INSTITUTO TECNOLÓGICO AUTÓNOMO DE MÉXICO

Plantilla para Tesis y Tesinas del ITAM

TESIS

QUE PARA OBTENER EL TÍTULO DE

Lic. Matemáticas Aplicas

PRESENTA

Imanol Nuñez Morales

ASESOR

–

MÉXICO, D.F. 2017

Resumen

Este documento presenta una plantilla para usar en las tesis y tesinas del ITAM. Se provee de manera gratuita y sin ninguna responsabilidad bajo la licencia *creative commons BY-SA 3.0*.

Abstract

In this work we present a template for thesis and titulation works presented at ITAM. It is provided freely and without any responsability under the *creative commons BY-SA 3.0*.

Índice general

Índice de figuras

Índice de tablas

Capítulo 1

Introducción

Formal mathematics is nature's way of letting you know how sloppy your mathematics is.

– Leslie Lamport

Este trabajo presenta una plantilla para las tesis y tesinas del Instituto Tecnológico Autónomo de México para los usuarios de LaTeX[1]. Nace de la necesidad de los matemáticos, actuarios e ingenieros (entre otras carreras) por utilizar un sistema de composición de textos adecuado para su trabajo de titulación. El objetivo es ayudar a la comunidad del ITAM a simplificar el proceso de escritura y edición de sus tesis, tesinas o casos. A continuación describimos a mayor detalle cada una de las partes de la plantilla.

1.0.1. Descripción de los archivos

1. El archivo **tesisITAM.cls** define una nueva clase de documento con el mismo nombre. Esta clase se basa en el tipo reporte, el cual se emplea comúnmente para reportes de trabajo, pequeños libros y tesis.

2. El archivo **macros.sty** define macros y operadores adicionales, como por ejemplo, el argumento mínimo argmín. Además provee un espacio para que el estudiante agregue sus definiciones propias.

3. Este archivo **introduction.tex** describe el funcionamiento de la plantilla y de los archivos.

4. El archivo principal **main.tex** es un ejemplo básico de un archivo de tesis para generar este ejemplo.

5. El archivo **portada.tex** contiene el código necesario para generar la portada. El archivo **derechos.tex** contiene el texto para cede de derechos de publicación hacia el ITAM.

1.0.2. Características de la plantilla

La plantilla se basa en el documento tipo reporte. Todas las opciones de la clase *report* pueden ser utilizadas sin ninguna modificación (a4paper, 10pt, etc...). El espaciado del texto se establece a 1.33. Se eliminó la identación del párrafo, y el espaciado entre párrafos fue disminuido. Las sub-secciones se numeran utilizando números romanos. Se define un encabezado para cada página (salvo las que inician un capítulo) donde aparece el número del capítulo y el nombre de la sección.

I. Paquetes importados

La plantilla utiliza los siguientes paquetes: *graphicx*, *amsopn*, *fancyhdr*, y *babel*. El paquete de manejo de idiomas babel es importado con los idiomas **english** y *spanish* por defacto. Además, se seleccionan las opciones de uso de punto decimal y de nombrar las tablas como tablas y no como cuadros.

II. Opciones de clase

Se declara una opción adicional de nombre **tesina** para cambiar la portada a que se declare el trabajo como tesina.

III. Campos para el autor

Para el autor, se define los siguientes campos: **title**,**author**,**degree**,**advisor**, y **year**. *Title* define el título del trabajo de titulación, este título se presenta en la portada y al ceder los derechos de publicación. *Author* define el nombre completo del autor de la tesis. *Degree* se utiliza para establecer la carrera de la cual se va a titular, y debe incluir el texto completo (*i.e.*, Licenciatura en X o Ingeniero en C). *Advisor* recibe el nombre completo (con grado académico) del asesor, tal como se presentará en la portada. *Year* recibe el año de titulación para presentarlo en la portada.

IV. Generando la portada

El comando **\maketitle** produce la portada con los campos previamente definidos. Utiliza el logo del ITAM, el cual se debe encontrar en el **PATH** de LaTeX(*e.g.*, en el fólder de Figures).

V. Cediendo derechos de publicación

El comando **\publicationrights** imprime una página con el texto oficial de cede de derechos. Los nombres de la tesis (tesina) y del autor se obtienen de los campos previamente definidos.

