Notas para tesis

I

27 de febrero de 2017

1. Prior y marginal

En estadística bayesiana se tiene que $x \sim p(x \mid \theta)$ en donde $\theta \sim \Pi(\theta)$. Si tenemos las observaciones $x_1, \ldots, x_n \overset{iid}{\sim} p(x \mid \theta)$, entonces la distribución *a posteriori* está dada por:

$$\Pi(\theta \mid x_1, \ldots, x_n) = \frac{p(x_1, \ldots, x_n \mid \theta)\, \Pi(\theta)}{p(x_1, \ldots, x_n)} \propto \Pi(\theta) \prod_{i=1}^{n} p(x_i \mid \theta). \tag{1}$$

La distribución marginal de x sin depender de θ está dada por

$$p(x) = \int_{\Theta} p(x \mid \theta)\, \Pi(\theta) d\theta, \tag{2}$$

por lo que dadas las distribuciones anteriores se tiene que la distribución marginal posterior está dada por

$$p(x \mid x_1, \ldots, x_n) = \int_{\Theta} p(x \mid \theta)\, \Pi(\theta \mid x_1, \ldots, x_n)\, d\theta. \tag{3}$$

2. Categórica-Dirichlet

En particular, si $X \in \{1, 2, \ldots, k\}$ donde $k < \infty$ tal que $P(X = j) = p_j \geq 0$ para $j = 1, \ldots, k$ donde $\sum_{i=1}^{k} p_i = 1$. Si suponemos que $p_1, \ldots, p_k$ son aleatorias entonces $\Pi(p_1, \ldots, p_k)$ tiene soporte sobre el simplex de dimensión $k-1$[1] puesto que uno de los valores está autodeterminado. La distribución que usualmente se usa en este caso es la distribución Dirichlet[2]:

$$\Pi\left(\underline{p} \mid \underline{\alpha}\right) = \frac{\Gamma\left(\sum_{i=1}^{k} \alpha_i\right)}{\prod_{i=1}^{k} \Gamma(\alpha_i)} \prod_{i=1}^{k} p_i^{\alpha_i - 1} \mathbb{1}_{S_{k-1}}\left(\underline{p}\right). \tag{4}$$

Con esto podemos se puede decir que $X \sim Cat\left(k, \underline{p}\right)$ donde

$$p\left(x \mid \underline{p}\right) = \prod_{i=1}^{k} p_j^{[x=j]} \mathbb{1}_{1,\ldots,k}(x), \tag{5}$$

[1] El simplex de dimensión m está definido como $S_m = \{y \in \mathbb{R}^m \mid y \geq 0,\, y^T e \leq 1\}$ con $e = (1, \ldots, 1)^T \in \mathbb{R}^m$.

[2] Para ahorrar espacio y facilitar la escritura usaré notación vectorial: $\underline{p} = (p_1, \ldots, p_k)$ y $\underline{\alpha} = (\alpha_1, \ldots, \alpha_k)$.

donde

$$[x=j]=\begin{cases} 0 & \text{si} \quad x\neq j \\ 1 & \text{si} \quad x=j \end{cases}$$

son los corchetes de Iverson.

Si se tienen observaciones $x_1,\ldots,x_n \overset{iid}{\sim} Cat\left(k,\underline{p}\right)$ donde $\underline{p}\sim Dirichlet\left(\underline{\alpha}\right)$ entonces la actualización, por (**??**), (**??**) y (**??**), está dada por:

$$\begin{aligned}\Pi\left(\theta\,|\,x_1,\ldots,x_n\right) &\propto \prod_{i=1}^{k} p_i^{\alpha_i-1}\prod_{j=1}^{n}\prod_{l=1}^{k} p_l^{[x_j=l]}\mathbb{1}_{S_{k-1}}\left(\underline{p}\right) \\ &= \prod_{i=1}^{k} p_i^{\alpha_i+\sum_{j=1}^{n}[x_j=i]-1}\mathbb{1}_{S_{k-1}}\left(\underline{p}\right),\end{aligned}$$

por lo que $\theta\,|\,x_1,\ldots,x_n \sim Dirichlet\left(\underline{\alpha}+\underline{n'}\right)$ en donde $\underline{n'}=\left(n'_1,\ldots,n'_k\right)$, con $n'_i=\sum_{j=1}^{n}\left[x_j=i\right]$, es el vector en donde la entrada i es el número de ocurrencias de ese caso en la muestra.

