

INSTITUTO TECNOLÓGICO AUTÓNOMO DE MÉXICO



# Título

TESIS

QUE PARA OBTENER EL TÍTULO

LICENCIADO EN ACTUARÍA

PRESENTA

ADRIANA PÉREZ ARCINIEGA SOBERÓN

ASESOR: ...

MÉXICO, D.F.

2016

Con fundamento en los artículos 21 y 27 de la Ley Federal del Derecho de Autor y como titular de los derechos moral y patrimonial de la obra titulada "**TÍTULO DE LA TESIS**", otorgo de manera gratuita y permanente al Instituto Tecnológico Autónomo de México y a la Biblioteca Raúl Baillères Jr., la autorización para que fijen la obra en cualquier medio, incluido el electrónico, y la divulguen entre sus usuarios, profesores, estudiantes o terceras personas, sin que pueda percibir por tal divulgación una contraprestación".

AUTOR

---

FECHA

---

FIRMA

# Índice general

<b>1. Introducción</b>	<b>5</b>
1.0.1. Introducción . . . . .	5
1.0.2. Definición de Variables Latentes . . . . .	5
<b>2. Procesos de Duración Marcada</b>	<b>6</b>
2.1. Introducción . . . . .	6
2.2. Definiciones . . . . .	7
2.3. Procesos de Duración Marcada . . . . .	11
2.4. Construcción de Procesos vía Variables Latentes . . . . .	16
2.5. Construcción del Modelo de Marcas . . . . .	22
2.6. Discusión . . . . .	23
<b>3. Inferencia y Predicción</b>	<b>25</b>
3.1. Introducción . . . . .	25
3.2. Verosimilitud Extendida . . . . .	26
3.3. Slice Sampler . . . . .	35
3.4. El Muestreador de Gibbs . . . . .	38
<b>4. Ilustración del Modelo con Datos</b>	<b>44</b>
4.1. Introducción . . . . .	44
4.2. Diabetes Mellitus en el mundo: Costos y Prevalencia . . . . .	44

4.3. Descripción de los datos . . . . .	52
4.4. Análisis Descriptivo . . . . .	52
4.5. Resultados . . . . .	52
<b>5. Conclusiones</b>	<b>53</b>
5.1. Qué me deja este trabajo? . . . . .	53
5.2. Visión Crítica Autónoma . . . . .	53
5.3. Trabajo Futuro . . . . .	53
<b>A. Paradigma Bayesiano</b>	<b>58</b>

# Agradecimientos

Muchas gracias a todos!

# Capítulo 1

## Introducción

### 1.0.1. Introducción

Introducción general a la tesis

### 1.0.2. Definición de Variables Latentes

$$P(x) = \int p(x|\theta)f(d\theta)$$

$\theta$  es una variable no observable o latente.

## Capítulo 2

# Procesos de Duración Marcada

### 2.1. Introducción

En este capítulo se hablará sobre los modelos de duración y de duración marcada y su aplicación en el objeto de este trabajo. Además de algunas propiedades, tales como la independencia, intercambiabilidad y, por supuesto, estacionariedad que son vitales para realizar inferencia y predicción de los datos.

Una vez que se tiene definido el proceso puntual se introduce el concepto de variables latentes para poder construir las distribuciones de las variables observables. Estas variables conectan las observaciones a través del tiempo, creando el modelo general de probabilidad que permite hacer inferencia y predicción sobre futuras observaciones. Tomando el modelo general de probabilidad, se eligen las distribuciones que mejor se adaptan al comportamiento de las variables.

Es importante remarcar que la historia de los procesos puntuales siempre ha estado unida a aquella de la estadística actuarial y de seguros, como nos mencionan [Daley and Vere-Jones, 2003] al referirse a las tablas de mortalidad como el primer estudio de procesos de intervalos. Por lo que el empleo de estos procesos como un método de tarificación es solamente otra colaboración en la larga lista de estas dos disciplinas.

## 2.2. Definiciones

Para el objeto de este estudio tenemos una muestra de microcostos de enfermedades crónicas de un cierto número de individuos a los que se les ha observado durante un período de tiempo. A su vez, cada uno de los individuos tiene asociadas covariables sociodemográficas, socio-económicas y médicas; estas covariables, por el alcance del trabajo no están incluidas en la modelación. Aunque las covariables asociadas a los individuos inciden en la trayectoria del padecimiento, estas relaciones son un tema a explorarse en un trabajo de investigación futuro.

De este modo, podríamos decir que tenemos  $i = 1, \dots, I$  individuos, donde  $I$  es el número total de individuos observados por un período de tiempo con costos asociados a su padecimiento. El objetivo es no solo modelar y predecir la duración y el costo de las etapas de estos padecimientos por individuo, sino también, introducir la estructura de dependencia entre éstas. De acuerdo a [Daley and Vere-Jones, 2003], las relaciones de dependencia entre las variables pueden llegar a ser muy complejas, por lo que la ejecución del modelo se vuelve más complicada.

Entonces, supongamos que empezamos el estudio de un individuo  $i$  en el tiempo  $t_{i0} = 0$ , es decir, este es el tiempo en el que el individuo  $i$  entra



al panel de estudio. La duración del estudio para el individuo es  $T_i$ , esto no quiere decir que no puedan ocurrir observaciones posteriores a  $T_i$ , a esto se le conoce como censuramiento de datos por la derecha.

Según [Paik Schoenberg, 2000], un proceso puntual es una medida aleatoria en un espacio métrico separado  $S$  tomando valores en los enteros no negativos  $Z^+$  donde  $N(t)$ , en un caso particular, es un proceso de conteo del número de puntos que ocurren antes del tiempo  $t$ .

Sea  $t_{ij} \in (t_{i0}, T_i]$  el momento en el que ocurre el  $j$ -ésimo cambio de tratamiento del  $i$ -ésimo individuo, por lo que definimos la variable aleatoria  $N(t)$  que cuenta el número de cortes o cambios en el intervalo.

Dado que la muestra consiste en microcostos a través del tiempo de un individuo  $i$ , decimos que a cada  $t_{ij}$  se le asocia la variable costo de tratamiento; es decir, a cada momento en que ocurre un cambio de tratamiento le corresponde un nuevo precio  $p_{ij}$ . De este modo, para cualquier  $i$  tenemos una sucesión de variables asociadas  $\{t_{i1}, p_{i1}\}, \{t_{i2}, p_{i2}\}, \dots, \{t_{ik}, p_{ik}\}$ . De este modo la sucesión de variables es una colección aleatoria de puntos en un espacio con una marca asociada a cada punto, así ya se pueden modelar los datos como en un proceso puntual marcado.

[Daley and Vere-Jones, 2003] definen el proceso puntual marcado como un proceso localizado en un espacio métrico completamente separado  $\chi$  y las marcas en otro espacio métrico completamente separado  $\kappa$ , entonces  $\{(\chi_i, \kappa_i)\}$  en  $\chi \times \kappa$  es un proceso puntual marcado con la propiedad adicional de que el proceso primario  $N(t)$  es a su vez un proceso puntual.

Lo que deseamos conocer es,

$$P(t_{i1}, \dots, t_{ik}, p_{i1}, \dots, p_{ik}) = P(t_{i1}, \dots, t_{ik}, p_{i1}, \dots, p_{ik} | N(t)) \quad (2.1)$$

Es decir, la función de distribución conjunta del tiempo de ocurrencia de los eventos y los precios asociados a estos es igual a la función de distribución de estas variables condicionados por la variable aleatoria del número de eventos en el intervalo  $(t_{i0}, T_i]$ . Sin embargo, dado que al usar las variables en sus valores absolutos estas pueden dar saltos muy altos entre si, se deben usar variables alternas.

Definimos las siguientes variables para un individuo  $i$ :

- $d_{ij} = t_{ij} - t_{ij-1}$ , donde  $d_{ij}$  es la duración entre los tiempos de ocurrencia de cada individuo.
- $c_{ij} = p_{ij} - p_{ij-1}$ , donde  $c_{ij}$  representa el costo, es decir, la diferencia entre los precios en cada tiempo de ocurrencia de cada individuo.

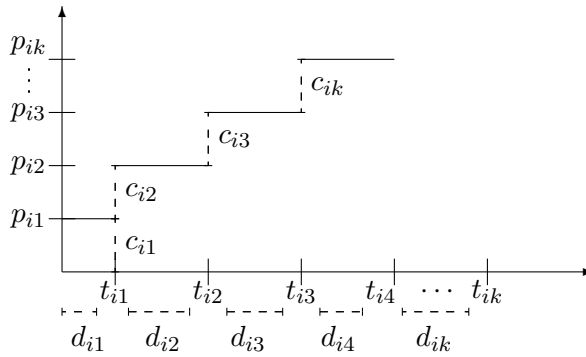


Figura 2.1: Trayectorias del individuo  $i$ .

De este modo,

$$P(t_{i1}, \dots, t_{ik}, p_{i1}, \dots, p_{ik} | N(t) = k) \cong P(d_{i1}, \dots, d_{ik}, c_{i1}, \dots, c_{ik} | N(t) = k) \quad (2.2)$$

Esto quiere decir que calcular la función de distribución conjunta de los tiempos de ocurrencia y los precios asociados a éstos es análogo a calcular la función de distribución conjunta de las duraciones y los costos asociados condicionados a la variable aleatoria del número de eventos en el intervalo de tiempo. Así pasamos de un proceso puntual marcado a uno de duración marcada.

El modelo de probabilidad consiste en dos variables, duración y costo. La relación entre estas dos variables es una de las siguientes opciones,

Las variables son independientes entre si, es decir,

$$\begin{aligned} P(d_{i1}, \dots, d_{ik}, c_{i1}, \dots, c_{ik} | N(t) = k) &= P(d_{i1}, \dots, d_{ik} | N(t) = k) \\ &\times P(c_{i1}, \dots, c_{ik} | N(t) = k) P(N(t) = k) \end{aligned}$$

Las variables no son independientes entre si. A su vez, la dependencia entre variables puede tener dos maneras de expresarse,

Las duraciones dependen de los costos. Es decir,

$$\begin{aligned} P((d_1, c_1), \dots, (d_k, c_k) | N(T) = k) &= P(d_1, \dots, d_k | N(T) = k, c_1, \dots, c_k) \\ &\times P(c_1, \dots, c_k | N(T) = k) \times P(N(T) = k) \end{aligned}$$

Los costos dependen de las duraciones, como en el artículo de [Engle

and Russell, 1998]. Es decir,

$$\begin{aligned} P((d_1, c_1), \dots, (d_k, c_k) | N(T) = k) &= P(c_1, \dots, c_k | N(T) = k, d_1, \dots, d_k) \\ &\times P(d_1, \dots, d_k | N(T) = k) \times P(N(T) = k) \end{aligned}$$

En este caso, los precios dependen del cambio de tratamiento, de manera análoga, la variable costo está asociada a la variable de duración. Es decir, que aunque la marca se localice en otro espacio métrico, esta sigue anclada al proceso puntual primario.