VI. Añadiendo el resumen

Comúnmente, el resumen (o *abstract*) se debe escribir tanto en español como en inglés. Para esto, se define el ambiente **abstract** que recibe como parámetro un idioma. En el caso de que el idioma no sea ni inglés ni español, se recomienda que el idioma haya sido previamente importado como opción del paquete babel. En otro caso, el paquete imprimirá una advertencia.

VII. Citando a los grandes

La plantilla define un nuevo ambiente para las citas al principio del capítulo **chapterquote** que recibe dos parámetros: el autor y el texto. Para un ejemplo, referirse al principio de este capítulo.

1.1. Licencia

Esta plantilla se distribuye bajo la licencia *creative commons BY-SA 3.0*. Esta licencia permite la modificación de cualquier aspecto de la plantilla siempre y cuando se respeten las siguientes condiciones:

1. Que se mantenga la atribución del trabajo original, es decir, que se mencionen los autores originales y su afiliación, tal como se hace en el archivo original.
2. Que todas las modificaciones se hagan públicas y libres de acceso, sin recibir ningún tipo de retribución por el uso o distribución de la plantilla.

1.2. Autor

Capítulo 2

Marco teórico

2.1. Procesos de Dirichlet

2.1.1. Revisión general

I. 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, $\mathcal{F}(\chi)$, 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.[1] 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.

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

II. 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.
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[2] 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*.

Por esto el proceso Dirichlet sirve como base prior y se usa como fuente para ser generalizada. Antoniak propuso que α también fuera un parámetro, indexado por $u \sim H$, por lo que se tiene una mezcla de procesos de Dirichlet, $P \in \int \mathcal{D}(\alpha_u)\, dH(u)$. Dalal propuso 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 priores en un espacio de funciones de densidad. Tomando $\alpha(\mathfrak{X})$ 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*).

[2] Investigar

III. 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 priores conjugada. Lijoi y Prünster identifican 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 $\mathbf{X} \sim P \in \mathcal{D}(\alpha)$, Ferguson demostró que la distribución posterior, dada la muestra es también un proceso de Dirichlet, en particular $P \,|\, \mathbf{X} \in \mathcal{D}(\alpha + \sum_{i=1}^{n} \delta_{x_i})$. Esta propiedad 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: $F_0(\cdot) = \overline{\alpha}(\cdot) = \frac{\alpha(\cdot)}{\alpha(\mathfrak{X})}$ y $M = \alpha(\mathfrak{X})$. F_0 se interpreta como una adivinanza a priori (o distribución base) de F y M es el tamaño poblacional a prior o el parámetro de precisión (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 $P(A)$ 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 medidas de probabilidad discretas.[3] 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 $\theta \sim P \in \int_U \mathcal{D}(\alpha_u)\, dH(u)$ una muestra de tamaño n, entonces $P \,|\, \theta \in \int_U \mathcal{D}(\alpha_u + \sum_{i=1}^{n} \delta_{\theta_i})\, dH_\theta(u)$. Estos procesos son útiles en problemas

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

de evaluación biológica, pero obtener la expresión explícita de la distribución posterior es difícil por lo que se confía en procedimientos computacionales.

El proceso de Dirichlet es no paramétrico en el sentido de que tiene un soporte amplio; por esto Dalal vio la necesidad de que la prior debía de tener 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 muestreo 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 $\sum_{j=1}^{\infty} p_j \leq 1$.

2.1.2. Proceso 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 medidas de probabilidad discretas.

I. 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 conjunto de todas las medidas de probabilidad en $(\mathfrak{X}, \mathcal{A})$. ($P$ es el parámetro y $(\Pi, \sigma(Pi))$ es el espacio 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 $S_{k-1} = \left\{\mathbf{x} \in \mathbb{R}^{k-1} \mid \mathbf{x} \geq \mathbf{0} \text{ y } \|\mathbf{x}\|_1 \leq 1\right\}$ 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. Tomando $\{X_n \,|\, n \geq 1\}$ 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

$$P\left(X_{n+1} \in A \,|\, X_1, \ldots, X_n\right) = \frac{\alpha_n(A)}{\alpha_n(\mathfrak{X})} = \frac{\alpha(A) + \sum_{i=1}^{n} \delta_{x_i}(A)}{\alpha(\mathfrak{X}) + n}. \tag{2.1}$$

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 (2.1) se puede expresar como

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

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 $\mathfrak{X}$,*

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

casi seguramente.