Como la distribución de la posterior sigue siendo Dirichlet, al igual que la prior, este modelo[3] se dice conjugado y de esto anterior es evidente que se puede generalizar al modelo *Multinomial-Dirichlet*. Al caso especial de este modelo, cuando $k=2$, se le conoce como *Bernoulli-Beta*.

3. Phadia

3.1. Revisión general

3.1.1. Pequeña introducción

Aquí se enlistan de forma general algunos procesos desarrollados para ser usados como distribuciones a priori en problemas no paramétricos desde el punto de vista bayesiano. Más adelante se detallarán algunos de estos y se dirán sus propiedades. En el enfoque bayesiano la función de distribución desconocida, de donde se obtiene una muestra se considera como un parámetro. Por ello se deben construir distribuciones a priori en el espacio de todas las funciones de distribución, $\mathcal{F}\left(\chi\right)$, o en el espacio de medidas de probabilidad, Π, definido sobre $(\mathfrak{X},\mathcal{A})$.

Mientras que la función de distribución es el parámetro de mayor interés en análisis bayesiano no paramétrico a veces es más útil discutir el proceso *a priori* en términos de una medida de probabilidad, P.[4] Definir cosas *a priori* es teóricamente complicado, por lo que se buscan funciones que cumplan ciertos requisitos para mejorar el asunto: el soporte de la función debe ser suficientemente amplio y la distribución posterior debe ser tratable analíticamente.

3.1.2. Métodos de construcción

El tener un prior en $\mathcal{F}$ o Π se puede clasificar en cuatro casos:

1. El primer caso es bajo la distribución conjunta de probabilidades aleatorias.

[3] Categórica-Dirichlet

[4] El proceso de Dirichlet está definido de esa forma.

2. El segundo y tercer caso es bajo conceptos de independencia.

3. El último caso se basa en un esquema de urnas de Polya.

Los primeros tres casos se sustentan primordialmente en propiedades de la distribución Dirichlet, aunque nuevos métodos han surgido tomando como base la construcción de *romper una rama*.

El primer método se debe a Ferguson y se desarrolla en términos de la distribución conjunta de las probabilidades de una partición medible en un conjunto arbitrario. El segundo método se fundamenta en la independencia de los incrementos sucesivos normalizados de una función de distribución F definida en $\mathbb{R}$. El tercer método se basa en la propiedad *tailfree* de la distribución Dirichlet.

Definición 1 (Tailfree). *Una medida de probabilidad aleatoria P se dice que es tailfree si dada $\{\pi_n\}$, una sucesión de particiones anidadas de $\mathbb{R}$, donde π_{n+1} es un refinamiento de π_n, se cumple que $\{P(B|A) \mid A \in \pi_n \text{ y } B \in \pi_{n+1}\}$ para $n = 1, 2, \ldots$ son independientes.*

El cuarto método de construcción se basa en construir una sucesión de variables aleatorias intercambiables por medio de un esquema de urnas de Polya y luego aplicando un teorema de de Finetti.

Definición 2 (Proceso de Polya). *Sea $\chi = \{1, 2, \ldots, k\}$, dada una urna con α_i bolas de un color ($i = 1, 2, \ldots, k$). Se saca una bola, creando una variable aleatoria X_1 en donde $P(X_1 = i) = \frac{\alpha_i}{\sum_{j=1}^{k} \alpha_j}$ y se remplaza con dos bolas del color que se sacó. En el siguiente paso se saca otra bola y ahora se tiene que $P(X_2 = j \mid X_1 = i) = \frac{\alpha_j + \delta_{ij}}{\sum \alpha_l + 1}$. Se repite este proceso indefinidamente para obtener una sucesión de variables aleatorias intercambiables tomando valores en χ.*