Por la estructura de construcción del modelo este se puede pensar, para un solo momento  $k$  en el tiempo de un individuo  $i$ , como

$$P(t_{ik}, p_{ik} | N(t)) \cong P(d_{ik}, c_{ik} | N(t)) = P(c_{ik} | d_{ik}) P(d_{ik}) \quad (2.3)$$

## 2.3. Procesos de Duración Marcada

Una vez que hemos definido qué es el proceso de duración y de duración marcada y cómo es que los datos que tenemos para este estudio se adaptan a este modelo, necesitamos especificar las propiedades que van a hacer posible la inferencia y la predicción. Estas propiedades son la independencia, la intercambiabilidad y, principalmente, la estacionariedad.

### Independencia

En una concepción tradicional, [Resnick, 1999] define la independencia de un número finito de eventos como:

**Definición 1.** *Los eventos  $A_1, \dots, A_n$  ( $n \geq 2$ ) son independientes si*

$$P\left(\bigcap_{i \in I} A_i\right) = \prod_{i \in I} P(A_i), \quad I \subset \{1, \dots, n\}$$

Los eventos son independientes si la probabilidad de la intersección de estos eventos o la probabilidad conjunta de los eventos es igual a la multiplicación de la probabilidad de los mismos.

Análogamente, podemos hacer la definición de independencia para el proceso de duración marcada. Recordemos que tenemos la función de probabilidad conjunta de las duraciones y los costos, por lo que la independencia en el proceso es:

$$P(d_1, c_1, \dots, d_k, c_k | N(t) = k) = \prod_{j=1}^{N(t)} P(d_j, c_j)$$

En este caso, la única diferencia reside en el hecho de que el número de funciones de probabilidad a multiplicar es a su vez una variable aleatoria, la cual se encarga de contar los cambios en el costo de tratamiento en el tiempo. El supuesto de independencia es útil para la inferencia de futuras observaciones.

## Intercambiabilidad

Otra propiedad muy importante para la inferencia y predicción de variables en un proceso de duración marcada es la intercambiabilidad que, de acuerdo a [Hahn and Zhang, 2012], se define como:

**Definición 2.** *Una sucesión de variables  $X = (X_1, X_2, \dots, X_n)$  es inter-*

cambiable si para cada operador de permutación  $\sigma(\cdot)$  de  $\{1, \dots, n\}$

$$P(X_1, X_2, \dots, X_n) = P(X_{\sigma(1)}, X_{\sigma(2)}, \dots, X_{\sigma(n)})$$

Si la sucesión de variables es independiente e idénticamente distribuida entonces es intercambiable. El concepto de intercambiabilidad está muy relacionado con la independencia, pues la independencia es un caso particular de la intercambiabilidad.

Para poder entender mejor la propiedad podemos citar el Teorema de Fenetti(1937) que nos dice,

**Teorema 3.** *Una sucesión infinita de variables aleatorias intercambiables  $\bar{X} = (X_1, X_2, \dots)$  es una mezcla de variables independientes e idénticamente distribuidas (i.i.d). Esto es, que existe un espacio de probabilidad  $(U, \Theta)$  tal que*

$$P(\bar{X} \in B) = \int_U P(\bar{X}(u) \in B) \Theta(du)$$

donde  $\bar{X}(u) = (X_1(u), X_2(u), \dots)$  es una secuencia de variables aleatorias i.i.d. y  $\Theta(\cdot)$  es una medida de probabilidad.

Esto se puede adaptar al proceso de duración marcada correspondiente a este análisis de la siguiente manera, tomando el Teorema de Fenetti

$$P(d_1, c_1, \dots, d_k, c_k | N(t) = k) = \int_{\Theta} \prod_{j=1}^{N(t)} P(d_j, c_j | \theta) \pi(\theta) d(\theta)$$

donde  $\theta$  es una variable aleatoria no observable y  $\pi(\theta)$  es una medida de probabilidad común a todas las variables aleatorias. Es decir, que a lo postulado en el apartado de independencia le agregamos la variable no

observable con su respectiva medida de probabilidad, sobre cuyo espacio de probabilidad está definida la integral. La variable no observable común a todas las variables aleatorias es un tema que se posteriormente se desarrollará con mayor profundidad.

## Estacionareidad

Una vez que han sido definidas la independencia y la intercambiabilidad faltaría definir la estacionareidad para poder hacer predicciones sobre futuras observaciones.

De manera intuitiva, podemos definir la estacionareidad en un proceso de duración cuando la función de probabilidad conjunta del proceso no cambia cuando es ésta es desplazada en el tiempo, lo cual indicaría que lo importante es la longitud de los intervalos, no la localización de los mismos. Sin embargo, de una manera más técnica, [Daley and Vere-Jones, 2003] definen la estacionareidad en un proceso como:

**Definicion 4.** *Un proceso puntual es estacionario por intervalos cuando para cada  $r = 1, 2, \dots$  y todos los enteros  $i_1, \dots, i_r$ , la distribución conjunta de  $\{\tau_{i_1+k}, \dots, \tau_{i_r+k}\}$  no depende de  $k$  ( $k = 0, \pm 1, \dots$ ).*

Esto implicaría que el orden de las observaciones importa y que las observaciones pasadas ayudan a construir la variable aleatoria. Es decir que con una sucesión de variables  $\vec{X} = (X_1, \dots, X_n)$  tendríamos que,

$$\begin{aligned} P(X_n, \dots, X_1) &= P(X_n | X_{n-1}, \dots, X_1) \times P(X_{n-1} | X_{n-2}, \dots, X_1) \times \dots \\ &\times P(X_2 | X_1) \times P(X_1) \end{aligned}$$

Así si la variable aleatoria depende de su historia, podríamos entonces

predecir observaciones futuras. Es decir, que para toda  $s \geq 0$

$$\begin{aligned} P(X_{n+1}, X_n, \dots, X_1) &= P(X_{n+1} | X_n, \dots, X_1) \\ &= P(X_{n+s+1} | X_{n+s}, \dots, X_{1+s}) \end{aligned}$$

De este modo, para el proceso de duración marcada la estacionareidad, junto con la noción desarrollada en la ecuación (3) sobre la relación de las marcas con el proceso de duración, se podría plantear como

$$\begin{aligned} P(d_1, c_1, \dots, d_k, c_k | N(t) = k) &= P(d_1, c_1) \prod_{j=2}^{N(t)} P(d_j, c_j | d_{j-1}, c_{j-1}) \\ &= P(c_1 | d_1) P(d_1) \prod_{j=2}^{N(t)} P(d_j | d_{j-1}) P(c_j | d_j, c_{j-1}) \end{aligned}$$

Lo que quiere decir que la función conjunta de probabilidad se puede definir con base a observaciones pasadas.

Citando a [Daley and Vere-Jones, 2003] el Proceso Puntual Marcado es estacionario si la estructura de probabilidad del proceso no cambia a pesar de los cambios que puedan existir en el espacio métrico  $\chi$ , es decir, en el espacio métrico del proceso de conteo primario. De acuerdo a lo desarrollado en la sección anterior se concluye que tanto el proceso primario como el Proceso Puntual Marcado, ambos son estacionarios, lo que permitiría la inferencia sobre el mismo.

Una vez que nuestro modelo de duración marcada cumple las propiedades descritas en esta sección podemos empezar a hacer inferencia sobre las variables y predecir las observaciones futuras. En la siguiente sección, desarrollaremos un modelo complementario de variables latentes que terminaría de conectar la idea de la variable no observable presentada en



el concepto de intercambiabilidad con el resto de la sucesión.

## 2.4. Construcción de Procesos vía Variables Latentes

Como formulado al final de la sección pasada, el modelo general de probabilidad se escribe como

$$P((d_{ij}, c_{ij})_{j=1}^{N(t)}) = P(c_1|d_1)P(d_1) \prod_{j=2}^{N(t)} P(d_j|d_{j-1})P(c_j|d_j, c_{j-1})$$

Se puede observar en el modelo general que, gracias a la estacionariedad del mismo, es necesario conocer las distribuciones de una variable en un momento  $j$  condicionada a la misma variable en el momento  $j - 1$ . Es decir, se necesitan conocer las distribuciones  $f(d_j|d_{j-1})$  y  $f(c_j|c_{j-1}, d_j)$ . El método de para conocer estas distribuciones es análogo entre ambas, por lo que se desarrollará el proceso en la distribución de las duraciones para después extenderla a aquella de los costos.

Para lograr construir estas distribuciones se supone que no se conoce la relación entre las variables observables, en este caso las variables observables son las duraciones, a través del tiempo. Para efecto de la construcción del modelo se supone también que existe una variable no observable, o latente, que conecta a las observaciones.

Tomando como base el método propuesto por [Pitt et al., 2002], la distribución  $f(d_j|d_{j-1})$  se puede reescribir como  $f_{Y|Z}(y|z)$  siendo  $Y = t_j$  y

$Z = t_{j-1}$ . Esta distribución a su vez se puede reescribir como

$$f_Y(y) = \int f_{Y|Z}(y|z)f_Z(z)dz$$

Para lograr la conexión entre las variables observables se necesita agregar otra variable, que le llamaremos latente, a esta distribución por lo que consideramos ahora las distribuciones de transición como

$$f_{Y|Z}(y|z) = \int f_{Y|\Theta}(y|\theta)f_{\Theta|Z}(\theta|z)d\lambda(\theta)$$

donde  $\lambda(\theta)$  es una medida de probabilidad correspondiente a la variable latente.

El punto crucial para asegurar la distribución de transición es construir una distribución conjunta  $f_{Y,\Theta}(y,\theta)$ , tal que las densidades condicionales sean  $f_{Y|\Theta}(y|\theta)$  y  $f_{\Theta|Z}(\theta|z)$  y con una distribución marginal  $f_Y(y)$ . Es decir, el proceso de transición se logra mediante el proceso de la variable latente, que aunque no es observable, este es conocido.

Según lo anterior, vemos que para conocer la distribución de una observación condicionada a la observación anterior se necesita introducir una variable no observable o latente. La introducción de esta variable se puede hacer mediante el Modelo Oculto de Markov, que de acuerdo con [Ghahramani, 2001], es un modelo donde la variable observable  $y_t$  y la variable no observable  $\theta_t$  son independientes y la variable no observable cumple la propiedad markoviana. Es decir, que dado el valor de  $\theta_{t-1}$ , el estado actual  $\theta_t$  es independiente de todos aquellos estados previos a

$t - 1$ . La función de distribución conjunta sería

$$P((y_i, \theta_i)_{i=1}^t) = P(y_1|\theta_1)P(\theta_1) \prod_{i=2}^{N(t)} P(y_i|\theta_i)P(\theta_i|\theta_{i-1})$$

Es decir,

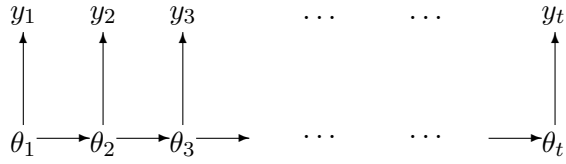


Figura 2.2: Representación gráfica de un Modelo Oculto de Markov.

