sin embargo, el proceso Poisson-Dirichlet que también es conocido en la literatura como el proceso de Pitman-Yor no puede ser construido de esta forma.

Este proceso fue introducido por Pitman, Yor y Perman en su artículo de 1992 titulado *Size-biased sampling of Poisson point processes and excursions*. Y posteriormente retomado en Pitman y Yor (1997). El proceso Poisson-Dirichlet (σ, θ) pertenece a los modelos de Poisson-Kingman; sin embargo, se puede construir a través de la normalización de una medida que no será completamente aleatoria y su análisis posterior se puede hacer de forma similar a lo realizado con los procesos de las secciones anteriores. Por esto último y por su importancia en la estadística bayesiana no paramétrica es que se hace mención de él en esta tesis.

Para su construcción se toman dos parámetros, $\sigma \in (0, 1)$ y $\theta > -\sigma$. De esta manera si \mathbb{P}_σ define la distribución de probabilidad de una medida completamente aleatoria σ -estable y se toma a $\mathbb{P}_{\sigma, \theta}$ absolutamente continua con respecto a \mathbb{P}_σ y tal que

$$\frac{d\mathbb{P}_{\sigma, \theta}(m)}{d\mathbb{P}_\sigma} = \frac{\Gamma(\theta + 1)}{\Gamma\left(\frac{\theta}{\sigma} + 1\right)} (m(\mathbb{R}))^{-\theta},$$

se tiene un cambio de medida mediante un tilting polinomial. Por lo que si $\Phi \sim \mathbb{P}_{\sigma, \theta}$ entonces Φ no es una medida completamente aleatoria pero es tal que al hacer la normalización

$$\mathcal{P} = \frac{\Phi}{\Phi(\mathbb{R})},$$

se sigue que \mathcal{P} se distribuye como un proceso Poisson-Dirichlet (σ, θ) .

Otro aspecto importante a mencionar del proceso Poisson-Dirichlet es que engloba los siguientes casos:

1. El proceso Dirichlet al hacer $\sigma \downarrow 0$.
2. El proceso σ -estable al hacer $\theta \downarrow 0$.

Por lo cual no será de extrañar que el comportamiento posterior estará relacionado con el del proceso gama generalizado; sin embargo, esto se verá más adelante. Por el momento, se hará énfasis en la obtención de la función de probabilidad para particiones intercambiables y la obtención de la distribución predictiva a través de ella.

A través del tilting polinomial y suponiendo que la distribución de la masa total, $\Phi(\mathbb{R}) = T$ es absolutamente continua con respecto a la medida de Lebesgue se tiene que

$$f_{\sigma, \theta}(T) = \frac{\Gamma(\theta + 1)}{\Gamma\left(\frac{\theta}{\sigma} + 1\right)} T^{-\theta} f_{\sigma}(T).$$

Por lo que la EPPF asociada se puede obtener de la siguiente manera,

$$\Pi_k^{(n)}(n_1, \dots, n_k) = \frac{\Gamma(\theta + 1)(-1)^{n-k}}{\Gamma\left(\frac{\theta}{\sigma} + 1\right) \Gamma(\theta + n)} \int_{\mathbb{R}_+} u^{\theta+n-1} e^{-\varphi(u)} \prod_{j=1}^k \varphi^{(n_j)}(u) du. \quad (3.8)$$

Donde, $\varphi(u) = u^{\sigma}$ es el exponente laplaciano de la distribución σ -estable y $\varphi^{(n_j)}$ denota la n_j -ésima derivada. Este resultado puede verificarse en Ghosal y van der Vaart (2017) y que por tratarse de una expresión relacionada con los procesos Poisson-Kingman no se entrará en detalle a su demostración; sin embargo, si se obtendrá una expresión cerrada para ella. Para calcularla primero se debe obtener una expresión para la n_j -ésima derivada del exponente laplaciano.

$$\begin{aligned}
\varphi^{(1)}(u) &= \sigma u^{\sigma-1} \\
\varphi^{(2)}(u) &= \sigma(\sigma-1)u^{\sigma-2} = (-1)\sigma(1-\sigma)u^{\sigma-2} \\
&\vdots \\
\varphi^{(n_j)}(u) &= (-1)^{n_j-1}\sigma(1-\sigma)_{n_j-1}u^{\sigma-n_j}.
\end{aligned}$$

Por lo que,

$$\prod_{j=1}^k \varphi^{(n_j)}(u) = (-1)^{n-k} \sigma^k u^{\sigma k-n} \prod_{j=1}^k (1-\sigma)_{n_j-1}.$$

Sustituyendo en la expresión (3.8) se tiene que,

$$\begin{aligned}
\Pi_k^{(n)}(n_1, \dots, n_k) &= \frac{\Gamma(\theta+1)}{\Gamma\left(\frac{\theta}{\sigma}+1\right)\Gamma(\theta+n)} \int_{\mathbb{R}_+} u^{\theta+n-1} e^{-\varphi(u)} \sigma^k u^{\sigma k-n} \prod_{j=1}^k (1-\sigma)_{n_j-1} du \\
&= \frac{\Gamma(\theta+1)\sigma^k}{\Gamma\left(\frac{\theta}{\sigma}+1\right)\Gamma(\theta+n)} \prod_{j=1}^k (1-\sigma)_{n_j-1} \int_{\mathbb{R}_+} u^{\theta+\sigma k-1} e^{-u^\sigma} du \\
&\quad (\text{haciendo } w = u^\sigma) \\
&= \frac{\Gamma(\theta+1)\sigma^{k-1}}{\Gamma\left(\frac{\theta}{\sigma}+1\right)\Gamma(\theta+n)} \Gamma\left(\frac{\theta}{\sigma}+k\right) \prod_{j=1}^k (1-\sigma)_{n_j-1} \\
&= \frac{\Gamma(\theta+1)\sigma^{k-1}}{\Gamma\left(\frac{\theta}{\sigma}+1\right)\Gamma(\theta+1+n-1)} \Gamma\left(\frac{\theta}{\sigma}+1+k-1\right) \prod_{j=1}^k (1-\sigma)_{n_j-1} \quad (3.9)
\end{aligned}$$

En esta última expresión se puede observar que

$$\frac{\Gamma(\theta+1)}{\Gamma(\theta+1+n-1)} = (\theta+1)_{(n-1)}.$$

Que representa el símbolo de Pochhammer que ya se había definido anteriormente. Y donde

$$\begin{aligned}
\frac{1}{\Gamma\left(\frac{\theta}{\sigma} + 1\right)} \Gamma\left(\frac{\theta}{\sigma} + 1 + k - 1\right) &= \left(\frac{\theta}{\sigma} + 1\right) \left(\frac{\theta}{\sigma} + 2\right) \cdots \left(\frac{\theta}{\sigma} + k - 1\right) \\
&= \left(\frac{\theta + \sigma}{\sigma}\right) \left(\frac{\theta + 2\sigma}{\sigma}\right) \cdots \left(\frac{\theta + \sigma(k-1)}{\sigma}\right) \\
&= \frac{1}{\sigma^{k-1}} \prod_{i=1}^{k-1} (\theta + i\sigma)
\end{aligned}$$

Por lo que, sustituyendo en (3.9) se obtiene finalmente que

$$\Pi_k^{(n)}(n_1, \dots, n_k) = \frac{\prod_{i=1}^{k-1} (\theta + i\sigma)}{(\theta + 1)_{(n-1)}} \prod_{j=1}^k (1 - \sigma)_{n_j - 1}$$