La distribución muestral de $X_1, X_2, \ldots$ converge casi seguramente a un vector $\theta = (\theta_1, \ldots, \theta_k)$ con distribución Dirichlet con parámetros $(\alpha_1, \ldots, \alpha_k)$, y dado θ, las X_i son independientes con $P(X_i = j) = \theta_j$. Un teorema de de Finetti[5] asegura que hay una medida μ tal que la distribución de probabilidad conjunta marginal bajo esta medida es la misma para cualquier permutación de las variables. Blackwell y MacQueen generalizaron lo anterior tomando un número continuo de colores α. Mostraron que la sucesión de $\frac{\alpha_n(\cdot)}{\alpha_n(\mathfrak{X})}$ (donde $\alpha_n(\cdot) = \alpha(\cdot) + \sum_{i=1}^{n} \delta_i(\cdot)$) converge a una medida discreta P, esta es el proceso de Dirichlet con parámetro α.

La representación de una medida de probabilidad como una mezcla contable ha sido útil para el desarrollo de nuevas aplicaciones. Ferguson propuso una definición alternativa del proceso de Dirichlet. $\sum_{i=1}^{\infty} p_i \delta_{\xi_i}$ en donde los puntos de acumulación de masa son aleatorios y provienen de la distribución $F_0 = \frac{\alpha(\cdot)}{\alpha(\mathfrak{X})}$, y los pesos p_i son aleatorios con las restricciones $0 \leq p_i \leq 1$ y $\sum_{i=1}^{\infty} p_i = 1$. El problema con esta representación es que Ferguson contruyó los pesos con la distribución gamma, lo cual lo hace poco práctico; Sethuraman ataca este problema generando los pesos por medio de variables aleatorias beta, esto renovó el interés en esta representación que da lugar a nuevos procesos llamados *Ferguson-Sethuraman*.

[5] Investigar

Por esto el proceso Dirichlet sirve como base prior y se usa como fuente para ser generalizada. Antoniak propusó que α también fuera un parámetro, indexado por $u \sim H$, por lo que se tiene una mezcla de procesos de Dirichlet, $P \in \int \mathcal{D}\left(\alpha_u\right) dH(u)$. Dalal propusó que la medida α fuera invariante dentro de un grupo de transformaciones creando el *Proceso de Dirichlet Invariante*. Lo, escribiendo $f(x) = \int K(x,u)dG(u)$, con K un kernel conocido y $G \in \mathcal{D}(\alpha)$, pudo poner priors en un espacio de funciones de densidad. Tomando $\alpha\left(\mathfrak{X}\right)$ como una función positiva en lugar de constante, Walker y Muliere generalizaron el proceso de Dirichlet para que el soporte incluya funciones absolutamente continuas (este proceso se llama *beta-Stacy*).

3.1.3. Procesos a priori

El proceso de Dirichlet de Ferguson cumple con los dos requisitos básicos para un proceso a priori: es simple, está definido en un espacio de probabilidad arbitrario y pertenece a una familia de prioris conjugada. Lijoi y Prünster indentifican la conjugacidad como estructural o paramétrica; en la primera la distribución posterior tiene la misma estructura que la prior mientras que en la segunda la distribución posterior es igual a la prior pero con cambios en los parámetros. El proceso de Dirichlet tiene un parámetro interpretable; dada una muestra aleatoria $\mathbf{X} \sim P \in \mathcal{D}\left(\alpha\right)$, Ferguson demostró que la distribución posterior, dada la muestra es también un proceso de Dirichlet, en particular $P \,|\, \mathbf{X} \in \mathcal{D}\left(\alpha + \sum_{i=1}^{n} \delta_{x_i}\right)$. Esta propieda hace posible la derivación de estimadores bayesianos no paramétricos de varias funciones de P actualizando α En realidad α puede ser visto como una representación de dos parámetros: $F_0\left(\cdot\right) = \overline{\alpha}\left(\cdot\right) = \frac{\alpha(\cdot)}{\alpha(\mathfrak{X})}$ y $M = \alpha\left(\mathfrak{X}\right)$. F_0 se interpreta como una adivinanza a priori (o distribución base) de F y M es el tamaño poblacionas a prior o el parámetro de presición (o nuestra creencia de que F_0 es correcta). La media posterior de F es una combinación convexa de F_0 y la función de distribución empírica.