El Modelo Oculto de Markov es un modelo muy flexible, según [Ghahramani, 2001], puede ser utilizado para modelar cualquier distribución con un número infinito de componentes y con las adecuaciones correspondientes se pueden modelar un sinnúmero de problemas dinámicos no-lineales; esta misma flexibilidad es lo que le resta fiabilidad a la inferencia. Aunado a esto, el modelo se basa en la relación existente entre los estados en el tiempo del parámetro, es decir,  $P(\theta_i|\theta_{i-1})$ , esta relación no es conocida en el problema de estudio del presente trabajo por lo que el Modelo Oculto de Markov no es el que mejor se adapta.

Tomando como base los Modelos Dinámicos Lineales descritos por [Harrison and West, 1999], el proceso de la variable latente para el caso de las duraciones, sería

Ecuación de Observación:  $d_j|\theta_j \sim f_{d|\Theta}(\cdot|\theta_j)$

Ecuación de Sistema:  $\theta_j|d_{j-1} \sim f_{\Theta|d}(\cdot|d_{j-1})$

Una vez que tenemos el modelo para la variable de las duraciones, lo hacemos extensivo para los costos. De este modo, el proceso de variables latentes quedaría

Ecuación de Observación:  $c_j | \gamma_j \sim f_{c|\gamma}(\cdot | \gamma_j)$

Ecuación de Sistema:  $\gamma_j | c_{j-1} \sim f_{\gamma|c}(\cdot | c_{j-1})$

Dado que el Modelo Oculto de Markov no necesariamente es estacionario afectando así la capacidad inferencial del mismo, el modelo que se construye a través de las estructuras de ecuaciones de observación y sistema que se acaban de definir, es estacionario por construcción y es capaz de modelar la observación actual en relación a la observación anterior. Tomando solamente la variable duración junto con su parámetro  $\theta$ , la representación gráfica de ambos modelos se compara del siguiente modo,

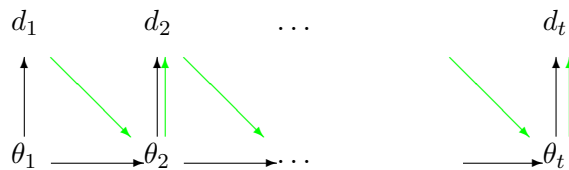


Figura 2.3: Comparación entre un Modelo Oculto de Markov y un modelo basado en Modelos Dinámicos Lineales

Donde las flechas negras representan un Modelo Oculto de Markov, don-

de la relación con el estado anterior se presenta solamente en los parámetros  $\theta_i|\theta_{i-1}$ ; mientras que las flechas verdes representan el modelo construido donde las observaciones están condicionadas al parámetro y el parámetro está condicionado por la observación anterior.

Ahora bien, utilizando las relaciones entre las variables observables y latentes, el proceso de transición para una observación en el modelo general de probabilidad con procesos de variables latentes se escribe asegurando la estacionariedad desde la construcción,

$$\begin{aligned}
P(c_i, d_i | c_{i-1}, d_{i-1}) &= \int P(c_i, d_i | \gamma, \theta) P(\gamma, \theta | c_{i-1}, d_{i-1}) \\
&= \int \int F(c_i, d_i | \gamma, \theta) \pi(\gamma | d_i, c_{i-1}) \pi(\theta | d_{i-1}) \\
&= \int \int F(c_i | \gamma, d_i) F(d_i | \theta) \pi(\gamma | d_i, c_{i-1}) \pi(\theta | d_{i-1}) \\
&= \int F(c_i | \gamma, d_i) \pi(\gamma | d_i, c_{i-1}) \int F(d_i | \theta) \pi(\theta | d_{i-1}) \\
&= P(c_i | d_i, c_{i-1}) P(d_i | d_{i-1})
\end{aligned}$$

De este modo, el modelo general de la probabilidad es

$$P((d_i, c_i)_{i=1}^k | N(t) = k) = P(c_1 | d_1) P(d_1) \prod_{i=2}^k P(c_i | d_i, c_{i-1}) P(d_i | d_{i-1}) \times P(N(t) = k)$$

Este es el modelo general de probabilidad del proceso puntual marcado de las duraciones y costos de las enfermedades crónico degenerativas. Al introducir las variables latentes se construyen funciones de transición más sólidas, tomando en cuenta los parámetros que influyen en el mismo proceso.

Como se ve en el modelo general de probabilidad, la estructura de de-

pendencia con las variables latentes y como, aunque los valores de estas variables no son observables, son influenciados por los valores de las variables observables. Esta estructura de dependencia, según [Gelman et al., 2014], puede denominarse como un modelo jerárquico con resultados observables condicionados a ciertos parámetros, los cuales, a su vez, están dados como variables aleatorias definidos a su vez con otros parámetros. Es decir,

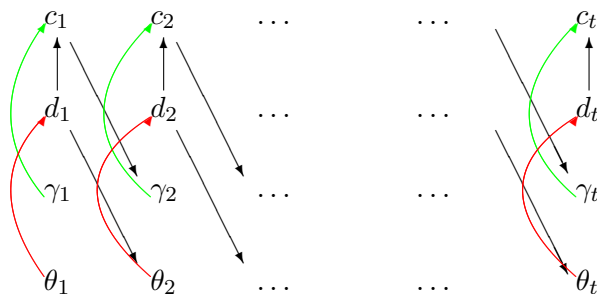


Figura 2.4: Representación gráfica de la estructura jerárquica del modelo general de probabilidad.

Es importante notar que la distribución de las variables latentes es arbitrario, por lo que con este modelo general de probabilidad resta determinar las distribuciones que mejor describan las características de la relación entre las variables latentes. En la siguiente sección se relizará una comparación entre las distintas distribuciones que trabajarían mejor con las características particulares de los datos para su modelado.

## 2.5. Construcción del Modelo de Marcas

Con el modelo general de probabilidad formulado en la sección anterior se sientan las bases para hacer inferencia sobre duración y costos de padecimientos crónico degenerativos. Igualmente, la determinación de las distribuciones de las variables latentes son cruciales para la predicción sobre futuras observaciones.

Según [Fader and Hardie, 2013] la distribución Gamma es muy útil para modelar gasto como en el campo actuarial son los siniestros, pues estos normalmente tienden a estar sesgados a la derecha además de que las propiedades de las convoluciones siguen los patrones de gasto si los siniestros se distribuyen Gamma. Aunado a esto, en una óptica bayesiana, la distribución Gamma tiene muchas distribuciones conjugadas como la distribución Poisson, la distribución exponencial, la distribución normal con media conocida, la distribución Pareto entre otras; que ayudan a la linealidad del modelo general de probabilidad.

Las duraciones se modelan mediante un modelo Gamma-Gamma, mientras que los costos con un modelo Gamma Inversa-Weibull. El modelo Gamma-Gamma de las duraciones se deriva del propio proceso de conteo que es una distribución conjugada.

$$P(d_i|d_{i-1}) = \int P(d_i|\theta_i)P(\theta_i|d_{i-1}) \quad i \neq 1$$

Donde,

$$P(d|\theta) \sim \text{Gamma}(d|\alpha_d, \theta)$$

$$P(\theta) \sim \text{Gamma}(\theta|\alpha_\theta, \beta_\theta)$$

De manera análoga, para los costos el modelo es Gamma Inversa-Weibull, la distribución Weibull es usada frecuentemente en el campo actuarial para modelar los siniestros, esto es porque al ser una distribución relacionada con la distribución exponencial, es útil también en la modelación de valores extremos. Por estas características, la distribución Weibull es la que modelará los costos en este trabajo, así se asegura que si alguno de los costos del modelo llegara a diferir mucho del resto, el modelo tenga la flexibilidad para incluir los datos. Tomando en cuenta que los costos se modelan con la distribución Weibull, por conveniencia, la variable latente se distribuye Gamma Inversa.

$$P(c_i|c_{i-1}) = \int P(c_i|\gamma_i, d_i)P(\gamma_i|d_i, c_{i-1}) \quad i \neq 1$$

Donde,

$$\begin{aligned} P(c|\gamma, d) &\sim Weibull(c|d, \gamma) \\ P(\gamma) &\sim Inv - Gamma(\gamma|\alpha_\gamma, \beta_\gamma) \end{aligned}$$

Con el modelo general de probabilidad y las distribuciones designadas ya es posible hacer inferencia sobre observaciones futuras, como se demostrará en el siguiente capítulo.

## 2.6. Discusión

Los beneficios que presentan los modelos de Procesos Puntuales Marcados con Variables Latentes para la estimación de costos totales de enfermedades crónico-degenerativas son la flexibilidad que brindan para modelar las estructuras de dependencia entre las marcas y el proceso puntual. Es decir, que es un modelo construido de manera específica



para ajustarse a las trayectorias de cada paciente con un padecimiento crónico degenerativo.

La especificidad del modelo aunado a la estacionareidad asegurada por construcción, se puede asegurar la inferencia de observaciones futuras. Aunque fuera del alcance de este proyecto de investigación, la inferencia es una herramienta poderosa para un estudio riguroso de tarificación en el sector asegurador para seguros de gastos médicos mayores. La verosimilitud del modelo y la manera de hacer inferencia se desarrollarán a profundidad en el siguiente capítulo, de acuerdo a las distribuciones designadas.

## Capítulo 3

# Inferencia y Predicción

### 3.1. Introducción

En el capítulo anterior se describió el modelo general de probabilidad que describe el proceso de duraciones y costos en un padecimiento crónico degenerativo, por lo que el siguiente paso sería hacer inferencia sobre el mismo. El objetivo de hacer inferencia es predecir futuras observaciones en base a los datos ya observados. En este capítulo se sentarán las bases para realizar esta inferencia.

El primer paso para realizar inferencia es la construcción de la función de verosimilitud, en este caso extendida a las variables latentes y a los parámetros correspondientes a sus distribuciones. Una vez que se determinaron las funciones de verosimilitud, se analizan los métodos de estimación que se podrían usar para la predicción de futuras observaciones. De este modo se puede completar la parte teórica de este trabajo de investigación.

### 3.2. Verosimilitud Extendida

Como especificado en la sección anterior, una vez que el modelo de probabilidad describe de manera precisa los datos del problema podemos empezar a hacer inferencia sobre observaciones futuras. La base sobre la que se puede hacer inferencia en base a los datos ya observados es la función de verosimilitud de la función de probabilidad.