De esta manera se tiene una expresión cerrada para la EPPF, que se puede observar genera una partición aleatoria de Gibbs con

$$V_{n,k} = \frac{\prod_{i=1}^{k-1} (\theta + i\sigma)}{(\theta + 1)_{(n-1)}}, \quad W_{n_i} = (1 - \sigma)_{n_i - 1}$$

y que además permite recuperar la distribución predictiva, ya que

$$\begin{aligned}
\mathbb{P}(X_{n+1} \neq Y_j \quad \forall j \in \{1, \dots, k\} | \mathbf{X}^{(n)}) &= \frac{\Pi_{k+1}^{(n+1)}(n_1, \dots, n_k, 1)}{\Pi_k^{(n)}(n_1, \dots, n_k)} \\
\mathbb{P}(X_{n+1} = Y_j | \mathbf{X}^{(n)}) &= \frac{\Pi_k^{(n)}(n_1, \dots, n_{j-1}, n_j + 1, n_{j+1}, \dots, n_k)}{\Pi_k^{(n)}(n_1, \dots, n_k)}.
\end{aligned}$$

Por lo que mediante unos cálculos sencillos se obtiene que

$$\mathbb{P}(X_{n+1} \neq Y_j \quad \forall j \in \{1, \dots, k\} | \mathbf{X}^{(n)}) = \frac{\theta + k\sigma}{\theta + n}.$$

Y

$$\mathbb{P}(X_{n+1} = Y_j | \mathbf{X}^{(n)}) = \frac{n_j - \sigma}{\theta + n}.$$

Y así la distribución predictiva tiene la siguiente representación,

$$\mathbb{P}(X_{n+1} \in dx | \mathbf{X}^{(n)}) = \frac{\theta + k\sigma}{\theta + n} P_0(dx) + \frac{1}{\theta + n} \sum_{i=1}^k (n_i - \sigma) \delta_{Y_i}(dx).$$

Que es una expresión bastante sencilla de trabajar y para la cual se tiene un algoritmo de muestreo de Backwell-MacQueen bastante parecido al proceso σ -estable, lo cual no es una sorpresa y cuyo algoritmo es:

1. Simular X_1 de P_0 .
2. Al momento $i + 1$, se denota por $Y_{1,i}, \dots, Y_{k_i,i}$ a los distintos valores observados en la colección X_1, \dots, X_i y con respectivas frecuencias $n_{1,i}, \dots, n_{k_i,i}$.
 - a. Simular ϵ_{i+1} de P_0 .

$$\text{b. } X_{i+1} = \begin{cases} \epsilon_{i+1} & \text{con probabilidad } \frac{\theta+k\sigma}{\theta+n} \\ Y_{1,i} & \text{con probabilidad } \frac{n_{1,i}-\sigma}{\theta+n} \\ \vdots & \\ Y_{k_i,i} & \text{con probabilidad } \frac{n_{k_i,i}-\sigma}{\theta+n}. \end{cases}$$

A partir de la distribución predictiva se puede observar que a diferencia del proceso Dirichlet, que es un caso particular, habrá dependencia en el número de grupos, lo cual proporciona mayor flexibilidad al momento de modelar fenómenos. Aunado a que tiene una estructura más sencilla que la del proceso gama generalizado, provoca que el proceso Poisson-Dirichlet tenga tanta importancia hoy en día en la estadística bayesiana no paramétrica.

Como ya se mencionó, el proceso Poisson-Dirichlet no pertenece a la clase de las NRMI; sin embargo, la distribución posterior puede ser estudiada de una manera similar a lo que se ha venido haciendo para los otros procesos, i.e., se puede escribir como una mezcla con respecto a una variable latente.

Para esto, sea U_k una variable aleatoria con densidad dada por

$$f_{\sigma, \theta, k}(u) = \frac{\sigma}{\Gamma\left(\frac{\theta}{\sigma} + k\right)} u^{\theta + k\sigma - 1} e^{-u^\sigma}.$$

Entonces se puede mostrar que

$$\Phi^{(U_k, \mathbf{X}^{(n)})} \stackrel{d}{=} \Phi^{(U_k)} + \sum_{i=1}^k J_i^{(U_k, \mathbf{X}^{(n)})} \delta_{Y_i}.$$

Donde $\Phi^{(U_k)}$ es un proceso gama generalizado con intensidad de Lévy dada por

$$\nu^{(U_k)}(ds, dx) = \frac{\sigma}{\Gamma(1 - \sigma)} s^{-1 - \sigma} e^{-U_k s} ds P_0(dx).$$

Y donde los brincos $J_i^{(U_k, \mathbf{X}^{(n)})}$ con localizaciones en las observaciones Y_i son variables aleatorias gama independientes de parámetros $(n_i - \sigma, U_k)$ y que son condicionalmente independientes de $\Phi^{(U_k)}$ dada la variable latente U_k . Que es una estructura similar a la que se vio para el proceso gama generalizado.

Capítulo 4

Medida completamente aleatoria de Bessel

En el capítulo anterior se estudiaron las medidas de probabilidad aleatorias asociadas a algunos de los procesos más utilizados en la literatura como lo son el proceso Dirichlet, el proceso σ -estable y el proceso Poisson-Dirichlet. Se pudo observar que muchas de las expresiones del análisis posterior como lo son la distribución predictiva y la función de probabilidad para particiones intercambiables tienen soluciones cerradas, lo cual provoca que sea sencillo el proceso de muestreo de estas medidas.

Una característica importante que comparten estos procesos es que todos tienen una medida de Lévy homogénea asociada, por lo cual pudiera pensarse que esto es lo que provoca que se tengan expresiones cerradas; sin embargo, como se verá a continuación, no tiene por qué ser cierto en todos los casos.

En este último capítulo se estudiará a la medida completamente aleatoria de Bessel que fue introducida por Argiento, Bianchini y Guglielmi (2016), la cual tiene una medida de Lévy homogénea y para la cual no se cuenta con expresiones cerradas en su análisis posterior, lo que sin duda recalca la importancia de la computación para la estadística bayesiana no paramétrica.

La medida completamente aleatoria de Bessel considera a la siguiente intensidad de Lévy

$$\nu(ds, dx) = \frac{1}{s} e^{-\omega s} I_\nu(s) ds \alpha(dx).$$

Donde $\omega > 1$ e $I_\nu(s)$ denota a la función modificada de Bessel de orden ν , que está definida como

$$I_\nu(s) = \sum_{i=0}^{\infty} \frac{\left(\frac{s}{2}\right)^{2i+\nu}}{i! \Gamma(\nu + i + 1)}.$$

De nueva cuenta la medida $\alpha(dx)$ puede ser considerada como un parámetro, la cual se debe elegir que sea no negativa y finita, por lo cual se puede centrar el análisis en la intensidad de los saltos

$$\rho(ds) = \frac{1}{s} e^{-\omega s} I_0(s) ds.$$

La cual puede ser expresada como

$$\rho(ds) = \frac{1}{s} e^{-\omega s} ds + \sum_{i=1}^{\infty} \frac{s^{2i-1} e^{-\omega s}}{2^{2i} \Gamma(i+1)^2} ds,$$

que representa la suma de procesos gama. Una de las ventajas en expresar de esta manera la intensidad de los saltos es que permite verificar con facilidad las condiciones que se piden para ρ . Para esto primero se debe verificar que $\rho(\mathbb{R}_+) = \infty$, lo cual ocurre ya que

$$\rho(\mathbb{R}_+) = \int_{\mathbb{R}_+} \frac{1}{s} e^{-\omega s} ds + \int_{\mathbb{R}_+} \sum_{i=1}^{\infty} \frac{s^{2i-1} e^{-\omega s}}{2^{2i} \Gamma(i+1)^2} ds$$