El proceso de Dirichlet es el único proceso a priori en el que la distribución de $P\left(A\right)$ depende únicamente del número de observaciones dentro de A y no de su ubicación y esto es considerado como una debilidad; pero el mayor problema es que el soporte solamente contiene medidad de probabilidad discretas.[6] El proceso de Dirichlet, a pesar de ser popular y satisfacer muchas demandas, no era adecuado para estimar densidades y algunos otros problemas. Para esto se crearon algunas extensiones.

Respecto a la estimación basada en datos censurados por la derecha, asumiendo el proceso Dirichlet como prior, se sigue que la distribución posterior es ua mezcla de procesos de Dirichlet. Esto llevó al desarrollo de mezclas de procesos de Dirichlet. (Antoniak) Este proceso también tiene la propiedad de conjugacidad; sea $\theta \sim P \in \int_U \mathcal{D}\left(\alpha_u\right) dH(u)$ una muestra de tamaño n, entonces $P \,|\, \theta \in \int_U \mathcal{D}\left(\alpha_u + \sum_{i=1}^{n} \delta_{\theta_i}\right) dH_\theta(u)$. Estos procesos son útiles en problemas de evaliación biológica, pero obtener la expresión explícita de la distribución posterior es díficil por lo que se confía en procedimientos computacionales.

El proceso de Dirichlet es no parámetrico en el sentido de que tiene un soporte amplir; por esto Dalal vio la necesidad de que la prior debía de tener

[6] Algunas aplicaciones recientes demuestran que este problema no es tan grave como podría parecer y que resulta benéfico.

una estructura inherente, como simetría, o alguna propiedad de invarianza. Esto lo llevó a definir un proceso invariante respecto a un grupo de transformaciones medibles que selecciona una función de distribución invariante con probabilidad uno. Eso lo llamó *proceso de Dirichlet invariante*, $\mathcal{DGI}(\alpha)$,con parámetro α. La propiedad de conjugacidad también se cumple para este tipo de procesos.

Teh *et al* proponen modelos jerárquicos donde los parámetros de las distribuciones a priori son asignados a priori con hiperparámetros. Un modelo general que contiene el proceso de Dirichlet, fue propuesto por Pitman, es el de *modelos de sampleo de especies* y la probabilidad está dada por

$$P(\cdot) = \sum_{j=1}^{\infty} p_j \delta_{\xi_j}(\cdot) + \left(1 - \sum_{j=1}^{\infty} p_j\right) Q(\cdot),$$

donde Q es una medida de probabilidad correspondiente a la distribución continua G, $\xi_i \sim G$ y $\sum_{j=1}^{\infty} p_j \leq 1$.

3.2. Proceso de Dirichlet

Este proceso es el más popular y usado como medida a priori en análisis no paramétrico bayesiano, este suceso se debe a su tratabilidad matemática; junto a sus generalizaciones, son de las priores más usadas e importantes en la modelación de datos covariados de grandes dimensiones. Un proceso de Dirichlet a priori con parámetro α para una función de distribución F es una medida de probabilidad en el espacio de funciones de distribución y tiene dos parámetros importantes: una distribución base F_0 y la precisión M. Ferguson define el proceso de Dirichlet en términos de una medida de probabilidad aleatoria; desafortunadamente la concentración del proceso es en medidad de probabilidad discretas.