La función de verosimilitud, según [Held and Sabanés Bové, 2014], se define como

**Definición 5.** *La función de verosimilitud  $V(\theta)$  es la función masa o la función de densidad de los datos observados  $x$ , entendidos en función del parámetro desconocido  $\theta$ .*

En este caso las variables observables se definen en función de las variables latentes, estas a su vez se describen en función de sus parámetros. De esto se desprende la noción de verosimilitud extendida para incluir las variables latentes. Como se especifica en [Pitt et al., 2002], la construcción de la función de verosimilitud resulta sencilla, incluso intuitiva. Sin embargo, la estimación de los parámetros mediante máxima verosimilitud no es tan sencilla pues no tiene una solución que se pueda expresar de manera analítica cerrada. Usando esta construcción, se escribe una función de verosimilitud para el modelo general de probabilidad de duraciones y costos

$$\begin{aligned}
 V(\{\theta_j\}, \{\gamma_j\}, \{d_j\}, \{c_j\}) &= f(d_1|\theta_1)f(c_1|d_1, \gamma_1)f(\gamma_1) \times \\
 &\quad \times \prod_{j=2}^{N(t)} f(d_j|\theta_j)f(\theta_j|d_{j-1})f(c_j|d_j, \gamma_j)f(\gamma_j|c_{j-1})
 \end{aligned}$$

Para poder calcular la función de verosimilitud que permite hacer in-

ferencia, es necesario conocer las distribuciones de las variables latentes en base a las observaciones anteriores para ambas variables observables, duraciones y costos.

Para la primera variable observable se toma en cuenta la relación

$$f_{\theta|d}(\theta|d) \propto f_{d|\theta}(d|\theta)f(\theta)$$

y que  $d|\theta \sim \text{Gamma}(d|\alpha_d, \theta)$  y  $\theta \sim \text{Gamma}(\theta|\alpha_\theta, \beta_\theta)$ .

$$\begin{aligned} f_{\theta|d}(\theta|d) &\propto \frac{\theta^{\alpha_d}}{\Gamma(\alpha_d)} d^{\alpha_d-1} e^{-\{\theta d\}} \times \frac{\beta_\theta^{\alpha_\theta}}{\Gamma(\alpha_\theta)} \theta^{\alpha_\theta-1} e^{-\{\beta_\theta \theta\}} \\ &= \frac{\beta_\theta^{\alpha_\theta}}{\Gamma(\alpha_d)\Gamma(\alpha_\theta)} d^{\alpha_d-1} \theta^{\alpha_d+\alpha_\theta-1} e^{-\{\theta(d+\beta_\theta)\}} \\ &\propto \theta^{\alpha_d+\alpha_\theta-1} e^{-\{\theta(d+\beta_\theta)\}} \\ &\Rightarrow \theta|d \sim \text{Gamma}(\alpha_d + \alpha_\theta, d + \beta_\theta) \end{aligned}$$

Para la variable de duraciones, la distribución de la variable latente que depende de la observación se puede expresar de una manera analítica cerrada como la distribución Gamma. Análogamente, para la variable de costos se vuelve a tomar en cuenta la misma relación y las distribuciones

$$c|d, \gamma \sim \text{Weibull}(c|d, \gamma) \quad \gamma \sim \text{InvGamma}(\gamma|\alpha_\gamma, \beta_\gamma)$$

$$\begin{aligned}
f_{\gamma|d,c}(\gamma|d,c) &\propto \frac{d}{\gamma^d} c^{d-1} e^{\{-(\frac{c}{\gamma})^d\}} \times \frac{\beta_\gamma^{\alpha_\gamma}}{\Gamma(\alpha_\gamma)} \left(\frac{1}{\gamma}\right)^{\alpha_\gamma+1} e^{\{-(\frac{\beta_\gamma}{\gamma})\}} \\
&= \frac{d\beta_\gamma^{\alpha_\gamma} c^{d-1}}{\Gamma(\alpha_\gamma)} \left(\frac{1}{\gamma}\right)^{d+\alpha_\gamma+1} e^{-((\frac{\beta_\gamma}{\gamma})+(\frac{c}{\gamma})^d)} \\
&\propto \left(\frac{1}{\gamma}\right)^{d+\alpha_\gamma+1} e^{-((\frac{\beta_\gamma}{\gamma})+(\frac{c}{\gamma})^d)}
\end{aligned}$$

Para la variable de costos, la distribución de la variable latente según la observación anterior no tiene una forma analítica cerrada como distribución, sin embargo, el kernel se puede simular con un slice sampler; este método se explicará con detalle en la siguiente sección.

Las distribuciones que resultan de los cálculos anteriores son la llave que se necesita para empezar a hacer inferencia, tomando estas distribuciones se redefine la verosimilitud como

$$\begin{aligned}
V(\{\alpha_d, \alpha_\theta, \beta_\theta, \alpha_\gamma, \beta_\gamma\}, \{\theta_i, \gamma_i\}_{i=1}^{N(t)} | \{d_j, c_j\}_{i=1}^{N(t)}) = \\
\prod_{i=1}^{N(t)} \text{Gamma}(d_i | \alpha_d, \theta_i) \text{Gamma}(\theta_i | \alpha_d + \alpha_\theta, d_{i-1} + \beta_\theta) \times \\
\text{Weibull}(c_i | d_i, \gamma_i) \left(\frac{1}{\gamma}\right)^{d_{i-1} + \alpha_\gamma + 1} e^{\{- (\frac{\beta_\gamma}{\gamma} + (\frac{c_{i-1}}{\gamma})^{d_i})\}} \times \\
\times \pi(\alpha_d) \pi(\alpha_\theta) \pi(\beta_\theta) \pi(\alpha_\gamma) \pi(\beta_\gamma)
\end{aligned}$$

Donde para las variables latentes,

$$\begin{aligned}
\pi(\theta_i|\alpha_d, \alpha_\theta, \beta_\theta) &\propto \text{Gamma}(d_i|\alpha_d, \theta_i) \times \text{Gamma}(\theta_i|\alpha_d + \alpha_\theta, d_{i-1} + \beta_\theta) \\
&= \frac{\theta_i^{\alpha_d}}{\Gamma(\alpha_d)} d_i^{\alpha_d-1} e^{-\theta_i d_i} \frac{(d_{i-1} + \beta_\theta)^{\alpha_d + \alpha_\theta}}{\Gamma(\alpha_d + \alpha_\theta)} \theta_i^{\alpha_d + \alpha_\theta} e^{\{d_{i-1} + \beta_\theta\}} \\
&= \frac{d_i^{\alpha_d-1} (d_{i-1} + \beta_\theta)^{\alpha_d + \alpha_\theta}}{\Gamma(\alpha_d) \Gamma(\alpha_d + \alpha_\theta)} \theta_i^{2\alpha_d + \alpha_\theta - 1} e^{\{-\theta_i(d_{i-1} + d_i + \beta_\theta)\}} \\
&\propto \theta_i^{2\alpha_d + \alpha_\theta - 1} e^{\{-\theta_i(d_{i-1} + d_i + \beta_\theta)\}} \\
&\Rightarrow \theta_i \sim \text{Gamma}(2\alpha_d + \alpha_\theta, d_i + d_{i-1} + \beta_\theta)
\end{aligned}$$

Y,

$$\begin{aligned}
\pi(\gamma_i|\alpha_\gamma, \beta_\gamma, d_i, c_{i-1}) &\propto \text{Weibull}(c_i|d_i, \gamma_i) \times \left(\frac{1}{\gamma_i}\right)^{d_i + \alpha_\gamma + 1} e^{\{-(\frac{\beta_\gamma}{\gamma_i} + (\frac{c_{i-1}}{\gamma_i})^{d_i})\}} \\
&= \frac{d_i}{\gamma_i^{d_i}} c_i^{d_i-1} e^{\{-(\frac{c_i}{\gamma_i})\}} \left(\frac{1}{\gamma_i}\right)^{d_i + \alpha_\gamma + 1} e^{\{-(\frac{\beta_\gamma}{\gamma_i} + (\frac{c_{i-1}}{\gamma_i})^{d_i})\}} \\
&= d_i c_i^{d_i-1} \left(\frac{1}{\gamma_i}\right)^{2d_i + \alpha_\gamma + 1} e^{\{-(\frac{\beta_\gamma}{\gamma_i} + (\frac{c_{i-1}}{\gamma_i})^{d_i} + (\frac{c_i}{\gamma_i})^{d_i})\}} \\
&\propto \left(\frac{1}{\gamma_i}\right)^{2d_i + \alpha_\gamma + 1} e^{\{-(\frac{\beta_\gamma}{\gamma_i} + (\frac{c_{i-1} + c_i}{\gamma_i})^{d_i})\}}
\end{aligned}$$

Las distribuciones que corresponden a los parámetros son,

$$\begin{aligned}
\pi(\alpha_d|\dots) &\propto \prod_{i=2}^{N(t)} \text{Gamma}(\theta_i|\alpha_d + \alpha_\theta, d_{i-1} + \beta_\theta) \times \text{Gamma}(\alpha_d|\alpha_0, \beta_0) \\
&= \prod_{i=2}^{N(t)} \frac{(d_{i-1} + \beta_\theta)^{\alpha_d + \alpha_\theta}}{\Gamma(\alpha_d + \alpha_\theta)} \theta_i^{\alpha_d + \alpha_\theta - 1} e^{\{-\theta_i(d_{i-1} + \beta_\theta)\}} \frac{\beta_0^{\alpha_0}}{\Gamma(\alpha_0)} \alpha_d^{\alpha_0 - 1} e^{\{-\alpha_d \beta_0\}} \\
&\propto \prod_{i=2}^{N(t)} \frac{(d_{i-1} + \beta_\theta)^{\alpha_d + \alpha_\theta}}{\Gamma(\alpha_d + \alpha_\theta)} \theta_i^{\alpha_d} \alpha_d^{\alpha_0 - 1} e^{\{-\alpha_d \beta_0\}}
\end{aligned}$$

$$\begin{aligned}
\pi(\alpha_\theta|\dots) &\propto \prod_{i=2}^{N(t)} \text{Gamma}(\theta_i|\alpha_d + \alpha_\theta, d_{i-1} + \beta_\theta) \times \text{Gamma}(\alpha_\theta|\alpha_0, \beta_0) \\
&= \prod_{i=2}^{N(t)} \frac{(d_{i-1} + \beta_\theta)^{\alpha_d + \alpha_\theta}}{\Gamma(\alpha_d + \alpha_\theta)} \theta_i^{\alpha_d + \alpha_\theta - 1} e^{\{-\theta_i(d_{i-1} + \beta_\theta)\}} \frac{\beta_0^{\alpha_0}}{\Gamma(\alpha_0)} \alpha_\theta^{\alpha_0 - 1} e^{\{-\alpha_\theta \beta_0\}} \\
&\propto \prod_{i=2}^{N(t)} \frac{(d_{i-1} + \beta_\theta)^{\alpha_d + \alpha_\theta}}{\Gamma(\alpha_d + \alpha_\theta)} \theta_i^{\alpha_\theta} \alpha_\theta^{\alpha_0 - 1} e^{\{-\alpha_\theta \beta_0\}}
\end{aligned}$$