Con

$$\int_{\mathbb{R}_+} \frac{1}{s} e^{-\omega s} ds = \infty.$$

Y

$$\begin{aligned}
\int_{\mathbb{R}_+} \sum_{i=1}^{\infty} \frac{s^{2i-1} e^{-\omega s}}{2^{2i} \Gamma(i+1)^2} ds &= \sum_{i=1}^{\infty} \frac{1}{2^{2i} \Gamma(i+1)^2} \int_{\mathbb{R}_+} s^{2i-1} e^{-\omega s} ds \\
&= \sum_{i=1}^{\infty} \frac{\Gamma(2i)}{\Gamma(i+1)^2 (2\omega)^{2i}} \\
&= -\log \left(\frac{1}{2} \left(\sqrt{\frac{\omega^2 - 1}{\omega^2}} + 1 \right) \right).
\end{aligned}$$

Que es finito ya que $\omega > 1$, por lo cual se tiene que $\rho(\mathbb{R}_+) = \infty$ y así se tiene una medida con actividad infinita. Otra condición a verificar es la de integrabilidad para ν , que se reduce a pedir que

$$\int_{\mathbb{R}_+} (s \wedge 1) \rho(ds) < \infty.$$

Lo cual es cierto ya que claramente se tiene que

$$\int_{\mathbb{R}_+} (s \wedge 1) \frac{1}{s} e^{-\omega s} ds = \int_0^1 e^{-\omega s} ds + \int_1^{\infty} \frac{1}{s} e^{-\omega s} ds < \infty$$

Y de forma similar a lo realizado para verificar que $\rho(\mathbb{R}_+) = \infty$ se tiene que

$$\int_0^1 \sum_{i=1}^{\infty} \frac{s^{2i} e^{-\omega s}}{\Gamma(i+1)^2 2^{2i}} ds < \sum_{i=1}^{\infty} \frac{\Gamma(2i+1)}{\Gamma(i+1)^2 2^{2i} \omega^{2i+1}} = \frac{\omega - \sqrt{\omega^2 - 1}}{\omega^2 \sqrt{\omega^2 - 1}} < \infty,$$

y

$$\int_1^{\infty} \sum_{i=1}^{\infty} \frac{s^{2i-1} e^{-\omega s}}{2^{2i} \Gamma(i+1)^2} ds < \sum_{i=1}^{\infty} \frac{\Gamma(2i)}{\Gamma(i+1)^2 (2\omega)^{2i}} = -\log \left(\frac{1}{2} \left(\sqrt{\frac{\omega^2 - 1}{\omega^2}} + 1 \right) \right) < \infty.$$

Por lo cual,

$$\int_{\mathbb{R}_+} (s \wedge 1) \sum_{i=1}^{\infty} \frac{s^{2i-1} e^{-\omega s}}{2^{2i} \Gamma(i+1)^2} ds < \infty.$$

Por lo que la condición de integrabilidad se cumple para esta intensidad de Lévy. El último paso antes de proceder al análisis posterior es la obtención del exponente laplaciano y verificar que sea finito, para lo cual se puede obtener primero la funcional de Laplace.

Para esto, sea f , una función medible y si Φ es la medida completamente aleatoria de Bessel se tiene que

$$\mathbb{E} \left[\exp \left(- \int_{\mathbb{R}} f(x) \Phi(dx) \right) \right] = \exp \left[- \int_{\mathbb{R}} \int_{\mathbb{R}_+} (1 - e^{-sf(x)}) \rho(ds) \alpha(dx) \right]$$

Donde

$$\begin{aligned} \int_{\mathbb{R}_+} (1 - e^{-sf(x)}) \rho(ds) &= \int_{\mathbb{R}_+} (1 - e^{-sf(x)}) \left(\frac{1}{s} e^{-\omega s} ds + \sum_{i=1}^{\infty} \frac{s^{2i-1} e^{-\omega s}}{\Gamma(i+1)^2 2^{2i}} ds \right) \\ &= \int_{\mathbb{R}_+} (1 - e^{-sf(x)}) \frac{e^{-\omega s}}{s} ds + \int_{\mathbb{R}_+} (1 - e^{-sf(x)}) \left(\sum_{i=1}^{\infty} \frac{s^{2i-1} e^{-\omega s}}{\Gamma(i+1)^2 2^{2i}} \right) ds \end{aligned} \quad (4.1)$$

La primer integral del lado derecho de la ecuación (4.1) se resuelve de manera similar a lo realizado para el proceso Dirichlet, ya que

$$\begin{aligned} \int_{\mathbb{R}_+} (1 - e^{-f(x)s}) \frac{e^{-\omega s}}{s} ds &= \int_{\mathbb{R}_+} \sum_{j=1}^{\infty} \frac{(-1)^{j+1} (f(x)s)^j}{j!} \cdot \frac{e^{-\omega s}}{s} ds \\ &= \sum_{j=1}^{\infty} \frac{(-1)^{j+1} f(x)^j}{j!} \int_{\mathbb{R}_+} s^{j-1} e^{-\omega s} ds \\ &= \sum_{j=1}^{\infty} \frac{(-1)^{j+1} f(x)^j \Gamma(j)}{j! \omega^j} \\ &= \log \left(1 + \frac{f(x)}{\omega} \right) \\ &= \log \left(\frac{\omega + f(x)}{\omega} \right) \end{aligned} \quad (4.2)$$

Y donde para la segunda integral se tiene que

$$\begin{aligned} &\int_{\mathbb{R}_+} (1 - e^{-sf(x)}) \left(\sum_{i=1}^{\infty} \frac{s^{2i-1} e^{-\omega s}}{\Gamma(i+1)^2 2^{2i}} \right) ds \\ &= \int_{\mathbb{R}_+} \sum_{i=1}^{\infty} \frac{s^{2i-1} e^{-\omega s}}{\Gamma(i+1)^2 2^{2i}} ds - \int_{\mathbb{R}_+} \sum_{i=1}^{\infty} \frac{s^{2i-1} e^{-s(f(x)+\omega)}}{\Gamma(i+1)^2 2^{2i}} ds \end{aligned} \quad (4.3)$$

Las integrales de la expresión (4.3) son sencillas de calcular y que ya se obtuvieron anteriormente, por lo que

$$\int_{\mathbb{R}_+} \sum_{i=1}^{\infty} \frac{s^{2i-1} e^{-\omega s}}{\Gamma(i+1)^2 2^{2i}} ds = \sum_{i=1}^{\infty} \frac{\Gamma(2i)}{\Gamma(i+1)^2 (2\omega)^{2i}} = -\log \left(\frac{1}{2} \left(\sqrt{\frac{\omega^2 - 1}{\omega^2}} + 1 \right) \right),$$

y

$$\begin{aligned} - \int_{\mathbb{R}_+} \sum_{i=1}^{\infty} \frac{s^{2i-1} e^{-s(f(x)+\omega)}}{\Gamma(i+1)^2 2^{2i}} s &= - \sum_{i=1}^{\infty} \frac{\Gamma(2i)}{\Gamma(i+1)^2 (2(\omega + f(x)))^{2i}} \\ &= \log \left(\frac{1}{2} \left(\sqrt{\frac{(\omega + f(x))^2 - 1}{(\omega + f(x))^2}} + 1 \right) \right). \end{aligned}$$

Por lo que utilizando las ecuaciones (4.2) y (4.3) se obtiene que (4.1) es igual a

$$\begin{aligned} &\log \left(\frac{\omega + f(x)}{\omega} \right) - \log \left(\frac{1}{2} \left(\sqrt{\frac{\omega^2 - 1}{\omega^2}} + 1 \right) \right) + \log \left(\frac{1}{2} \left(\sqrt{\frac{(\omega + f(x))^2 - 1}{(\omega + f(x))^2}} + 1 \right) \right) \\ &= \log \left(\frac{\left(\frac{\omega + f(x)}{\omega} \right) \left(\sqrt{\frac{(\omega + f(x))^2 - 1}{(\omega + f(x))^2}} + 1 \right)}{\left(\sqrt{\frac{\omega^2 - 1}{\omega^2}} + 1 \right)} \right) \\ &= \log \left(\frac{f(x) + \omega + \sqrt{(f(x) + \omega)^2 - 1}}{\omega + \sqrt{\omega^2 - 1}} \right). \end{aligned}$$