II. 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

$$X_{n+1} \,|\, X_1, \ldots, X_n \sim \frac{M}{M+n} \overline{\alpha} + \frac{n}{M+n} \frac{1}{n} \sum_{j=1}^{K} n_j \delta_{X_j^*}. \tag{2.2}$$

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 $\overline{\alpha}$.

III. 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 $P \mid \theta \sim \mathcal{D}(\alpha_\theta)$. 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 $f(x) = \sum_{i=1}^{\infty} p_i \mathcal{N}(x \mid \mu_i, \sigma_i)$. Esto puede ser escrito como $f(x) = \int \mathcal{N}(x \mid \mu, \sigma)\, dG(\mu, \sigma)$. 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 $Y_i \mid \mu_i \sim \mathcal{N}(\mu_i, 1)$, $\mu_i \mid G \overset{iid}{\sim} G$ 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 $\left(\mu_i, \sigma_i^2\right)$ se tienen observaciones independientes $Y_1, \ldots, Y_n$ tales que $Y_i \mid \left(\mu_i, \sigma_i^2\right) \sim \mathcal{N}\left(\mu_i, \sigma_i^2\right)$ y $\nu_i = \left(\mu_i, \sigma_i^2\right)$ son muestreados de una distribución a priori G sobre $\mathbb{R}\times\mathbb{R}_{>0}$. Los autores asumen que $G \sim \mathcal{D}\left(MG_0\right)$ donde G_0 es la distribución a priori sobre $\mathbb{R}\times\mathbb{R}_{>0}$. Por ser un proceso Dirichlet $\nu_{n+1} \mid \nu_1, \ldots, \nu_n$ seguirá una distribución de la forma (2.2). De lo anterior proceden a derivar la distribución de $Y_{n+1} \mid \nu_1, \ldots, \nu_n$, que resulta ser una mezcla de n normales y una t de Student, con esto prubas que $Y_{n+1} \mid Y_1, \ldots, Y_n$ tiene la distribución predictiva $\int P\left(Y_{n+1} \mid \nu\right) dP\left(\nu \mid Y_1, \ldots, Y_n\right)$.

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 priores discretas; si los pesos definidos por $B\left(1, \alpha\left(\mathfrak{X}\right)\right)$ 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.

IV. 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 $\gamma\left(A\right) = \gamma\left(gA\right)$ 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.

Algunas propiedades de este proceso son las siguientes:

1. Sea $P \in \mathcal{DGI}(\alpha)$ y X una muestra de tamaño 1 de P, entonces para $A \in \sigma(\mathfrak{X})$ se tiene que

$$P(X \in A) = P(X \in gA) = \frac{\alpha(A)}{\alpha(\mathfrak{X})} \text{ para cualquier } g \in \mathcal{G}.$$

2. Si $P \in \mathcal{DGI}(\alpha)$ entonces P es una medida de probabilidad discreta con probabilidad 1.

3. Se tiene la siguiente propiedad de conjugacidad:

   Teorema 2 (Dalal) *Sea $P \in \mathcal{DGI}(\alpha)$ y sea $X_1, \ldots, X_n$ una muestra de tamaño n de P. Entonces la distribución posterior de P dadas $X_1, \ldots, X_n$ es $\mathcal{DGI}\left(\alpha + \sum_{i=1}^{n} \delta_{X_i}^{g}\right)$ donde $\delta_{X_i}^{g} = \frac{1}{k} \sum_{j=1}^{k} \delta g_j X_i$.*

2.1.3. Mezclas de procesos Dirichlet

Hay algunas situaciones en las que el proceso Dirichlet no es adecuado como en el siguiente problema de *bioassay*. Sea $F(t)$ la probabilidad de una respuesta positiva de un animal a una medicina administrada con nivel $t \geq 0$. Se asume que $F(0) = 0$ y que $F(t)$ es no decreciente con $\lim_{t \to \infty} F(t) = 1$. Para entender propiedades de F se hace un experimento con n animales, administrándoles

$t_1, \ldots, t_n$ de la medicina y observando las variables independientes $Y_1, \ldots, Y_n$ que representan si la respuesta al tratamiento es positiva. Este problema se puede tratar desde un punto de vista bayesiano no paramétrico escogiendo como prior un proceso Dirichlet con parámetro α para F. Pero lo que queremos es, dados los datos, obtener la distribución posterior de F y esta no es un proceso Dirichlet.