$$\begin{aligned}
\pi(\beta_\theta|\dots) &\propto \prod_{i=2}^{N(t)} \text{Gamma}(\theta_i|\alpha_d + \alpha_\theta, d_{i-1} + \beta_\theta) \times \text{Gamma}(\beta_\theta|\alpha_0, \beta_0) \\
&= \prod_{i=2}^{N(t)} \frac{(d_{i-1} + \beta_\theta)^{\alpha_d + \alpha_\theta}}{\Gamma(\alpha_d + \alpha_\theta)} \theta_i^{\alpha_d + \alpha_\theta - 1} e^{\{-\theta_i(d_{i-1} + \beta_\theta)\}} \frac{\beta_0^{\alpha_0}}{\Gamma(\alpha_0)} \beta_\theta^{\alpha_0 - 1} e^{\{-\beta_\theta \beta_0\}} \\
&\propto \prod_{i=2}^{N(t)} (d_{i-1} + \beta_\theta)^{\alpha_d + \alpha_\theta} \beta_\theta^{\alpha_0 - 1} e^{\{-\beta_\theta(\theta_i + \beta_0)\}}
\end{aligned}$$

$$\begin{aligned}
\pi(\alpha_\gamma|\dots) &\propto \prod_{i=2}^{N(t)} \pi(\gamma_i|\alpha_\gamma, \beta_\gamma, d_i, c_{i-1}) \times \text{Gamma}(\alpha_\gamma|\alpha_0, \beta_0) \\
&= \prod_{i=2}^{N(t)} \left(\frac{1}{\gamma_i}\right)^{d_i + \alpha_\gamma + 1} e^{\{-(\frac{\beta_\gamma}{\gamma_i} + (\frac{c_{i-1}}{\gamma_i})^{d_i})\}} \frac{\beta_0^{\alpha_0}}{\Gamma(\alpha_0)} \alpha_\gamma^{\alpha_0 - 1} e^{\{-\alpha_\gamma \beta_0\}} \\
&\propto \prod_{i=2}^{N(t)} \left(\frac{1}{\gamma_i}\right)^{\alpha_\gamma} \alpha_\gamma^{\alpha_0 - 1} e^{\{-\alpha_\gamma \beta_0\}}
\end{aligned}$$



$$\begin{aligned}
\pi(\beta_\gamma|\dots) &\propto \prod_{i=2}^{N(t)} \pi(\gamma_i|\alpha_\gamma, \beta_\gamma, d_i, c_{i-1}) \times \text{Gamma}(\beta_\gamma|\alpha_0, \beta_0) \\
&= \prod_{i=2}^{N(t)} \left(\frac{1}{\gamma_i}\right)^{d_i+\alpha_\gamma+1} e^{\{-(\frac{\beta_\gamma}{\gamma_i}+(\frac{c_{i-1}}{\gamma_i})^{d_i})\}} \frac{\beta_0^{\alpha_0}}{\Gamma(\alpha_0)} \beta_\gamma^{\alpha_0-1} e^{\{-\beta_\gamma\beta_0\}} \\
&\propto \beta_\gamma^{\alpha_0-1} e^{\{-\beta_\gamma(\frac{1}{\gamma_i}+\beta_0)\}} \\
&\Rightarrow \beta_\gamma \sim \text{Gamma}(\alpha_0, \frac{1}{\gamma_i} + \beta_0)
\end{aligned}$$

Los parámetros a estimar son aquellos correspondientes a las variables latentes, cuyo kernel se estima para cada momento  $i$  ( $i = 1, \dots, N(t)$ ), y a las variables observables del modelo. Dado el uso de las variables latentes no existirá un óptimo global para las variables, razón por la cual se requiere el uso de métodos numéricos no solo para la estimación de los parámetros y variables latentes sino para analizar las distribuciones que no tienen expresión analítica cerrada, como aquellas de los parámetros  $\alpha_d, \alpha_\theta, \alpha_\gamma$  y la variable latente  $\gamma_i$ .

Es importante mencionar que estas distribuciones que no tienen forma analítica cerrada son funciones log-cóncavas. Las funciones log-cóncavas son, de acuerdo con [Bagnoli and Bergstrom, 2005], funciones que se grafican con una curva cóncava en los números reales positivos y cuyo logaritmo es también una función cóncava. Estas funciones tienen el siguiente teorema

**Teorema 6.** *Sea  $F$  sea una función doblemente diferenciable que toma valores positivos con soporte en  $(a, b)$  y sea  $t$  una función doblemente*

diferenciable y monotónica que va de  $(a', b')$  a  $(a, b) = (t(a'), t(b'))$ . Se define la función  $\hat{F}$  con soporte en  $(a', b')$  para toda  $x \in (a', b')$ ,  $\hat{F}(x) = F(t(x))$ . Si  $F$  es log-cóncava y  $t$  una función cóncava, entonces  $\hat{F}$  es log-cóncava.

Este teorema tiene como corolario,

**Corolario 7.** Sea  $F$  una función con soporte en  $(a, b)$ . Sea  $t$  una transformación lineal de la línea real hacia si misma, y se define una función  $\hat{F}$  con soporte en  $(t(a), t(b))$  tal que  $\hat{F}(x) = F(t(x))$ . Si  $F$  es log-cóncava entonces  $\hat{F}$  es log-cóncava.

Según [Bagnoli and Bergstrom, 2005], la distribución Weibull cumple con las características de log-concavidad si su parámetro de forma es mayor o igual a uno, para el caso de la distribución de los costos este parámetro corresponde a la duración, que por definición es mayor a uno. También la distribución Gamma de  $\alpha_d, \alpha_\theta, \alpha_\gamma$  debe tener el parámetro de forma mayor a uno, como sucede para la distribución Weibull, esta condición se cumple por construcción. De este modo aseguramos la log-concavidad para estas distribuciones que no tienen una fórmula analítica cerrada y que se muestrearan de manera previa a la estimación de parámetros para el modelo general de probabilidad.

[Pitt et al., 2002] especifica que la estimación de máxima verosimilitud puede resolverse mediante el algoritmo EM, aunque también, debido a que las densidades son dos condicionales de la densidad conjunta puede ligarse con el Muestreador Gibbs.

El algoritmo EM es un algoritmo para calcular el estimador de máxima verosimilitud, que de acuerdo con [Held and Sabanés Bové, 2014], se define como

**Definición 8.** *El Estimador de Máxima Verosimilitud (EMV)  $\hat{\theta}_{MV}$  del parámetro  $\theta$  se obtiene maximizando la función de verosimilitud.*

$$\hat{\theta}_{MV} = \max_{\theta \in \Theta} L(\theta)$$

Según [Dempster et al., 1977] el algoritmo EM calcula el EML mediante iteraciones, cada iteración consiste en un paso dónde se calcula la esperanza y en otro se maximiza la misma, de ahí el nombre de EM. Este algoritmo se relaciona con las variables latentes suponiendo dos variables  $x$  y  $y$  las cuales se relacionan  $x \rightarrow y(x)$ , donde  $y$  son los datos observables.

De este modo, análogamente a lo expresado en el capítulo anterior por [Pitt et al., 2002], se proponen las siguientes funciones de densidad  $f(x|\phi)$  y  $g(y|\phi)$ ; en las cuales, de acuerdo a [Dempster et al., 1977] los datos completos (variables latentes)  $f(x|\cdot)$  se relacionan con los datos incompletos (variables observadas)  $g(y|\cdot)$  mediante

$$g(y|\phi) = \int_{x(y)} f(x|\phi) dx$$

El algoritmo EM se dedica a encontrar un valor de  $\phi$  que maximice  $g(y|\phi)$  dada la  $y$  observada usando la familia asociada de  $f(x|\phi)$ . Una de las caracterizaciones más simples supone  $\phi^{(p)}$  es el valor actual de  $\phi$  después de  $p$  iteraciones y  $t(x)$  como el estadístico suficientes de los datos completos, es decir, el estimador de la variable latente; por lo que la siguiente iteración se puede desglosar en los siguientes dos pasos:

- Paso E: Estimar los estadísticos suficientes de los datos completos.

$$t^{(p)} = E[t(x)|y, \phi^{(p)}]$$

- Paso M: Determinar  $\phi^{(p+1)}$  como solución a la ecuación

$$E[t(x)|\phi] = t^{(p)}$$

Es decir, que si suponemos que  $t^{(p)}$  es el estadístico suficiente calculado de  $x$  observada en la distribución  $f(x|\phi)$  entonces la ecuación definida en el Paso M se define como el EMV. Este concepto se hace general al definir la siguiente función

$$Q(\phi'|\phi) = E[\log f(x|\phi')|y, \phi]$$

Esta función se asume que existe para toda pareja  $(\phi', \phi)$ . Se define la iteración EM para  $\phi^{(p)} \rightarrow \phi^{(p+1)}$ ,

- Calcular  $Q(\phi, \phi^{(p)})$ .
- Determinar  $\phi^{(p+1)}$  tal que maximice  $Q(\phi, \phi^{(p)})$ .

La idea central es tomar una  $\phi'$  que maximice  $\log f(x|\phi)$ , dado que esta distribución y su correspondiente logaritmo no necesariamente se conoce, se puede maximizar los datos observados y  $\phi^{(p)}$ .

El algoritmo EM es muy útil pues por su estructura iterativa puede dar resultados a modelos de probabilidad muy complejos, además de que al igual que el Muestreador de Gibbs, utiliza una estructura subyacente o de variables latentes. En la siguiente sección se explorará con detalle el Muestreador de Gibbs y sus implicaciones con el modelo general de probabilidad relativo a este trabajo.

### 3.3. Slice Sampler

En la función de verosimilitud definida en la sección anterior, las distribuciones de la variable latente  $\gamma_i$  y de los parámetros  $\alpha_d, \alpha_\theta, \alpha_\gamma$  no se pueden expresar de una forma analítica cerrada. Es por esto que antes de empezar a estimar la función de verosimilitud es necesario muestrear primero estas distribuciones para lograr que los métodos de estimación funcionen de la mejor manera. Este primer acercamiento se hace mediante el Slice Sampler.

De acuerdo con [Neal, 2003] existen varias aplicaciones para esta manera de muestrear, desde la manera más simple con una distribución univariada como alternativa al Muestreador de Gibbs hasta la más compleja donde se pueden adaptar relaciones de dependencia entre las variables, permitiendo mayor flexibilidad al Muestreador de Gibbs. Esta es la clase de Slice Sampler que se utilizará en este trabajo de investigación.

La idea detrás del concepto del Slice Sampler es que si queremos muestrear la distribución de una variable  $x$  que toma valores en el conjunto  $\mathbb{R}^n$ , cuya densidad sea proporcional a alguna función  $f(x)$ , entonces se muestrea de forma uniforme en la región por debajo de la curva de  $f(x)$  con dimensión  $(n + 1)$ . Esta idea se puede concretar introduciendo una variable real auxiliar  $y$  y definiendo la distribución conjunta de  $x$  y  $y$  que es uniforme sobre la región debajo de la curva de  $f(x)$ , tal que,  $U = \{(x, y) : 0 < y < f(x)\}$ . Es decir, que para muestrear  $x$ , muestrearemos conjuntamente  $(x, y)$  para después ignorar  $y$ .