De esta manera la funcional de Laplace está dada por

$$\exp \left[-\theta \int_{\mathbb{R}} \log \left(\frac{f(x) + \omega + \sqrt{(f(x) + \omega)^2 - 1}}{\omega + \sqrt{\omega^2 - 1}} \right) P_0(dx) \right]. \quad (4.4)$$

Y a partir de ella podemos obtener el exponente laplaciano

$$\varphi(u) = \log \left(\frac{u + \omega + \sqrt{(u + \omega)^2 - 1}}{\omega + \sqrt{\omega^2 - 1}} \right)^\theta < \infty.$$

Con lo cual se cumplen todas las condiciones necesarias para poder definir una medida de probabilidad aleatoria mediante la normalización del proceso de Bessel. Más aún, de la ecuación (4.4) se puede caracterizar la distribución de $\Phi(B)$ a través de su transformada de Laplace. Tomando a $f = \lambda \mathbb{1}_B$, se obtiene que

$$\mathbb{E}(e^{-\lambda\Phi(B)}) = \left(\frac{\lambda + \omega + \sqrt{(\lambda + \omega)^2 - 1}}{\omega + \sqrt{\omega^2 - 1}} \right)^{-\theta P_0(B)}.$$

Para el análisis posterior se debe recordar que

$$\Phi^{(U_n, \mathbf{X}^{(n)})} \stackrel{d}{=} \Phi^{(U_n)} + \sum_{i=1}^{n(\pi)} J_i^{(U_n, \mathbf{X}^{(n)})} \delta_{Y_i}.$$

Donde $\Phi^{(U_n)}$ tiene una intensidad de Lévy asociada dada por

$$\begin{aligned} \nu^{(U_n)}(ds, dx) &= e^{-U_n s} \rho(ds) \alpha(dx) \\ &= \left(\frac{1}{s} e^{-(\omega+U_n)s} + \sum_{i=1}^{\infty} \frac{s^{2i-1} e^{-(\omega+U_n)s}}{\Gamma(i+1)^2 2^{2i}} \right) ds \alpha(dx) \\ &= \frac{1}{s} e^{-(\omega+U_n)s} I_0(s) ds \alpha(dx). \end{aligned}$$

Que tiene la misma estructura que $\nu(ds, dx)$. Y donde los saltos $J_i^{(U_n, \mathbf{X}^{(n)})}$ tienen una densidad $f_j(s)$ tal que

$$\begin{aligned} f_j(s) &\propto s^{n_j} e^{-U_n s} \rho(ds) \\ &= s^{n_j-1} e^{-(\omega+U_n)s} + \sum_{i=1}^{\infty} \frac{s^{2i+n_j-1} e^{-(\omega+U_n)s}}{\Gamma(i+1)^2 2^{2i}} \\ &= \sum_{i=0}^{\infty} \frac{s^{2i+n_j-1} e^{-(\omega+U_n)s}}{\Gamma(i+1)^2 2^{2i}}. \end{aligned}$$

Por lo que la constante de normalización es igual a

$$\begin{aligned}
\int_{\mathbb{R}_+} \sum_{i=0}^{\infty} \frac{s^{2i+n_j-1} e^{-(\omega+U_n)s}}{\Gamma(i+1)^2 2^{2i}} &= \sum_{i=0}^{\infty} \frac{\Gamma(2i+n_j)}{\Gamma(i+1)^2 2^{2i} (\omega+U_n)^{2i+n_j}} \\
&= \frac{\Gamma(n_j)}{(\omega+U_n)^{n_j}} \sum_{i=0}^{\infty} \frac{\Gamma(2i+n_j)}{2^{2i} \Gamma(n_j)} \frac{1}{\Gamma(i+1)^2} \frac{1}{(\omega+U_n)^{2i}} \quad (4.5)
\end{aligned}$$

Donde se puede notar que

$$\begin{aligned}
\left(\frac{n_j}{2}\right)_i \cdot \left(\frac{n_j+1}{2}\right)_i &= \prod_{l=0}^{i-1} \left(\frac{n_j}{2} + l\right) \prod_{l=0}^{i-1} \left(\frac{n_j+1}{2} + l\right) \\
&= \prod_{l=0}^{i-1} \left(\frac{n_j+2l}{2}\right) \prod_{l=0}^{i-1} \left(\frac{n_j+1+2l}{2}\right) \\
&= \frac{1}{2^{2i}} \prod_{l=0}^{2i-1} (n_j + l) \\
&= \frac{\Gamma(n_j+2i)}{2^{2i} \Gamma(n_j)}.
\end{aligned}$$

Por lo que la ecuación (4.5) se puede reescribir como

$$\begin{aligned}
&\frac{\Gamma(n_j)}{(\omega+U_n)^{n_j}} \sum_{i=0}^{\infty} \left(\frac{n_j}{2}\right)_i \left(\frac{n_j+1}{2}\right)_i \frac{1}{\Gamma(i+1)^2} \frac{1}{(\omega+U_n)^{2i}} \\
&= \frac{\Gamma(n_j)}{(\omega+U_n)^{n_j}} {}_2F_1\left(\frac{n_j}{2}, \frac{n_j+1}{2}; 1; \frac{1}{(\omega+U_n)^2}\right).
\end{aligned}$$

Donde ${}_2F_1(a, b; c; z)$ representa la función hipergeométrica. De esta manera los saltos resultan ser una mezcla de variables aleatorias gammas, por lo cual obtener una caracterización de $\Phi^{(U_n, \mathbf{X}^{(n)})}$ a través de su transformada de Laplace ya no resulta tan sencillo como en casos anteriores y llevará a resultados difíciles de interpretar en comparación con las medidas del Capítulo 3. Sin embargo, es claro que puede hacerse y que una expresión numérica puede encontrarse para la función hipergeométrica cuando se tienen parámetros fijos.

Dado lo anterior, se continuará con el análisis posterior, para lo cual se necesitan los siguiente elementos:

$$e^{-\varphi(u)} = \left(\frac{u + \omega + \sqrt{(u + \omega)^2 - 1}}{\omega + \sqrt{\omega^2 - 1}} \right)^{-\theta},$$

$$\tau_{n_j}(u) = \int_{\mathbb{R}_+} s^{n_j} e^{-us} \rho(ds) = \frac{\Gamma(n_j)}{(\omega + u)^{n_j}} {}_2F_1 \left(\frac{n_j}{2}, \frac{n_j + 1}{2}; 1; \frac{1}{(\omega + u)^2} \right),$$

por lo que

$$\prod_{j=1}^{n(\pi)} \tau_{n_j}(u) = \frac{1}{(\omega + u)^n} \prod_{j=1}^{n(\pi)} \Gamma(n_j) {}_2F_1 \left(\frac{n_j}{2}, \frac{n_j + 1}{2}; 1; \frac{1}{(\omega + u)^2} \right).$$

De esta manera se sigue que la distribución condicional de la variable latente U_n dada la muestra es tal que

$$\begin{aligned} f_{U_n|\mathbf{X}^{(n)}}(u) &\propto u^{n-1} \prod_{j=1}^{n(\pi)} \tau_{n_j}(u) e^{-\varphi(u)} \\ &= \frac{u^{n-1}}{(\omega + u)^n} \cdot \prod_{j=1}^{n(\pi)} \Gamma(n_j) {}_2F_1 \left(\frac{n_j}{2}, \frac{n_j + 1}{2}; 1; \frac{1}{(\omega + u)^2} \right) \\ &\quad \cdot \left(\frac{u + \omega + \sqrt{(u + \omega)^2 - 1}}{\omega + \sqrt{\omega^2 - 1}} \right)^{-\theta}. \end{aligned}$$

En el Capítulo 2 quedó de manifiesto la importancia de $f_{U_n|\mathbf{X}^{(n)}}$ para la obtención de los pesos que permiten caracterizar a la distribución predictiva, ya que se puede recordar que están dados por

$$\begin{aligned} \omega^{(n)} &= \frac{\theta}{n} \int_{\mathbb{R}_+} u \tau_1(u|x) f_{U_n|\mathbf{X}^{(n)}}(u) du \\ \omega_i^{(n)} &= \int_{\mathbb{R}_+} u \frac{\tau_{n_i+1}(u|Y_i)}{\tau_{n_i}(u|Y_i)} f_{U_n|\mathbf{X}^{(n)}}(u) du. \end{aligned}$$

Ahora bien, no es difícil darse cuenta que la obtención de dichas integrales se tiene que hacer de manera numérica, por lo cual en este caso no se cuenta con expresiones cerradas.