3.2.1. Definición

Sea P una medida de probabilidad en $(\mathfrak{X}, \mathcal{A})$ en donde $\mathfrak{X}$ es un espacio métrico separable y $\mathcal{A} = \sigma(\mathfrak{X})$, además sea Π el conunto de todas las medidas de probabilidad en $(\mathfrak{X}, \mathcal{A})$. ($P$ es el parámetro y $(\Pi, \sigma(Pi))$ es el esspacio parametral.) Se denota como F a la función de distribución correspondiente a P y $\mathcal{F}$ el espacio de todas las funciones de distribución.

Sea $D(\gamma_1, \ldots, \gamma_k)$ la distribución Dirichlet $k-1$-dimensional con densidad

$$f(x_1, \ldots, x_{k-1}) = \frac{\Gamma(\gamma_1 + \cdots + \gamma_k)}{\Gamma(\gamma_1) \cdots \Gamma(\gamma_k)} \prod_{i=1}^{k-1} x_i^{\gamma_i - 1} \left(1 - \sum_{i=1}^{k-1} x_i\right)^{\gamma_k - 1},$$

sobre $S_{k-1} = \left\{\mathbf{x} \in \mathbb{R}^{k-1} \,|\, \mathbf{x} \geq \mathbf{0} \text{ y } \|\mathbf{x}\|_1 \leq 1\right\}$ y $\gamma_i \in \mathbb{R}_{>0}$.

Decimos que P es una medida de probabilidad aleatoria en $(\mathfrak{X}, \mathcal{A})$ si para cualquier $A \in \mathcal{A}$, $P(()A)$ es aleatoria con valores en $[0,1]$, $P(\mathfrak{X}) = 1$ casi seguramente y P es finitamente aditiva.

Definición 3 (Proceso de Dirichlet)**.** *(Ferguson) Sea α una medida finita no negativa en $(\mathfrak{X}, \mathcal{A})$. Una probabilidad aleatoria P es un proceso de Dirichlet en $(\mathfrak{X}, \mathcal{A})$ con parámetro α si para cada $k > 0$ entero y una partición medible $(A_1, \ldots, A_k)$ de $\mathfrak{X}$, el vector $(P(A_1), \ldots, P(A_k))$ tiene la distribución $D(\alpha(A_1), \ldots, \alpha(A_k))$.*

Una definición alternativa está dada por una mezcla contable en la que los pesos se derivan de un proceso gamma y masas puntuales en puntos aleatorios. Definió el proceso Dirichlet como un proceso gamma con los incrementos divididos por su suma. Ishwaran y James describen una formulación alternativa para los pesos, en donde usan exponenciales con parámetro 1. Otro proceso de construcción se da por Sethuraman y Tiwari en el que los pesos se derivan usando la distribución beta con parámetros 1 y $\alpha(\mathfrak{X})$; su representación de una medida de probabilidad P con prior Dirichlet $\mathcal{D}(\alpha)$ es

$$P(A) = \sum_{j=1}^{\infty} p_j \delta_{\xi_j}(A), \qquad A \in \mathcal{A},$$

con $\xi_j \overset{iid}{\sim} \overline{\alpha}(\cdot)$ tomando valores en $\mathfrak{X}$, $p_1 = V_1$ y $p_j = V_j \prod_{i=1}^{j-1}(1 - V_i)$ con $V_j \overset{iid}{\sim} B(1, \alpha(\mathfrak{X}))$. Esta construcción se llama construcción de romper una rama. Esta representación demuestra que el proceso de Dirichlet escoge ua distribución discreta con probabilidad 1.

Blackwell y MacQueen también proponen una definición alternativa. Sea $\{X_n \mid n \geq 1\}$ una sucesión de variables aleatorias tomando valores en $\mathfrak{X}$ como sigue: para cada $A \in \mathcal{A}$ sea $P(X_1 \in A) = \frac{\alpha(A)}{\alpha(\mathfrak{X})}$ y

$$P(X_{n+1} \in A \mid X_1, \ldots, X_n) = \frac{\alpha_n(A)}{\alpha_n(\mathfrak{X})} = \frac{\alpha(A) + \sum_{i=1}^{n} \delta_{x_i}(A)}{\alpha(\mathfrak{X}) + n}. \qquad (6)$$