Para el problema anterior y otros problemas relacionados, la distribución posterior de F dados los datos resulta ser una mezcla de procesos Dirichlet. *Grosso modo*, una mezcla de procesos Dirichlet es un proceso Dirichlet donde el parámetro α es aleatorio y se la da una cierta distribución. Dalal y Hall mostraron que un modelo paramétrico bayesiano se puede aproximar por un modelo bayesiano no paramétrico con mezclas de modelos Dirichlet, por lo que el prior asigna casi todo su peso a vecindades de los modelos paramétricos; además cualquier prior paramétrica o no paramétrica se puede aproximar por una mezcla de procesos Dirichlet. Esto se está usando para modelar datos de dimensiones altas a gran escala.

La mezcla más simple de procesos Dirichlet sería la que escoja $P \in \mathcal{D}(\alpha_1)$ con probabilidad π y que escoja $P \in \mathcal{D}(\alpha_2)$ con probabilidad $1 - \pi$. Este tipo de mezclas no se debe confundir con las que propuso Lo en la que un kernel conocido se mezcla respecto a una distribución no paramétrica.

I. Definición

Primero se presenta la defినción de una medida de transición.

Definición 6 (Antoniak) *Sean $(\mathfrak{X}, \sigma(\mathfrak{X}))$ y $(U, \sigma(U))$ dos espacios medibles. Una medida de transición es un mapeo de $U \times \sigma(\mathfrak{X})$ a $[0, \infty)$ tal que se cumplan los dos siguientes requisitos.*

1. *Para cada $u \in U$, $\alpha(u, \cdot)$ debe ser una medida finita no negativa y no nula en $(\mathfrak{X}, \sigma(\mathfrak{X}))$.*

2. *Para cada $A \in \sigma(\mathfrak{X})$, $\alpha(\cdot, A)$ debe ser medible en $(U, \sigma(U))$.*

Por conveniencia, en lugar de usar $\alpha(u, \cdot)$, se utiliza la notación $\alpha_u(\cdot)$ y esta medida sirve como parámetro para el proceso Dirichlet. Ahora se da la definición de una mezcla de procesos Dirichlet.

Definición 7 (Antoniak) *Se dice que P es una mezcla de procesos Dirichlet en $(\mathfrak{X}, \sigma(\mathfrak{X}))$ con distribución de mezcla H en $(U, \sigma(U))$ y medida de transición α_u si para toda $k = 1, 2, \ldots$ y toda partición medible $A_1, \ldots, A_k$ de $\mathfrak{X}$, se tiene que el vector aleatorio $(P(A_1), \ldots, P(A_k))$ tiene la distribución como mezcla*

$$\int_U D(\alpha_u(A_1), \ldots, \alpha_u(A_k))\, dH(u),$$

en donde $D(\alpha_1, \ldots, \alpha_k)$ es la distribución Dirichlet k-dimensional. Por conveniencia se escribe $P \in \int D(\alpha_u(\cdot))\, dH(u)$.

Si se considera que u es una variable aleatoria con distribución H entonces P, condicional a u, es un proceso Dirichlet con parámetro α_u. Una muestra de una mezcla de procesos Dirichlet tiene una definición similar a la muestra de un proceso Dirichlet.

Definición 8 (Antoniak) *Sea P una mezcla de procesos Dirichlet en $(\mathfrak{X}, \sigma(\mathfrak{X}))$ con medida de transición α_u y distribución de mezcla H. $X_1, \ldots, X_n$ es una muestra de P si para cada entero positivo m y conjuntos medibles $A_1, \ldots, A_m$, $C_1, \ldots, C_n$ se cumple lo siguiente:*

$$P\{X_1 \in C_1, \ldots, X_n \in C_n \mid u, P(A_1), \ldots, P(A_m), P(C_1), \ldots, P(C_n)\}$$
$$= \prod_{j=1}^{n} P(C_j) \text{ casi seguramente.}$$

II. Propiedades

Antoniak estableció las siguientes propiedades.