Esto mismo ocurre cuando se trabaja con la EPPF, ya que está dada por

$$\begin{aligned}
\Pi_k^{(n)}(n_1, \dots, n_k) &= \frac{1}{\Gamma(n)} \int_{\mathbb{R}_+} u^{n-1} e^{-\varphi(u)} \prod_{j=1}^k \int_{\mathbb{R}} \tau_{n_j}(u|x) \alpha(dx) du \\
&= \frac{\theta^k}{\Gamma(n)} \int_{\mathbb{R}_+} u^{n-1} \left(\frac{\omega + \sqrt{\omega^2 - 1}}{u + \omega + \sqrt{(u + \omega)^2 - 1}} \right)^\theta \frac{1}{(\omega + u)^n} \\
&\quad \times \prod_{j=1}^k \Gamma(n_j) {}_2F_1 \left(\frac{n_j}{2}, \frac{n_j + 1}{2}; 1; \frac{1}{(\omega + u)^2} \right) du
\end{aligned}$$

Por lo cual siguiendo este acercamiento es imposible simplificar más las expresiones. Aunado a lo anterior las rutinas numéricas también se vuelven más complejas, lo cual resalta la importancia de la computación para esta rama de la estadística. Sin embargo, no es todo lo que se puede hacer, como ya se vio en el Capítulo 2 cuando no se tienen expresiones cerradas para la EPPF, se puede encontrar una versión de ella condicionada a la variable latente que permitía expresarla como la de una partición de Gibbs.

De esta manera se puede recordar que

$$\Pi^{(n)}(\pi|u) = \frac{e^{-\varphi(u)} \prod_{j=1}^n \kappa_{n_j}(u)}{\int_{\mathbb{R}_+} t^n e^{-ut} f_T(t) dt},$$

donde

$$\kappa_{n_j}(u) = \int_{\mathbb{R}} \tau_{n_j}(u|x) \alpha(dx).$$

Y como se está trabajando con una medida de Lévy homogénea se tiene que τ_{n_j} no depende de x , por lo que $\kappa_{n_j}(u) = \theta \tau_{n_j}(u)$. Por lo que recordando la ecuación (2.13) se tiene que para la medida completamente aleatoria de Bessel

$$\mathbb{P}(X_{n+1} \in dx | \mathbf{X}^{(n)}, U_n = u) \propto \kappa_1(u) \tau_1(u) P_0(dx) + \sum_{j=1}^k \frac{\tau_{n_j+1}(u)}{\tau_{n_j}(u)} \delta_{Y_j}(dx).$$

Donde

$$\begin{aligned}
\tau_1(u) &= \frac{1}{(\omega + u)} \cdot {}_2F_1(.5, 1; 1; (\omega + u)^2) \\
&= \frac{1}{\omega + u} \cdot \frac{1}{\left(1 - \frac{1}{(\omega + u)^2}\right)^{1/2}} \\
&= \frac{1}{\omega + u} \cdot \left(\frac{(\omega + u)}{\sqrt{(\omega + u)^2 - 1}}\right) \\
&= \frac{1}{\sqrt{(\omega + u)^2 - 1}}.
\end{aligned}$$

Y de manera inmediata se sabe que

$$\kappa_1(u) = \frac{\theta}{\sqrt{(\omega + u)^2 - 1}}.$$

Por lo que estaremos obteniendo una nueva observación con un peso proporcional a

$$\kappa_1(u)\tau_1(u) = \frac{\theta}{(\omega + u)^2 - 1}.$$

Y donde se tendrá una observación repetida con un peso proporcional a

$$\begin{aligned}
\frac{\tau_{n_j+1}(u)}{\tau_{n_j}(u)} &= \frac{\frac{\Gamma(n_j+1)}{(\omega+u)^{n_j+1}} \cdot {}_2F_1\left(\frac{n_j+1}{2}, \frac{n_j+2}{2}; 1; \frac{1}{(\omega+u)^2}\right)}{\frac{\Gamma(n_j)}{(\omega+u)^{n_j}} \cdot {}_2F_1\left(\frac{n_j}{2}, \frac{n_j+1}{2}; 1; \frac{1}{(\omega+u)^2}\right)} \\
&= \frac{n_j}{(\omega + u)} \frac{{}_2F_1\left(\frac{n_j+1}{2}, \frac{n_j+2}{2}; 1; \frac{1}{(\omega+u)^2}\right)}{{}_2F_1\left(\frac{n_j}{2}, \frac{n_j+1}{2}; 1; \frac{1}{(\omega+u)^2}\right)}
\end{aligned}$$

Y con lo anterior realizar un esquema de muestreo como el presentado en la página 49. Es importante señalar que ambos acercamientos traerán consigo complejidades al momento de establecer la rutina numérica, caso totalmente contrario al del proceso Dirichlet, σ -estable y Poisson-Dirichlet; sin embargo, se pueden realizar y como prueba de ello está la aplicación presente en el trabajo de Argiento, Bianchini y Guglielmi (2016).

Finalmente y como último comentario acerca de la medida completamente aleatoria de Bessel, se tiene la siguiente proposición que es ciertamente interesante y cuya demostración se puede ver en el trabajo de Argiento, Bianchini y Guglielmi (2016).

Proposición 4.1. *Sea (n_1, \dots, n_k) una composición de n . Entonces la EPPF asociada a la medida de probabilidad aleatoria de Bessel es tal que*

$$\lim_{\omega \rightarrow +\infty} \Pi_k^{(n)}(n_1, \dots, n_k)_{Bessel} = \Pi_k^{(n)}(n_1, \dots, n_k)_{Dirichlet}$$

Por lo que se puede apreciar el porque el proceso Dirichlet es considerado como uno de los procesos clave dentro de la estadística bayesiana no paramétrica.

Conclusiones

Teóricamente, la estadística bayesiana no paramétrica es una gran alternativa a la hora de manejar grandes cantidades de datos ya que se trabaja con espacios infinito dimensionales, lo cual brinda una nueva forma de modelar problemas de interés en áreas como la genética y las finanzas. Sin embargo, no fue hasta años recientes y con ayuda del gran avance en la computación que se ha podido explotar de manera más amplia toda la teoría de esta rama de la estadística.

Uno de los procesos más importantes es el proceso Dirichlet introducido por Ferguson (1973), siendo clave para el posterior desarrollo de la estadística bayesiana no paramétrica. Para su construcción se pueden seguir varios caminos, siendo uno de ellos la normalización de un subordinador gama cuando se trabaja en la recta real. Este acercamiento no se retomó hasta años después y fue el tema principal de este trabajo en donde se cubrió la teoría sobre el denominado análisis posterior y se estudiaron algunas de las medidas que se pueden construir de esta manera.