A esta sucesión se le llama *sucesión de Polya con parámetro* α; Blackwell y MacQueen probaron que la sucesión $\frac{\alpha_n(\cdot)}{\alpha_n(\mathfrak{X})}$ converge a una medida discreta P y P es el proceso de Dirichlet con parámetro α. Podemos notar que (**??**) se puede expresar como

$$P(X_{n+1} \in A \mid X_1, \ldots, X_n) = \sum_{i=1}^{n} \frac{1}{\alpha(\mathfrak{X}) + n} \delta_{x_i}(A) + \frac{\alpha(\mathfrak{X})}{\alpha(\mathfrak{X}) + n} \overline{\alpha}(A).$$

Definición 4. *(Ferguson) Sea P una medida de probabilidad aleatoria en $(\mathfrak{X}, \mathcal{A})$. $X_1, \ldots, X_n$ es una muestra de P si para cada $m > 0$ entero y conjuntos medibles $A_1, \ldots, A_m, C_1, \ldots, C_n$ de $\mathfrak{X}$,*

$$P(X_1 \in C_1, \ldots X_n \in C_n | P(A_1), \ldots, P(A_n), P(C_1), \ldots, P(C_n)) = \prod_{j=1}^{n} P(C_j),$$

casi seguramente.

3.2.2. Propiedades

Para estas propiedades se asume que $P \in \mathcal{D}(\alpha)$ y dada P $X_1, \ldots, X_n$ es una muestra de P. Además se define $M = \alpha(\mathfrak{X})$.

1. El proceso de Dirichlet escoge una medida de probabilidad discreta con probabilidad 1; esto es cierto aunque α sea continua.

2. Por la discreción del proceso de Dirichlet, la distribución predictiva de una observación futura es dada por

$$X_{n+1} \mid X_1, \ldots, X_n \sim \frac{M}{M+n} \overline{\alpha} + \frac{n}{M+n} \frac{1}{n} \sum_{j=1}^{K} n_j \delta_{X_j^*}. \qquad (7)$$

3. El siguiente teorema establece la conjugacidad del proceso de Dirichlet con respecto a observaciones no censuradas:

   Teorema 1 (Ferguson). *Sea $P \in \mathcal{D}(\alpha)$ y dada P sea $X \sim P$ una muestra aleatoria de tamaño uno, entonces la distribución marginal de X es $\overline{\alpha}$; además la distribución posterior de P dada X es $\mathcal{D}(\alpha + \delta_x)$. Si $X_1, \ldots, X_n$ es una muestra aleatoria de P entonces la distribución posterior de P dada la muestra es $\mathcal{D}(\alpha + \sum_{i=1}^{n} \delta_{x_i})$.*

4. Para tomar una muestra de un proceso Dirichlet se puede realizar el procedimiento de romper una rama o la extensión de la urna de Polya. La diferencia en estos métodos es la exactitud del primero contra la aproximación del segundo. Su deficiencia en generar P es que se necesitan infinitas repeticiones y eso no es posible, por lo que se toma un criterio de paro en el tamaño. Otro método de aproximación es tomar los pesos de una distribución Dirichlet simétrica con parámetro $\frac{\alpha(\mathfrak{X})}{N}$ y los puntos de masa se obtienen de $\overline{\alpha}$.

3.2.3. Mezclas y generalizaciones del proceso de Dirichlet

Para resolver problemas de estimación bayesianos Antoniak definió mezclas de procesos Dirichlet indexando el parámetro α por $\theta \sim H(\theta)$, por lo que $P \,|\, \theta \sim \mathcal{D}(\alpha_\theta)$. Por otra parte, Lo se dio cuenta que al tratar con funciones de densidad el proceso de Dirichlet no era adecuado, por lo que propusó otro tipo de mezclas; modeló una función de densidad aleatoria en $\mathbb{R}$ como $f(x) = \int K(x,s)dG(s)$ en donde $K(x,s)$ es un kernel conocido en $\mathbb{R} \times \mathfrak{X}$ y G es un proceso Dirichlet.