Sin embargo, dado que generar puntos independientes muestreados de la distribución  $U$  no siempre es sencillo, se puede definir una Cadena de Markov que converja a esta distribución uniforme. Esto se hace median-

te el muestreo alternado de la distribución condicional de  $y|x$  la cual es uniforme en el intervalo  $(0, f(x))$  y de la distribución condicional de  $x|y$  la cual también es uniforme sobre la región  $S = \{x_y < f(x)\}$ , la cual es nombrada como la *rebanada* definida por  $y$ . El procedimiento para construir la Cadena de Markov para una distribución univariada  $f(x)$ , tomando el valor inicial  $x_0$ , de acuerdo con [Neal, 2003], es:

- Se extrae un valor real  $y$  uniforme en  $(0, f(x))$ , definiendo una *rebanada* horizontal  $S = \{x : y < f(x)\}$ . Es importante notar que  $x_0$  está siempre dentro de  $S$ .
- Se encuentra un intervalo  $I = (L, R)$  en la vecindad de  $x_0$  que contenga toda, o la mayor parte, de la *rebanada*.
- Se extrae un nuevo punto,  $x_1$ , de la parte de la *rebanada* que está dentro del intervalo  $I$ .

En el primer paso del procedimiento se elige la variable auxiliar que es característica al Slice Sampler, pues este valor no es necesario de una iteración de la Cadena de Markov a la siguiente; mientras que los siguientes dos pasos pueden ser implementados de muchas maneras en tanto que la Cadena de Markov resultante permita que la distribución definida  $f(x)$  permanezca invariante. El siguiente problema que se enfrenta es delimitar el intervalo, pues necesita ser lo suficientemente grande para que el nuevo punto ( $x_1$ ) esté lo más lejos del anterior ( $x_0$ ) dentro de la misma *rebanada*, pero el intervalo tampoco puede salirse de la misma pues eso volvería ineficiente al muestreo.

Para garantizar convergencia en la distribución  $f(x)$  y su invarianza, la Cadena de Markov debe ser ergódica. Según [Neal, 2003], para demostrar que la función de distribución permanece invariante, suponemos que el estado inicial  $x_0$  se distribuye  $f(x)$ , en el primer paso del procedimiento

la distribución conjunta es con las variables  $x_0$  y  $y$ , por lo que al actualizar  $x_0$  a  $x_1$  la función de distribución conjunta se mantiene invariante, ignorando así la variable  $y$ . La distribución de  $x_1$  es la distribución marginal de la conjunta, es decir, la función  $f(x)$  definida. De este modo, lo único que se necesita demostrar que la selección de  $x_0$  y  $x_1$  en los siguientes dos pasos del procedimiento deja a la distribución conjunta de  $x$  y  $y$  invariante, y con la distribución condicional sobre  $S = \{x : y < f(x)\}$ , es decir, la *rebanada* definida por  $y$ . Esta invarianza se puede demostrar si la probabilidad de que  $x_1$  sea el próximo estado dado que el estado actual es  $x_0$  es igual a la probabilidad de que  $x_0$  sea el próximo estado dado que el estado actual es  $x_1$ , para cualquier  $x_0$  y  $x_1$  en  $S$ .

Este es el procedimiento que aplica para cuando la distribución es univariada, sin embargo, si esta llegara a ser una distribución multivariada ( $x = (x_1, \dots, x_n)$ ) existen dos caminos para hacer el Slice Sampler: el primero es muestrear para cada variable por separado; para esto es necesario poder calcular la función  $f_i(x_i)$  y que esta sea proporcional a  $p(x_i | \{x_j\}_{j \neq i})$ , donde  $\{x_j\}_{j \neq i}$  son los valores de las variables. El otro camino es seguir el mismo procedimiento que se definió para las distribuciones univariadas, solamente que en el segundo paso se reemplaza el intervalo con un hiperrectángulo  $H = \{x : L_i < x_i < R_i, \quad i = 1, \dots, n\}$  donde  $L_i$  y  $R_i$  definen la extensión del hiperrectángulo a lo largo del eje para la variable  $x_i$ .

En el caso de las distribuciones de los parámetros  $\alpha_d, \alpha_\theta, \alpha_\gamma$  y de la variable latente  $\gamma_i$  son distribuciones con una sola variable, por lo que el Slice Sampler a utilizar para estimar es el de una sola variable. Ahora, una vez estimado estas distribuciones es necesario utilizar el Muestreador de Gibbs que se desarrolla en la siguiente sección para la estimación del modelo general de probabilidad.

### 3.4. El Muestreador de Gibbs

Como mencionado en la sección anterior, la estimación de parámetros del modelo general de probabilidad que se utilizará en este trabajo no se puede hacer a través de métodos analíticos tradicionales, por lo que se utilizarán métodos numéricos. Entre estos se encuentran el algoritmo EM y el Muestreador de Gibbs, el algoritmo EM fue descrito en la sección anterior y aunque útil para el análisis del modelo presentado, el Muestreador de Gibbs tiene una interpretación más simple.

De acuerdo con [Gelman et al., 2014], la simulación a través de las cadenas de Markov también llamadas Cadenas de Markov vía simulación Monte Carlo (MCMC, por sus siglas en inglés) es un método basado en las realizaciones de los valores de  $\theta$  de distribuciones aproximadas y después corrigiendo esas realizaciones para poder mejorar la distribución posterior de  $p(\theta|y)$ . Estas realizaciones se realizan de manera iterativa. Una de estas simulaciones MCMC es el Muestreador de Gibbs.

Normalmente el Muestreador Gibbs se asocia con la estadística bayesiana aunque puede ser útil también en la visión clásica de la estadística, según [Casella and George, 1992] este algoritmo es una técnica que genera variables aleatorias indirectamente de distribuciones marginales sin tener que calcular la densidad, debido a que se basa en las propiedades principales de las Cadenas de Markov como la estacionariedad para simplificar cálculos y tener estimados más precisos.

Siguiendo la ilustración de [Casella and George, 1992], supongamos que tenemos una distribución conjunta  $f(\theta, y_1, y_2, \dots, y_p)$

$$f(\theta) = \int \cdots \int f(\theta, y_1, y_2, \dots, y_p) dy_1, dy_2, \dots, dy_p$$



Si el interés se encuentra en la marginal  $f(\theta)$  y ésta es demasiado complicada para calcularse directamente, con el Muestreador de Gibb se puede generar una muestra  $\theta_1, \dots, \theta_m \sim f(\theta)$  sin la necesidad de calcular la distribución marginal. Esto permite obtener información de la misma con alto grado de precisión.

Para ejemplificar mejor el mecanismo del Muestreador de Gibbs se toman dos variables aleatorias  $(\Theta, Y)$ . El algoritmo genera una muestra de  $f(\theta)$  muestreando de las distribuciones condicionales  $f(\theta|y)$  y  $f(y|\theta)$  que son la que normalmente se conocen en los modelos estadísticos. Esta muestra se obtiene mediante, lo que [Casella and George, 1992] nombra como, una secuencia de Gibbs  $(Y'_0, \theta'_0, Y'_1, \theta'_1, \dots, Y'_k, \theta'_k)$  que de manera iterativa genera variables aleatorias a partir de valores iniciales especificados ( $Y'_0 = y'_0$ ).

El proceso iterativo es como sigue

$$\begin{aligned}\theta'_j &\sim f(\theta|Y'_j = y'_j) \\ Y'_{j+1} &\sim f(y|\theta'_j = x'_j)\end{aligned}$$

Si la muestra es suficientemente grande, es decir, que si  $k \rightarrow \infty$  la distribución de  $\theta'_k$  convergerá con la verdadera distribución marginal de  $\theta$ .

El Muestreador de Gibbs puede pensarse como una implementación práctica del concepto de que solo conociendo las distribuciones marginales se puede determinar la distribución conjunta. Esto sería cierto en la mayoría de los casos bivariados, el procedimiento no es tan directo para los casos multivariados.

De acuerdo con [Casella and George, 1992] para el caso bivariado, suponemos dos variables aleatorias  $\theta, Y$ , de las cuales se conocen sus distribuciones condicionales  $f_{\theta|Y}(\theta|y)$  y  $f_{Y|\theta}(y|\theta)$ . A partir de estas podríamos calcular la función marginal de  $\theta$  y la distribución conjunta de ambas variables, mediante el siguiente argumento:

$$f_{\theta}(\theta) = \int f_{\theta Y}(\theta, y) dy$$

donde la distribución conjunta es aún desconocida, tomando el hecho que  $f_{\theta Y}(\theta, y) = f_{\theta|Y}(\theta|y)f_Y(y)$  tendríamos que,

$$f_{\theta}(\theta) = \int f_{\theta|Y}(\theta|y)f_Y(y) dy$$

Asimismo, si sustituimos la distribución marginal de  $y$  ( $f_Y(y)$ ) con el mismo argumento utilizado para la distribución marginal de  $\theta$ , se tiene que

$$\begin{aligned} f_{\theta}(\theta) &= \int f_{\theta|Y}(\theta|y)f_{Y|\theta}(y|t)f_{\theta}(t) dt dy \\ &= \int \left[ \int f_{\theta|Y}(\theta|y)f_{Y|\theta}(y|t) dy \right] f_{\theta}(t) dt \end{aligned}$$

Esta ecuación es una forma limitada de la iteración de Gibbs, ilustrando como las distribuciones condicionales producen una distribución marginal. Aunque la distribución conjunta de las variables determinan las distribuciones condicionales y marginales, no siempre las condicionales determinen de manera tan directa la distribución marginal. Esto es cierto no solo para los casos bivariados, sino que se extiende a los multivariados.

En cuantas más variables existan, el problema se vuelve más complejo pues la relación entre las condicionales, marginales y conjuntas se vuelve más intrincada. Por ejemplo, la relación *condicional*  $\times$  *marginal* =

*conjunta* no se sostiene para todas las condicionales y marginales. Pero se pueden hacer varios conjuntos de variables y construir las ecuaciones integrales para calcular la distribución marginal de interés.

Para casos multivariados [Casella and George, 1992] supone las variables aleatorias  $X, Y, Z$  con interés en la distribución  $f_X(x)$ . Para esto, se toman las variables  $(Y, Z)$  como una sola variable, lo que resultaría en

$$f_X(x) = \int \left[ \int \int f_{X|YZ}(x|y, z) f_{Y|X}(y, z|x) dy dz \right] f_X(x) dx$$

De esta manera, muestreando iterativamente de  $f_{X|YZ}$  y  $f_{Y|X}$  resultarían en una serie de variables aleatorias que convergen en  $f_X(x)$ . Por otro lado, el Muestreador de Gibb muestrearía iterativamente las distribuciones  $f_{X|YZ}, f_{Y|XZ}, f_{Z|X}$ , de tal modo que en la  $j$ -ésima iteración se tendría,

$$\begin{aligned} X'_j &\sim f(x|Y'_j = y'_j, Z'_j = z'_j) \\ Y'_{j+1} &\sim f(y|X'_j = x'_j, Z'_j = z'_j) \\ Z'_{j+1} &\sim f(z|X'_j = x'_j, Y'_{j+1} = y'_{j+1}) \end{aligned}$$

Este esquema de iteraciones nos produce una secuencia de Gibbs,

$$Y'_0, Z'_0, X'_0, Y'_1, Z'_1, X'_1, \dots$$

con la misma propiedad de convergencia que en el caso bivariado, ente más grande es la  $k$ ,  $X'_k = x'_k$  es un punto de la distribución marginal  $f(x)$ .