La normalización de medidas completamente aleatorias es una forma de construir medidas de probabilidad aleatorias en espacios generales y se ha podido ver a lo largo de estos cuatro capítulos aspectos teóricos de gran interés para la estadística bayesiana no paramétrica. Se pudieron observar ventajas y desventajas; siendo una de las ventajas el hecho de que algunos de los procesos más utilizados en la literatura pueden ser construidos de esta manera; sin embargo, no engloba a todos y como ejemplo de esto se tiene al proceso Poisson-Dirichlet.

Otro aspecto importante a mencionar es el hecho de que los procesos estudiados a lo largo de los capítulos 3 y 4 pertenecen a la clase caracterizada por tener asociada una medida de Lévy homogénea; lo que en algunos casos permitió encontrar expresiones cerradas y fáciles de interpretar para la distribución predictiva, lo cual a su vez permitió establecer algoritmos de muestreo sencillos del tipo Blackwell-MacQueen. Esto último en particular para el proceso Dirichlet, el proceso σ -estable y el Poisson-Dirichlet. Sin embargo, no siempre se puede garantizar esto para los procesos con medidas homogéneas tal y como se vio para el proceso de Bessel, lo cual hace énfasis en la importancia de la computación para la estadística bayesiana no paramétrica.

También se pudo hacer notar la importancia de la función de probabilidad para particiones intercambiables (EPPF), ya que con ella se puede caracterizar el sistema de ecuaciones predictivas, lo que ciertamente facilita para algunos procesos, como el Poisson-Dirichlet, la obtención de los pesos de la distribución predictiva. Sin entrar en mucho detalle se vio que algunas de estas funciones de probabilidad para particiones intercambiables pertenecen a una clase denominada del tipo Gibbs siendo esto el objetivo de condicionar con la variable latente cuando no se tuvieran expresiones cerradas.

Es claro que aún queda mucho por hacer y mucha teoría por desarrollar. Pero de la misma manera se está avanzado a un ritmo formidable. Ya no sólo se trabajan con los procesos clásicos, se están buscando nuevas medidas y ciertamente hay mucho margen para hacerlo. El proceso de Bessel es un claro ejemplo de ello al considerar una mezcla infinita de procesos gamas. La estadística bayesiana no paramétrica va más allá de la normalización de medidas completamente aleatorias, por lo cual es clave para un correcto desarrollo el completo entendimiento de los demás acercamientos.

Apéndice A

Teorema de representación

El teorema de representación para variables aleatorias intercambiables tiene un papel fundamental dentro de la estadística bayesiana, por lo cual este apéndice está dedicado únicamente a su demostración. Este resultado corresponde al Teorema 1.4 del Capítulo 1 cuyo enunciado es el siguiente.

Teorema A.1. *Teorema de representación*

Sea $X^{(\infty)} = \{X_i\}_{i=1}^{\infty}$ una colección de variables aleatorias definidas en un espacio de probabilidad $(\Omega, \mathcal{F}, \mathbb{P})$, con valores en un espacio polaco \mathbb{X} que está dotado con su σ -álgebra de Borel \mathcal{X} . Entonces $X^{(\infty)}$ se dice intercambiable si y sólo si existe una medida \mathcal{Q} en $\mathcal{P}_{\mathbb{X}}$ tal que para toda $n \geq 1$ y $A_1, \dots, A_n \in \mathcal{X}$ se cumple que

$$\mathbb{P}(X_1 \in A_1, \dots, X_n \in A_n) = \int_{\mathcal{P}_{\mathbb{X}}} \prod_{i=1}^n P(A_i) \mathcal{Q}(dP).$$

Donde $\mathcal{P}_{\mathbb{X}}$ denota al espacio de todas las medidas de probabilidad sobre $(\mathbb{X}, \mathcal{X})$.

Esta es una de las muchas formas en las que se puede encontrar enunciado el teorema de representación y al cual se dará una demostración basada en teoría de martingalas y basada en el artículo de Kingman (1978) y en los libros de Klenke (2014) y Durrett (2010).

Demostración. Dada la sucesión de variables aleatorias $\mathbf{X} = \{X_i\}_{i=1}^{\infty}$ se dice que $\phi(\mathbf{X})$ es una variable aleatoria n -simétrica si es invariante ante permutaciones de las primeras n -entradas del vector aleatorio. Por lo cual se define a \mathcal{F}_n como la mínima σ -álgebra que hace medibles a todas las funciones n -simétricas. Esta σ -álgebra cumple con ser decreciente, i.e., $\mathcal{F}_{n+1} \subseteq \mathcal{F}_n$.

Considerando a f una función tal que $\mathbb{E}(f(X_1)) < \infty$ y a $\phi(\mathbf{x})$ una función acotada y n -simétrica y por intercambiabilidad se tiene que para $1 \leq j \leq n$

$$\begin{aligned}\mathbb{E}(f(X_j)\phi(\mathbf{X})) &= \mathbb{E}(f(X_1)\phi(X_j, X_2, \dots, X_{j-1}, X_1, X_{j+1}, \dots)) \\ &= \mathbb{E}(f(X_1)\phi(\mathbf{X})).\end{aligned}$$

De esta manera se tiene que

$$\mathbb{E}\left(\frac{1}{n}\sum_{j=1}^n f(X_j)\phi(\mathbf{X})\right) = \mathbb{E}(f(X_1)\phi(\mathbf{X})).$$

Si se toma a un evento $A \in \mathcal{F}_n$ se puede elegir a $\phi(\mathbf{X})$ como la indicadora de A , que es claramente una función acotada y n -simétrica, por lo que

$$\mathbb{E}\left(\frac{1}{n}\sum_{j=1}^n f(X_j)\phi(\mathbf{X})\right) = \mathbb{E}\left(\frac{1}{n}\sum_{j=1}^n f(X_j)\mathbb{1}_A\right) = \mathbb{E}(f(X_1)\mathbb{1}_A). \quad (\text{A.1})$$

Más aún, se tiene que $\frac{1}{n}\sum_{j=1}^n f(X_j)$ es una variable n -simétrica, por lo cual es \mathcal{F}_n medible y cumple con la ecuación (A.1), que es la definición de esperanza condicional, por lo que

$$\frac{1}{n}\sum_{j=1}^n f(X_j) = \mathbb{E}(f(X_1)|\mathcal{F}_n). \quad (\text{A.2})$$

La expresión (A.2) define una martingala reversa con respecto a $\{\mathcal{F}_n\}_{n \in \mathbb{N}}$ y más aún utilizando el teorema de convergencia para martingalas reversas se tiene que

$$\frac{1}{n} \sum_{j=1}^n f(X_j) \xrightarrow{\text{c.s.}} \mathbb{E}(f(X_1) | \mathcal{F}_\infty),$$

donde $\mathcal{F}_\infty = \bigcap_{n \in \mathbb{N}} \mathcal{F}_n$. Si $f(x) = \mathbb{1}_{(-\infty, x]}$ se tiene que

$$F_n(x) = \frac{1}{n} \sum_{j=1}^n \mathbb{1}_{\{X_j \leq x\}} \xrightarrow{\text{c.s.}} F(x) = \mathbb{P}(X_1 \leq x | \mathcal{F}_\infty).$$

Lo cual nos da una pauta de lo que se debe hacer cuando se busque una generalización para n -variables. Para esto se define a

$$A_n(f) = \frac{1}{\binom{n}{k}} \sum_{i \in I_{n,k}} f(X_{i_1}, \dots, X_{i_k}).$$

Con f una función medible de $\mathbb{R}^k \rightarrow \mathbb{R}$ y donde $I_{n,k}$ denota al conjunto de índices distintos $1 \leq i_1, \dots, i_k \leq n$. Por intercambiabilidad se tiene que $A_n(f)$ es una función n -simétrica y por ende \mathcal{F}_n medible por lo que

$$\begin{aligned} A_n(f) &= \mathbb{E}(A_n(f) | \mathcal{F}_n) \\ &= \frac{1}{\binom{n}{k}} \sum_{I_{n,k}} \mathbb{E}(f(X_{i_1}, \dots, X_{i_k}) | \mathcal{F}_n) \\ &= \mathbb{E}(f(X_1, \dots, X_k) | \mathcal{F}_n). \end{aligned}$$