Ferguson consideró una mezcla contable de densidades normales formulando la función de densidad como $f(x) = \sum_{i=1}^{\infty} p_i \mathcal{N}(x \,|\, \mu_i, \sigma_i)$. Esto puede ser escrito como $f(x) = \int \mathcal{N}(x \,|\, \mu, \sigma)\, dG(\mu, \sigma)$. Para esto toma la representación de Sethuraman, definiendo los pesos p_j con parámtro M y (μ_j, σ_j) son *iid* con respecto al prior gamma-normal; esto prueba que G es un proceso Dirchlet con parámetro $\alpha = MG_0$, donde G_0 es el prior para (μ, σ).

Mezclas gaussianas también surgen con Escobar; sea $Y_i \,|\, \mu_i \sim \mathcal{N}(\mu_i, 1)$, $\mu_i \,|\, G \overset{iid}{\sim} G$ con μ_i y G desconocidas. El objetivo es estimar las medias por medio de las observaciones. Junto con West, también uso un modelo de mezclas gaussianas para estimación de densidades; dado (μ_i, σ_i^2) se tienen observaciones independientes $Y_1, \ldots, Y_n$ tales que $Y_i \,|\, (\mu_i, \sigma_i^2) \sim \mathcal{N}(\mu_i, \sigma_i^2)$ y $\nu_i = (\mu_i, \sigma_i^2)$ son muestreados de una distribución a priori G sobre $\mathbb{R} \times \mathbb{R}_{>0}$. Los autores asumen que $G \sim \mathcal{D}(MG_0)$ donde G_0 es la distribución a priori sobre $\mathbb{R} \times \mathbb{R}_{>0}$. Por ser un proceso Dirichlet $\nu_{n+1} \,|\, \nu_1, \ldots, \nu_n$ seguirá una distribución de la forma (**??**). De lo anterior proceden a derivar la distribución de $Y_{n+1} \,|\, \nu_1, \ldots, \nu_n$, que resulta ser una mezcla de n normales y una t de Student, con esto prubas que $Y_{n+1} \,|\, Y_1, \ldots, Y_n$ tiene la distribución predictiva $\int P(Y_{n+1} \,|\, \nu)\, dP(\nu \,|\, Y_1, \ldots, Y_n)$.

Un tercer tipo de mezcla conlleva los modelos jerárquicos donde los parámetros de la distribución a priori tienen priores asignadas con hiperparámetros.

Por otro lado el proceso Dirichlet se prestó para muchas generalizaciones o se vio que era un caso particular de otros procesos. Como ejemplo de esto segundo, el proceso de Dirichlet es claramente un caso particular de el proceso Dirichlet invariante y de mezclas de procesos Dirichlet; si se define en $\mathbb{R}$ el proceso es neutral a la derecha, una transformación del proceso beta da como resultado el proceso Dirichlet que también resulta ser un caso particular de proceso beta-Stacy. Varios de los procesos relacionados al proceso Dirichlet fueron resultado de la representación de Sethuraman. Si la suma contable se trunca a $N < \infty$ términos, N fijo o aleatorio, se genera una clase de distribuciones piores discretas; si los pesos definidos por $B(1, \alpha(\mathfrak{X}))$ se obtienen por una distribución beta con dos parámetros otro grupo de priores emergen. Otro grupo surge indexando las masas con covariables; el cuarto grupo es resultado de otro tipo de extensión, si δ se cambia por una medida de probabilidad G no degenerada se tiene el proceso Dirichlet de kernels.

3.2.4. Proceso Dirichlet Invariante

Sea $\mathcal{G} = \{g_1, \ldots, g_k\}$ un grupo de transformaciones medibles en un espacio p-dimensional euclidiano. U conjunto $B \in \mathfrak{X}$ se dice $\mathcal{G}$-invariante si $B = gB$ para toda $g \in \mathcal{G}$, una medida finita no negativa γ se dice $\mathcal{G}$-invariante si $\gamma(A) = \gamma(gA)$ para toda $g \in \mathcal{G}$ y todo $A \in \mathfrak{X}$. Una partición medible de $\mathfrak{X}$ se dice $\mathcal{G}$-invariante si $A_j = gA_j$ para toda $g \in \mathcal{G}$ y $j = 1, \ldots, m$.