De este modo queda evidenciada la utilidad del Muestreador de Gibbs en el ahorro de cálculos y la precisión de sus resultados. Como mencio-

nado en la sección anterior, esta técnica inferencial es muy útil tanto en la estadística bayesiana como en la clásica, en la primera para calcular la distribución posterior y en la última, para calcular la función de verosimilitud. Según [Gelman et al., 2014], la clave del éxito de este método es la iteración en la cual las distribuciones aproximadas mejoran hasta converger en la distribución deseada.

Tomando la verosimilitud extendida definida anteriormente, primero se fijan los valores iniciales para los parámetros  $\alpha_d^{(0)}, \alpha_\theta^{(0)}, \beta_\theta^{(0)}, \alpha_\gamma^{(0)}, \beta_\gamma^{(0)}$  y para las variables latentes  $\{\theta_i^{(0)}\}_{i=1}^{N(t)}, \{\gamma_i^{(0)}\}_{i=1}^{N(t)}$  y para cada  $k = 1, \dots, N(t)$  tenemos la siguiente distribución que es proporcional a la verosimilitud,

$$\pi(\alpha_d^{(k)}, \alpha_\theta^{(k)}, \beta_\theta^{(k)}, \alpha_\gamma^{(k)}, \beta_\gamma^{(k)} | \{\theta_i^{(k-1)}\}_{i=1}^{N(t)}, \{\gamma_i^{(k-1)}\}_{i=1}^{N(t)}, \{d_i, c_i\}_{i=1}^{N(t)})$$

Este es el principio necesario para utilizar el Muestreador de Gibbs de modo que se estimen los parámetros en base a las variables latentes que a su vez se estiman en base a las observaciones para que con los parámetros estimados se estimen las variables latentes que ayuden a predecir futuras observaciones. Una vez que se tiene el concepto general de la aplicación de esta técnica de muestreo para el objetivo de este trabajo, es necesario considerar las maneras de implementarlo.

Para la implementación del Muestreador de Gibbs en este trabajo se utilizará el programa JAGS (Solo Otro Muestreador de Gibbs, por sus siglas en inglés). Este programa es una extensión del programa BUGS (Inferencia Bayesiana Usando el Muestreador de Gibbs, por sus siglas en inglés), que normalmente es utilizado para la exploración de modelos Markov multivariados en el contexto de estadística actuarial. JAGS es un programa que automáticamente construye las cadenas de Markov vía simulación Monte Carlo (MCMC) para modelos multivariados.

Para lograr expresar estos problemas con BUGS, explicado por sus creadores [Plummer et al., 2003], se necesitan tomar en cuenta dos distribuciones:

- Una distribución que describa la probabilidad de estar en el estado  $j$  en el tiempo  $t$  dado que el sujeto estaba en el estado  $i$  al tiempo 0.
- Una distribución de supervivencia que describa el tiempo en un estado absorbente, que bien, es conocida o censurada a la derecha.

La necesidad de estas distribuciones aunado a la necesidad de una herramienta que ayude a explorar modelos gráficos es lo que da origen a JAGS. Es decir, el programa toma la descripción del modelo general de probabilidad multivariado y regresa un muestreo de MCMC de la distribución posterior.

De este modo, se establece no solo las bases para la inferencia y predicción de futuras observaciones en base a la verosimilitud extendida y la resolución de su función sino también la implementación numérica de la misma. Una vez que se definieron estas herramientas, lo que resta es la adaptación del modelo a la ilustración con los datos, esto se explorará en el siguiente capítulo.

## Capítulo 4

# Ilustración del Modelo con Datos

### 4.1. Introducción

En los capítulos anteriores se define el modelo general de probabilidad y cómo este puede hacer inferencia sobre observaciones. El modelo está definido para describir padecimientos crónico degenerativas, sus duraciones y los costos asociados a las mismas. En este capítulo, este modelo será aplicado al padecimiento específico de diabetes mellitus tipo II con datos de...

### 4.2. Diabetes Mellitus en el mundo: Costos y Prevalencia

En el panorama de salud mundial actual, las enfermedades crónicas han tomado un papel preponderante al ser padecimientos que no tienen cura, que se pueden prolongar de manera indefinida y requieren moni-

toreo y atención constante. Uno de los principales problemas de estos padecimientos consiste en la gran carga económica que estos presentan, junto con sus complicaciones; tanto a pacientes como a los proveedores de servicios de salud. Además de la decreciente calidad de vida que experimenta el paciente, aunado a la cantidad de procedimientos médicos que va necesitando y que muchas veces, su proveedor de servicios de salud ya no es capaz de suministrar.

Uno de los padecimientos con más prevalencia y que es más propenso a complicaciones es la diabetes mellitus. La insulina es la hormona que regula el azúcar en la sangre, entonces la diabetes mellitus es un padecimiento en el cual el páncreas no produce o produce poca insulina, o bien, las células del cuerpo no responden de manera normal a la insulina que se produce. Un elevado nivel de glucosa en la sangre puede, a largo plazo, derivar en una serie de complicaciones oftalmológicas, renales, cardíacas y de circulación.

La [OrganizacionMundialdeSalud, ] (OMS) reconoce dos tipos de diabetes: Tipo I y Tipo II. La diabetes mellitus Tipo I (DM-TI) consiste en la falta de producción de insulina en el cuerpo, normalmente se diagnostica en edades tempranas y sus causas aún son desconocidas, por lo que no existe ninguna clase de tratamiento preventivo. La diabetes mellitus Tipo II (DM-TII) se debe a que el cuerpo no puede procesar correctamente la insulina, esto debido la combinación de sobrepeso y falta de actividad física; antes este padecimiento era exclusivo de la edad adulta, sin embargo, ahora se observa también en niños. La mayoría de enfermos son de tipo dos según el Atlas de la Diabetes publicado por [Federación Internacional de la Diabetes, 2015], donde se reporta que entre el 87 % y el 91 % de los enfermos de diabetes tiene DM-TII; esto provoca que no solo sea mayor la carga económica asociada a ellos, sino que además

son más propensas a desarrollar complicaciones, aunados a la falta de autocuidado del paciente.

Según la OMS, la prevalencia mundial de esta enfermedad casi se duplicó en las personas mayores de 18 años, de 4.7 % en 1980 a 8.5 % en 2014; además de provocar más de 2.2 millones de muertes tan solo en el año 2016 y ser una de las mayores causas de ceguera, fallas renales y amputación de extremidades. En las estadísticas reportadas por la OMS, la diabetes es la cuarta enfermedad que causa más muertes entre las no transmisibles mediante algún agente infeccioso, con un 6.01 % de la mortalidad.

De manera particular, se puede ver la prevalencia de la diabetes en los distintos países. En el Reino Unido la prevalencia en el año 2010/2011 fue de 3,818,545 personas de las cuales el 89.6 % se refieren a un diagnóstico de DM-TII; y según el estudio realizado por [Hex et al., 2012] en el año 2035/2036 esta prevalencia aumentará a 6,289,925 mientras que las proporciones entre tipos de diabetes se mantienen. De acuerdo a datos de la OMS, de los 172 países en los que se analizan las estadísticas de mortalidad debido a diabetes mellitus, el Reino Unido ocupa el lugar 167 con 4.2 muertes por 100,000 personas. A pesar de que sus índices de mortalidad son relativamente bajos, comparados con el resto de los países, el problema con la creciente prevalencia de la enfermedad radica también en las complicaciones asociadas a la misma.

De acuerdo con [Bolaños et al., 2010] para el año 2030 habrán alrededor de 366 millones de pacientes diagnosticados con diabetes en el mundo, de los cuales el 70 % se encontrarán en países de ingresos de medios a bajos, entre ellos los países de Latinoamérica. En específico en México se observan resultados preocupantes, en la Encuesta Nacional de Salud



y Nutrición (ENSANUT) ([Gutiérrez et al., ]) de 2012 la prevalencia de diabetes se reporta de 9.2 %, lo cual es un aumento considerable con lo reportado en la ENSANUT 2006 (7 %) y la ENSANUT 2000 (5.8 %). Viendo estos resultados, el aceleramiento en las tasas de obesidad y los padecimientos asociados el Instituto Nacional de Salud Pública decidió hacer otra encuesta intermedia a la que denominó ENSANUT de Medio Camino, la cual se realizó en el año 2016 con un resultado en la prevalencia de 9.4 %, el cual confirma la tendencia creciente de la misma. En las estadísticas de mortalidad de la OMS, México ocupa el décimo lugar con 90.5 muertes derivadas del diagnóstico de diabetes mellitus por 100,000 personas y las estadísticas nacionales muestran que la diabetes mellitus es la tercera causa de mortalidad. En nuestro país, este padecimiento toma un cariz más apremiante debido a la cantidad de gente diagnosticada y el porcentaje de mortalidad asociado.

A pesar que se preveé que la mayoría de los enfermos de diabetes se encuentren en países en vías de desarrollo, también hay que tomar en cuenta que dado los factores de riesgo que facilitan el desarrollo de DM-TII, la más común; son un estilo de vida sedentario y obesidad. Estos factores normalmente se pueden encontrar en países con economías desarrolladas, tales como EE.UU. Tomando como referencia el reporte elaborado en 2014 por el Centro para el Control y Prevención de Enfermedades ([CDC, ], por sus siglas en inglés) la prevalencia de diabetes es de 9.3 % de la población, es decir, 29.1 millones de personas; aunque el 27.8 % de estas personas aún no estén diagnosticadas. Por otro lado, el estudio de [AmericanDiabetesAssociation et al., 2013] (ADA, por sus siglas en inglés) estima que la prevalencia de este padecimiento para el año 2012 sea de 7 %, que aunque difiere de lo reportado por la CDC, sigue siendo consistente con las mismas proyecciones. Se estima que cada año se diagnostican 1.4 millones de pacientes con diabetes. Este padeci-

miento está listado como la séptima causa de muerte en EE.UU., ya sea como causa principal del deceso o como causa subyacente. En relación a lo reportado por la OMS, EE.UU. se coloca en la posición 124 con 13.40 muertes entre 100,000 personas.

Independientemente del país del que se hable, se puede concluir que la diabetes mellitus, particularmente la Tipo II, es uno de los más grandes retos de salud pública que se enfrentan actualmente. Lo amenazante del padecimiento no es sólo el crecimiento constante en sus tasas de prevalencia sino también las complicaciones que éste presenta y que conducen, eventualmente, a la muerte. Como consecuencia, se tiene que el gasto asociado a este padecimiento y a sus complicaciones se vuelve cada vez más grande y muchas veces imposible de sostener por las instituciones proveedoras de servicios de salud.