Utilizando de nueva cuenta el teorema de convergencia para martingalas reversas se tiene que

$$A_n(f) \xrightarrow{\text{c.s.}} \mathbb{E}(f(X_1, \dots, X_k) | \mathcal{F}_\infty).$$

Lo que se quiere demostrar es que $\mathbb{E}(f(X_1, \dots, X_k) | \mathcal{F}_\infty) = \mathbb{E}(f(X_1, \dots, X_k))$ para lo cual se toman funciones medibles $f : \mathbb{R}^{k-1} \rightarrow \mathbb{R}$ y $g : \mathbb{R} \rightarrow \mathbb{R}$ y se define a $\phi(x_1, \dots, x_k) = f(x_1, \dots, x_{k-1})g(x_k)$ y a $\phi_j(x_1, \dots, x_k) = f(x_1, \dots, x_{k-1})g(x_j)$

De esta manera

$$\begin{aligned}
(n)_{k-1}A_n(f) \cdot nA_n(g) &= \sum_{i \in I_{n,k-1}} f(X_{i_1}, \dots, X_{i_{k-1}}) \sum_{j=1}^n g(X_j) \\
&= \sum_{i \in I_{n,k}} f(X_{i_1}, \dots, X_{i_{k-1}})g(X_{i_k}) + \sum_{i \in I_{n,k-1}} \sum_{j=1}^{k-1} f(X_{i_1}, \dots, X_{i_{k-1}})g(X_j).
\end{aligned}$$

Por lo que,

$$\begin{aligned}
A_n(\phi) &= \frac{1}{(n)_k} \sum_{i \in I_{n,k}} \phi(X_{i_1}, \dots, X_{i_k}) \\
&= \frac{1}{(n)_k} \sum_{i \in I_{n,k}} f(X_{i_1}, \dots, X_{i_{k-1}})g(X_{i_k}) \\
&= \frac{(n)_{k-1}n}{(n)_k} A_n(f)A_n(g) - \frac{(n)_{k-1}}{(n)_k} \sum_{j=1}^{k-1} A_n(\phi_j) \\
&= \frac{n}{n-k+1} A_n(f)A_n(g) - \frac{1}{n-k+1} \sum_{j=1}^{k-1} A_n(\phi_j). \tag{A.3}
\end{aligned}$$

Al hacer n tender a infinito el segundo término del lado derecho de la ecuación (A.3) se va a cero y utilizando el teorema de convergencia para martingalas reversas en $A_n(\phi)$, $A_n(f)$ y $A_n(g)$ se tiene que

$$\mathbb{E}(f(X_1, \dots, X_{k-1})g(X_k)|\mathcal{F}_\infty) = \mathbb{E}(f(X_1, \dots, X_{k-1})|\mathcal{F}_\infty)\mathbb{E}(g(X_k)|\mathcal{F}_\infty).$$

De manera inductiva se sigue que

$$\mathbb{E}\left(\prod_{j=1}^k f_j(X_j) \middle| \mathcal{F}_\infty\right) = \prod_{j=1}^k \mathbb{E}(f_j(X_j)|\mathcal{F}_\infty).$$

Por lo que la sucesión de variables aleatorias es independiente e idénticamente distribuida dada \mathcal{F}_∞ .

Finalmente, si tomamos a la función de distribución empírica,

$$Q_n = \frac{1}{n} \sum_{j=1}^n \delta_{X_j},$$

se tiene que existe una distribución condicional $Q := \mathbb{P}(X_1 \leq x | \mathcal{F}_\infty)$, tal que para eventos $A_1, \dots, A_n \in \mathcal{X}$, se tiene que $\mathbb{P}(X_i \in A_i | \mathcal{F}_\infty) = Q(A_i)$, por lo que si se elige a $P \sim Q$ se tiene que

$$\begin{aligned} \mathbb{P}(X_1 \in A_1, \dots, X_n \in A_n | P) &= \mathbb{E}(\mathbb{P}(X_1 \in A_1, \dots, X_n \in A_n | \mathcal{F}_\infty) | P) \\ &= \mathbb{E} \left(\prod_{j=1}^n P(A_j) \middle| P \right) \\ &= \prod_{j=1}^n P(A_j). \end{aligned}$$

Con lo cual se demuestra la ida del teorema de representación, para la vuelta basta notar que el producto es una función simétrica e invariante ante permutaciones, por lo cual si se tiene la representación integral dada en el teorema se cumple que las variables aleatorias son intercambiables. □

Apéndice B

En este apéndice se mencionan bajo la nomenclatura de la presente tesis los resultados descritos por James (2005) para la demostración de los teoremas de la sección 2.2.

El primero de ellos es una proposición relacionada con un cambio de medida exponencial. Para esto se debe considerar un proceso Poisson, Π definido en un espacio polaco $(\mathbb{X}, \mathcal{X})$, de intensidad ν y cuya ley está descrita por \mathbb{P}_ν . También se debe considerar el espacio $(\mathbb{M}_\mathbb{X}, \mathcal{M}_\mathbb{X})$ descrito como en la sección 2.1. y el espacio de funciones con soporte acotado denotado por $BM(\mathbb{X})$ y de manera equivalente a $BM_+(\mathbb{X})$ como el espacio de funciones no negativas de soporte acotado.

Proposición B.1. *Para toda $f \in BM_+(\mathbb{X})$ y $g \in (\mathbb{M}_\mathbb{X}, \mathcal{M}_\mathbb{X})$ se tiene*

$$\int_{\mathbb{M}_\mathbb{X}} g(\Pi) e^{-\Pi(f)} \mathbb{P}_\nu(d\Pi) = e^{-\varphi(f)} \int_{\mathbb{M}_\mathbb{X}} g(\Pi) \mathbb{P}_{e^{-f}\nu}(d\Pi).$$

Donde $\varphi(f)$ es el exponente laplaciano y

$$\Pi(f) := \int_{\mathbb{X}} f(x) \Pi(dx).$$

En otras palabras, se sigue que $e^{-\Pi(f)} \mathbb{P}_\nu(d\Pi) = e^{-\varphi(f)} \mathbb{P}_{e^{-f}\nu}(d\Pi)$. El resultado se puede extender a cualquier función f medible y no negativa tal que $\int_{\mathbb{X}} (1 - e^{-f(x)}) \nu(dx) < \infty$.

Si ahora, se toma a $\mathbb{W} = \mathbb{R}_+ \times \mathbb{X}$ como el espacio producto, a Π un proceso Poisson en este espacio con intensidad $\nu(ds, dx) = \rho_x(ds)\alpha(dx)$ y a Φ una medida aleatoria (descrita como en el Capítulo 2) en \mathbb{X} entonces,

$$\Phi(dx) = \int_{\mathbb{R}_+} s\Pi(ds, dx).$$

Se puede construir una clase de medidas de probabilidad aleatorias que sean discretas en \mathbb{X} como

$$\mathcal{P}_\Phi(dx) = q(x, \Phi)\Phi(dx) = q(x, \Phi) \int_{\mathbb{R}_+} s\Pi(ds, dx). \quad (\text{B.1})$$

Donde q es una función medible y estrictamente positiva que haga que \mathcal{P}_Φ esté bien definida. La representación en B.1 sirve para construir una nueva clase de medidas de probabilidad aleatorias en el espacio producto $\mathbb{R}_+ \times \mathbb{X}$, definidas como

$$\tilde{\mathcal{P}}_\Phi(ds, dx) = q(x, \Phi)s\Pi(ds, dx). \quad (\text{B.2})$$