Definición 5 (Dalal)**.** *Una medida de probabilidad aleatoria $\mathcal{G}$-invariante es un proceso Dirichlet invariante si existe una medida, α, $\mathcal{G}$-invariante en $(\mathfrak{X}, \sigma(\mathfrak{X}))$ tal que para cada partición medible $\mathcal{G}$-invariante de $\mathfrak{X}$ la distribución conjunta de $(P(A_1), \ldots, P(A_m))$ es $D(\alpha(A_1), \ldots, \alpha(A_m))$. Se denota $P \in \mathcal{DGI}(\alpha)$.*

Tiwari extendió la represetación de Sethuraman al proceso Dirichlet invariante. Sea α una medida $\mathcal{G}$-invariante en $(\mathfrak{X}, \sigma(\mathfrak{X}))$, sean $(p_1, p_2, \ldots)$ y $(\xi_1, \xi_2, \ldots)$ dos sucesiones independientes de variables aleatorias *iid*, con las condiciones dadas en la representación contable del proceso Dirichlet, entonces la medida de probabilidad aleatoria P dada por

$$P(A) = \sum_{j=1}^{\infty} p_j \frac{1}{k} \sum_{i=1}^{k} \delta_{g_i, \xi_j}(A), \qquad A \in \sigma(\mathfrak{X}),$$

es un proceso Dirichlet invariante con parámetro α.

4. Muestreo

Suponiendo que $P \in \mathcal{D}(\alpha)$ tiene representación de Sethuraman con *stick-breaking* entonces se puede introducir una variable latente u para controlar el número de elementos que componen la distribución. Sea $u \sim U(0,1)$, entonces

$$P(\cdot \mid u) = \sum_{j=1}^{\infty} w_j \delta_{\theta_j}(\cdot)\, \mathbb{1}(u < w_j)$$

tiene un número finito de componentes si se asume que $\{\mathrm{E}[w_i]\}_{i=1}^{\infty}$ es una sucesión decreciente.

Demostración. Por hipótesis se tiene que los pesos aleatorios cumplen $w_j \geq 0$ para toda j y $\sum_{j=1}^{\infty} w_j = 1$. Ahora, como $\{\mathrm{E}[w_i]\}_{i=1}^{\infty}$ es decreciente se puede tomar la sucesión como los pesos. Ahora basta probar que $\{w_j\}_{j=1}^{\infty}$ es una sucesión que converge a 0, pues si $u \sim U(0,1)$ solamente un número finito de pesos será mayor que u.

Sea $\varepsilon > 0$, como $\sum_{j=1}^{\infty} w_j = 1$ se sigue que existe $N = N_\varepsilon \in \mathbb{N}$ tal que para toda $n \geq N$ entera se cumple que

$$\left| \sum_{i=1}^{n} w_j - 1 \right| < \frac{\varepsilon}{2}.$$

Luego, si $n > N$ se obtiene lo siguiente:

$$\begin{aligned} |w_n| &= \left| \sum_{j=1}^{n} w_j - \sum_{i=1}^{n-1} w_i \right| \\ &\leq \left| \sum_{j=1}^{n} w_j \right| + \left| \sum_{i=1}^{n-1} w_i \right| \\ &< \varepsilon. \end{aligned}$$

Por lo tanto, si $u \in (0,1)$ y se define $A = \{j \in \mathbb{N} \mid w_j > u\}$, entonces $\#(A) \leq N_u < \infty$. $\square$

Es claro que el proceso original P se puede recuperar de la siguiente manera.

$$P(\cdot) = \lim_{u \to 0} P(\cdot \mid u)$$

Esto se puede extender a una mezcla contable de kernels. Es decir, si a

$$P(x) = \sum_{i=1}^{\infty} w_i K(x \mid \theta_i)$$

le agregamos la variable latente u entonces