1. Sea $P \in \mathcal{D}(\alpha)$ y H una media de probabilidad fija en $(\mathfrak{X}, \sigma(\mathfrak{X}))$, si $\alpha_u(A) = \alpha(A) + \delta_u(A)$, entonces el proceso P^* que escoge u respecto a H y P de un proceso Dirichlet con parámetro α_u es una mezcla de procesos Dirichlet.

2. Si $P \in \int \mathcal{D}(\alpha_u(\cdot))\, dH(u)$ y X es una muestra de tamaño uno de P, entonces para todo conjunto medible A se cumple que

$$P(X \in A) = \int_U \frac{\alpha_u(A)}{\alpha_u(\mathfrak{X})} dH(u).$$

3. Sean $P \in \mathcal{D}(\alpha)$, X una muestra de tamaño uno de P y A un conjunto medible tal que $\alpha(A) > 0$. Entonces la distribución condicional de P dado que $X \in A$ es una mezcla de procesos Dirichlet con medida de transición $\alpha_u = \alpha + \delta_u$ para $u \in A$ y distribución de mezcla $H_A(\cdot) = \frac{\alpha(\cdot)}{\alpha(A)}$. Si $A = \mathfrak{X}$ entonces esto se reduce al proceso de Dirichlet.

4. Las mezclas de procesos Dirichlet satisfacen la siguiente propiedad de conjugacidad.

   Teorema 3 (Antoniak) *Sea $\mathbf{X} = (X_1, \ldots, X_n)$ una muestra de tamaño n de $P \in \int_U \mathcal{D}(\alpha_u)\, dH(u)$, entonces$P \mid \mathbf{X} \in \int_U \mathcal{D}(\alpha_u + \sum_{i=1}^{n} \delta_{\theta_i})\, dH_{\mathbf{X}}$, donde $H_{\mathbf{X}}$ es la distribución condicional de u dada $\mathbf{X}$.*

Apéndice A

Elementos de probabilidad y estadística

A.1. Prior y marginal

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

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

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

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

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

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

A.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]:

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

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

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

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 (A.1), (A.4) y (A.5), está dada por:

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

por lo que $\theta \mid x_1, \ldots, x_n \sim Dirichlet\left(\underline{\alpha} + \underline{n'}\right)$ en donde $\underline{n'} = (n'_1, \ldots, n'_k)$, con $n'_i = \sum_{j=1}^{n} [x_j = i]$, 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*.

[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 $\underline{\alpha} = (\alpha_1, \ldots, \alpha_k)$.

[3]Categórica-Dirichlet

A.3. Modelo Normal-Wishart Inversa

Sea $V \in \mathbb{R}^{p \times p}$ una matriz simétrica positiva definida. Se dice que V sigue ladistribución Wishart p-dimensional con matriz de escala Σ y n grados de libertad si $p \leq n$ y la distribución conjunta de los elementos de V tiene la función de densidad

$$p(V) = \frac{c\,|V|^{\frac{n-p-1}{2}}}{|\Sigma|^{\frac{n}{2}}} \exp\left(-\frac{1}{2}\mathrm{tr}\left(\Sigma^{-1}V\right)\right) \tag{A.6}$$

si $V > 0$ y $\Sigma > 0$ y $p(V) = 0$ en otro caso. c es una constante numérica definida por

$$c^{-1} = 2^{\frac{np}{2}}\pi^{\frac{p(p-1)}{4}} \prod_{j=1}^{p} \Gamma\left(\frac{n+1-j}{2}\right).$$

Si $p > n$ entonces la distribución es singular y no cuenta con densidad. Simbólicamente se escribe $V \sim \mathcal{W}(\Sigma, p, n)$.