Recordando que dada una muestra $X_1, \dots, X_n | \mathcal{P}_\Phi \stackrel{iid}{\sim} \mathcal{P}_\Phi$, ésta admite una representación en términos de $Y_1, \dots, Y_{n(\pi)}$, que son las distintas observaciones y π una partición de $\{1, \dots, n\}$, con correspondientes saltos J_j . Se denota a $W_i = (J_i, X_i)$, para $i = 1, \dots, n$, como los puntos en el espacio producto y a $W_j^* = (J_{j,n}, Y_j)$, con $j = 1, \dots, n(\pi)$, como los valores únicos de esta muestra. Se puede definir la siguiente medida aleatoria,

$$\tilde{\Phi}_n(dx) = \Phi(dx) + \sum_{j=1}^{n(\pi)} J_{j,n} \delta_{Y_j}(dx).$$

Y

$$\psi_n(\mathbf{J}, \mathbf{Y}) = \left(\prod_{j=1}^{n(\pi)} J_{j,n}^{n_j} \right) \phi_n(\mathbf{J}, \mathbf{Y}).$$

Donde

$$\phi_n(\mathbf{J}, \mathbf{Y}) = \int_{\mathbb{M}_\mathbb{X}} \left(\prod_{j=1}^{n(\pi)} (q(Y_j, \tilde{\Phi}_n)^{n_j}) \right) \mathbb{P}_\nu(d\Pi)$$

Y se llega al siguiente teorema.

Teorema B.2. *Sea \mathcal{P}_Φ como en B.2, con Π un proceso Poisson en $\mathbb{R}_+ \times \mathbb{X}$ y de intensidad ν . Sea $\mathbf{X} = (X_1, \dots, X_n)$ tal que $X_1, \dots, X_n | \mathcal{P}_\Phi \stackrel{iid}{\sim} \mathcal{P}_\Phi$, entonces se siguen los siguientes resultados.*

1. *La distribución posterior de $\Pi | \mathbf{X}$ corresponde a la ley condicional de la medida aleatoria $\tilde{\Pi}_n = \Pi + \sum_{j=1}^{n(\pi)} \delta_{J_j, Y_j}$, donde la ley condicional de Π dado \mathbf{J}, \mathbf{X} está dada por*

$$f(d\Pi | \mathbf{J}, \mathbf{X}) = (\phi_n(\mathbf{J}, \mathbf{X}))^{-1} \left(\prod_{j=1}^{n(\pi)} (q(Y_j, \tilde{\Phi}_n)^{n_j}) \right) \mathbb{P}_\nu(d\Pi).$$

Más aún, la distribución condicional de $\mathbf{J} | \mathbf{X}$ es proporcional a $\phi_n(\mathbf{J}, \mathbf{X}) \prod_{j=1}^{n(\pi)} J_{j,n}^{n_j} \rho_{Y_j}(dJ_{j,n})$. La ley de $\tilde{\Phi}_n(dx) = \int_{\mathbb{R}_+} s \tilde{\Pi}_n(ds, dx)$, dada \mathbf{X} , está determinada por la ley de $\tilde{\Pi}_n | \mathbf{X}$, y corresponde a la distribución posterior de $\Phi | \mathbf{X}$.

2. *La distribución posterior de $\mathcal{P}_\Phi | \mathbf{X}$ es equivalente a la distribución condicional, dado \mathbf{X} , de la medida de probabilidad aleatoria $\mathcal{P}_{\tilde{\Phi}_n}(dx) = q(y, \tilde{\Phi}_n) \tilde{\Phi}_n(dx)$.*
3. *Finalmente, la distribución (intercambiable) marginal de \mathbf{X} está dada por*

$$\mathbb{P}_\nu(d\mathbf{X}) = \int_{\mathbb{R}_+^{n(\pi)}} \phi_n(\mathbf{s}, \mathbf{X}) \prod_{j=1}^{n(\pi)} s_{j,n}^{n_j} \rho_{y_j}(ds_{j,n}) \alpha(dy_j).$$

Y la EPPF se puede expresar como

$$f(n_1, \dots, n_{n(\pi)}) = \int_{\mathbb{R}_+^{n(\pi)} \times \mathbb{X}^{n(\pi)}} \phi_n(\mathbf{s}, \mathbf{x}) \prod_{j=1}^{n(\pi)} s_{j,n}^{n_j} \rho_{y_j}(ds_{j,n}) \alpha(dy_j). \quad (\text{B.3})$$

Bibliografía

- Argiento, R., I. Bianchini y A. Guglielmi (2016). *Posterior sampling from ϵ -approximation of normalized completely random measure mixtures*. Electronic Journal of Statistics. Vol. 10, 3516-3547.
- Brix, A. (1999). *Generalized gamma measures and shot-noise Cox processes*. Advances in Applied Probability, 31, 929-953
- Durrett, R. (2010). *Probability: Theory and Examples*. Cambridge University Press. 4ta edición.
- Ferguson, T. (1973). *A Bayesian Analysis of Some Nonparametric Problems*. The Annals of Statistics, Vol.1, No.2, 209-230.
- Ghosal, S. y A. van der Vaart. (2017). *Fundamentals of Nonparametric Bayesian Inference*. Cambridge University Press.
- Gil Leyva, María F. (2016). *Random Partitions Models* (Tesis de maestría). México. Universidad Nacional Autónoma de México.
- González-Barrios, José M. (2011). *Lecture Notes on Probability Theory*. México. Serie Monografías. Volumen 14, No.31.
- Hjort, N. L., C. Holmes, P. Müller y S. G. Walker (2010). *Bayesian Nonparametrics*. Nueva York, Estados Unidos: Cambridge University Press.

- James, Lancelot F. (2005). *Bayesian Poisson Process Partition Calculus with an Application to Bayesian Lévy Moving Averages*. The Hong Kong University of Science and Technology. The Annals of Statistics 2005, Vol. 33, No. 4, 1771-1799.
- Kingman, J.F.C. (1967). *Completely Random Measures*. Pacific Journal of Mathematics. Vol. 21, No.1, 59-78.
- Kingman, J.F.C. (1975). *Random discrete distributions (with discussion)*. Journal of the Royal Statistical Society, Series B,37, 1-22.
- Kingman, J.F.C. (1978). *Uses of Exchangeability*. The Annals of Probability, Vol. 6, No. 2, 183-197.
- Kingman, J.F.C. (1993). *Poisson Processes*. Nueva York, Estados Unidos: Oxford University Press.
- Klenke, A. (2014). *Probability Theory*. Alemania. Springer. 2da edición.
- Kyprianou, Andreas E. (2006). *Introductory Lectures on Fluctuations of Lévy Processes with Applications*. Alemania. Springer.
- Lijoi A., e I. Prünster. (2009) *Distributional properties of means of random probability measures*. Statistics Surveys. Vol 3 47-95
- Lijoi A., L. James e I. Prünster. (2009) *Posterior Analysis for Normalized Random Measures with Independent Increments*. Scandinavian Journal of Statistics. Vol 36: 76-97.
- Pitman, J. y M. Yor. (1997). *The Two-Parameter Poisson-Dirichlet Distribution Derived from a Stable Subordinator*. The Annals of Probability. Vol. 25, No. 2, 855-900.
- Pitman, J. (2006) *Combinatorial Stochastic Processes*. Ecole d'Été de Probabilités de Saint-Flour XXXII-2002. Springer
- Sato Ken-Iti. (1990). *Lévy Processes and Infinitely Divisible Distributions*. Cambridge, UK: Cambridge University Press.

Steutel Fred W. y Klaas V. Harin. (2004) *Infinite Divisibility of Probability Distributions on the Real Line*. Nueva York, Estados Unidos. Marcel Dekker, Inc.

Uribe B., Gerónimo. (2013) *Procesos Estocásticos I. Semestre 2013-II*. México. Instituto de Matemáticas, Universidad Nacional Autónoma de México.