Si definimos $U = V^{-1}$ entonces de puede probar que U tiene la siguiente densidad.

$$p(U) = \frac{c\,|S|^{\frac{n}{2}}}{|U|^{\frac{n+p+1}{2}}} \exp\left(-\frac{1}{2}\mathrm{tr}\left(SU^{-1}\right)\right), \tag{A.7}$$

con c definido igual que en (A.6) y $S = \Sigma^{-1}$. Se dice que U sigue la distribución Wishart inversa. Simbólicamente $U \sim \mathcal{W}^{-1}(S, p, n)$.

En el caso de que $X \sim \mathcal{N}(\mu, \Sigma)$ es p-dimensional donde μ y Σ son desconocidas. Como distribución a priori para μ y Σ se puede utilizar la Normal-Wishart inversa. Para esto se necesitan cumplir las siguientes condiciones:

$$\mu \,|\, \mu_0, \Sigma, \lambda \sim \mathcal{N}\left(\mu_0, \frac{1}{\lambda}\Sigma\right), \tag{A.8}$$

$$\Sigma \,|\, S, \nu \sim \mathcal{W}^{-1}(S, p, n). \tag{A.9}$$

Entonces si la densidad de $\mu, \Sigma \,|\, \mu_0, S, \nu, \lambda$ está dada por

$$f(\mu, \Sigma \,|\, \mu_0, S, \nu, \lambda) = \mathcal{N}\left(\mu \,|\, \mu_0, \frac{1}{\lambda}\Sigma\right)\mathcal{W}^{-1}(\Sigma \,|\, S, \nu),$$

se dice que μ, Σ tiene una distribución Normal-Wishart inversa con parámetros $\mu_0, S,\ \nu,\ \lambda$. Simbólicamente se escribe $\mu, \Sigma \sim \mathcal{N}\mathcal{W}^{-1}(\mu_0, S, \nu, \lambda)$.

Ahora supongamos que con μ y Σ desconocidas se tiene una colección de n observaciones tales que $X_i \overset{iid}{\sim} \mathcal{N}(\mu, \Sigma)$ se tiene que

$$
\begin{array}{rcl}
p(\mu, \Sigma \mid X_1, \ldots, X_n) & \propto & p(X_1, \ldots, X_n \mid \mu, \Sigma)\, p(\mu, \Sigma) \\
& \propto & |\Sigma|^{-\frac{n}{2}} \exp\left(-\frac{1}{2}\sum_{i=1}^{n}(X_i - \mu)^T \Sigma^{-1}(X_i - \mu)\right) \\
& & \cdot |\Sigma|^{-\frac{1}{2}} \exp\left(-\frac{\lambda}{2}(\mu - \mu_0)^T \Sigma^{-1}(\mu - \mu_0)\right) \\
& & \cdot |\Sigma|^{-\frac{\nu+p+1}{2}} \exp\left(-\frac{1}{2}\mathrm{tr}\left(S\Sigma^{-1}\right)\right) \\
& \propto & |\Sigma|^{-\frac{1}{2}} \exp\left(-\frac{\lambda_n}{2}(\mu - \mu_n)^T \Sigma^{-1}(\mu - \mu_n)\right) \\
& & \cdot |\Sigma|^{-\frac{\nu+n+p+1}{2}} \exp\left(S_n \Sigma^{-1}\right) \\
& \propto & \mathcal{N}\left(\mu \mid \mu_n, \frac{1}{\lambda_n}\Sigma\right) \mathcal{W}^{-1}(\Sigma \mid S_n, \nu_n),
\end{array}
$$

en donde

$$
\begin{array}{rcl}
\mu_n & = & \frac{\lambda\mu_0 + n\overline{X}}{\lambda + n}, \\
\lambda_n & = & \lambda + n, \\
\nu_n & = & \nu + n \qquad \text{y} \\
S_n & = & S + \sum_{i=1}^{n}\left(X_i - \overline{X}\right)\left(X_i - \overline{X}\right)^T + \frac{\lambda n}{\lambda + n}\left(\overline{X} - \mu_0\right)\left(\overline{X} - \mu_0\right)^T.
\end{array}
$$

Entonces se puede observar que la distribución posterior de los parámetros μ y Σ sigue siendo una Normal-Wishart inversa.

Bibliografía

[1] Leslie Lamport. *LaTeX: a document preparation system.* Addison-Wesley, 1994.