En el último Atlas de la Diabetes elaborado por la [Federación Internacional de la Diabetes, 2015] (IDF, por sus siglas en inglés), se estima que el gasto asociado a la diabetes mellitus es de USD 673,000 millones, es decir, el 12 % del total del gasto mundial en salud. Esto quiere decir que en la mayoría de los países gastan entre un 5 % y 20 % de su presupuesto de salud, exclusivamente en el tratamiento de la diabetes mellitus. Esto debido a que los gastos en servicios de salud promedios de las personas con diabetes es entre dos y tres veces mayores que aquellos de las personas sin diabetes. En cara a estos datos, se vuelven prioritarias las estrategias de prevención y reducción de costos. Es importante destacar que los costos asociados a cualquier padecimiento, particularmente a la diabetes, deben tomarse en cuenta los costos directos del padecimiento (diagnóstico inicial, consultas, medicamentos, hospitalizaciones), los costos de las complicaciones (consultas, medicamentos, hospitalizaciones) y los costos indirectos (costos en productividad, muerte prematura,

ausentismo, etc.).

En el análisis realizado por [Hex et al., 2012] sobre la prevalencia y costos de la diabetes mellitus en el Reino Unido se estima un costo total de la diabetes en el año 2010/2011 de £23,631 millones, con una proyección de costos para el año 2035/2036 de £39,753 millones. El gasto de 2010/2011 es el 10 % del presupuesto asignado al Servicio Nacional de Salud (NHS, por sus siglas en inglés); mientras que la proyección de 2035/2036 llega hasta el 17 % de este mismo presupuesto. Lo que se muestra revelador de este estudio, en el cálculo de los costos de 2010/2011, son las diferencias entre los costos directos, de complicaciones e indirectos. Los costos directos son solo el 9 % de los costos totales, los costos de complicaciones 33 % mientras que lo más oneroso son los costos indirectos con una participación del 59 %; estas proporciones se mantienen con una leve variación para las proyecciones de 2035/2036. Igualmente se mantiene el supuesto que la DM-TII es mucho más costosa, pues el gasto total calculado en este tipo específico de diabetes oscila entre el 89 % y 92 %. Esto quiere decir que aunque el padecimiento en sÃ no tenga la mayor parte asociada de la carga económica, es la causa subyacente de muchas complicaciones y diagnósticos más costosos.

La región de Norteamérica y el Caribe tienen la mayor prevalencia de diabetes con un 12.9 % de la población adulta afectada, esto de acuerdo a la [Federación Internacional de la Diabetes, 2015]. En esta región el gasto total asociado está entre los USD 348,000 y USD 610,000 millones, este gasto representa más de la mitad del presupuesto mundial para la diabetes. Particularmente en México, según el estudio de [Barcelo et al., 2003], México es el país con los costos anuales más elevados en comparación con el resto de Latinoamérica. Al igual, que el análisis realizado para el Reino Unido, el costo anual de México de USD 15,118 millones,

este se puede dividir entre USD 1,974 millones de costos directos (diabetes y complicaciones) y USD 13,144 millones de costos indirectos. Estos costos asociados fueron calculados con una prevalencia del 4.1 %, la cual está desactualizada, por lo que se podría asumir que el gasto actual está por arriba. De nuevo y de manera más actual, el minucioso reporte de FUNSALUD realizado por [Barraza-Lloréns et al., 2015], estima que los costos asociados, particularmente para la DM-TII, sean de \$362,800 millones de pesos en el año base, de los cuales el 49 % se refieren a costos directos y el 51 % a costos indirectos. Para el año 2018, este mismo estudio proyecta un gasto total de \$506,000 millones de pesos; el aumento se explica con las proyecciones demográficas y de prevalencia estimada. Este estudio confirma que la DM-TII es mucho más cara y común que la Tipo I y pone en evidencia la importancia de mejor planeación financiera para hacer frente a esta epidemia de salud pública.

Según la FID, del gasto asociado a la región de Norteamérica, los EE.UU. representan la mayor parte del mismo con un gasto estimado en 2015 de USD 320,000 millones. De igual modo, el estudio de la [American Diabetes Association et al., 2013] estima que el gasto asociado con la diabetes mellitus en el 2012 es de USD 245,000 millones, de este monto el 72 % son costos directos del padecimiento y sus complicaciones y el 28 % restante se refiere a los costos indirectos; a diferencia del Reino Unido o México donde los costos indirectos sobrepasaban los costos directos. Los costos en salud de las personas diagnosticadas con diabetes son 2.3 veces más altos que los de aquellos no diagnosticados y según la Agencia para la Investigación y Calidad para el Servicios de Salud ([AHRQ, ], por sus siglas en inglés), la diabetes mellitus es el cuarto padecimiento más costoso en los EE.UU. Además, de acuerdo al estudio de [Zhuo et al., 2014], entre más joven sea el paciente a la edad de diagnóstico mucho mayor será el costo de salud acumulado. Debido a que EE.UU.

es uno de los países con mayor prevalencia y costos asociados a la diabetes mellitus en el mundo, la correcta cuantificación de los mismos es crucial para la correcta planeación financiera de los proveedores de salud.

Como podemos ver por las cifras reportadas, el problema del gasto asociado a la diabetes mellitus es uno de los más grandes retos que existen en el horizonte de la salud pública. Uno de los mecanismos principales para disminuir la prevalencia y por ende los gastos, se reducen a la prevención y a la procuración de un estilo de vida más sano; sobre todo para la DM-TII. Sin embargo, aunado a las recomendaciones para el paciente también se necesitan nuevas maneras de hacer estos gastos más efectivos mediante la correcta localización del rubro que genera más presión en el gasto y el uso más eficiente de los recursos existentes, para así poder hacerlos más accesibles tanto a los pacientes como a las instituciones proveedoras de salud.

Como se menciona en [Cichon et al., 1999] las herramientas estadísticas son de gran ayuda para entender la realidad actual y poder elaborar proyecciones futuras, especialmente en el campo de la medición de costos en el sector salud. Son muchas las preguntas que rodean la diabetes mellitus, acerca de su prevalencia, su morbilidad; sin embargo, la que se intenta contestar es cuánto cuesta la enfermedad? o cuánto costará?. Como respuesta a estas preguntas necesitamos no solo cuantificar los gastos asociados, sino hacer un análisis más profundo sobre la naturaleza de los mismos, para así poder pensar en estrategias para reducirlos y generar mayor bienestar en la población. En este trabajo de investigación se propone un nuevo modelo estadístico para modelar los gastos asociados a la diabetes mellitus, el cual se describe en el siguiente capítulo.

### **4.3. Descripción de los datos**

Descripción del comportamiento de los datos con estos modelos, particularmente los datos de diabetes mellitus en el estado de Michigan EEUU.

### **4.4. Análisis Descriptivo**

### **4.5. Resultados**

Descripción de los resultados inferenciales y predictivos.

## Capítulo 5

# Conclusiones

### 5.1. Qué me deja este trabajo?

### 5.2. Visión Crítica Autónoma

### 5.3. Trabajo Futuro

Como este trabajo sienta las bases para nuevos modelos de tarificación.

# Bibliografía

- [AHRQ, ] AHRQ. Medical expenditure panel survey, meps. [https://meps.ahrq.gov/mepsweb/data\\_stats/quick\\_tables\\_search.jsp?component=2&subcomponent=1](https://meps.ahrq.gov/mepsweb/data_stats/quick_tables_search.jsp?component=2&subcomponent=1). Accessed: 30-01-2017.
- [AmericanDiabetesAssociation et al., 2013] AmericanDiabetesAssociation et al. (2013). Economic costs of diabetes in the us in 2012. *Diabetes care*, 36(4):1033–1046.
- [Bagnoli and Bergstrom, 2005] Bagnoli, M. and Bergstrom, T. (2005). Log-concave probability and its applications. *Economic theory*, 26(2):445–469.
- [Barcelo et al., 2003] Barcelo, A., Aedo, C., Rajpathak, S., and Robles, S. (2003). The cost of diabetes in latin america and the caribbean. *Bulletin of the world health organization*, 81(1):19–27.
- [Barraza-Lloréns et al., 2015] Barraza-Lloréns, M., Guajardo-Barrón, V., Picó, J., García, R., Hernández, C., Mora, F., Athié, J., Crable, E., and Urtiz, A. (2015). Carga económica de la diabetes mellitus en méxico, 2013. *México, DF: Funsalud*.
- [Bolaños et al., 2010] Bolaños, R. d. l. Á. R., Shigematsu, L. M. R., Ruíz, J. A. J., Márquez, S. A. J., and Ávila, M. H. (2010). Costos directos de atención médica en pacientes con diabetes mellitus tipo 2 en



- méxico: análisis de microcosteo. *Rev Panam Salud Publica*, 28(6):412–20.
- [Casella and George, 1992] Casella, G. and George, E. I. (1992). Explaining the gibbs sampler. *The American Statistician*, 46(3):167–174.
- [CDC, ] CDC. National diabetes statistics report, 2014. <https://www.cdc.gov/diabetes/pdfs/data/2014-report-estimates-of-diabetes-and-its-burden-in-the-united-states.pdf>. Accessed: 30-01-2017.
- [Cichon et al., 1999] Cichon, M., Newbrander, W., Yamabana, H., Weber, A., Normand, C., Dror, D., and Preker, A. (1999). *Modelling in health care finance: A compendium of quantitative techniques for health care financing*. International Labour Organization.
- [Daley and Vere-Jones, 2003] Daley, D. and Vere-Jones, D. (2003). *An Introduction to the Theory of Point Processes: Volume I: Elementary Theory and Methods*. 2nd edition edition.
- [Dempster et al., 1977] Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the royal statistical society. Series B (methodological)*, pages 1–38.
- [Engle and Russell, 1998] Engle, R. F. and Russell, J. R. (1998). Autoregressive conditional duration: a new model for irregularly spaced transaction data. *Econometrica*, pages 1127–1162.
- [Fader and Hardie, 2013] Fader, P. S. and Hardie, B. G. (2013). The gamma-gamma model of monetary value.

- [Federación Internacional de la Diabetes, 2015] Federación Internacional de la Diabetes, B. (2015). Atlas idf diabetes, 2015. *Available from:[Last accessed: Enero 2017]*.
- [Gelman et al., 2014] Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2014). *Bayesian data analysis*, volume 2. Chapman & Hall/CRC Boca Raton, FL, USA.
- [Ghahramani, 2001] Ghahramani, Z. (2001). An introduction to hidden markov models and bayesian networks. *International journal of pattern recognition and artificial intelligence*, 15(01):9–42.
- [Gutiérrez et al., ] Gutiérrez, J., Rivera-Dommarco, J., Shamah-Levy, T., Villalpando-Hernández, S., Franco, A., Cuevas-Nasu, L., et al. Encuesta nacional de salud y nutrición. ensanut 2012 resultados nacionales. cuernavaca, méxico: Instituto nacional de salud pública (mx), 2012 [accedido en enero 2017].
- [Hahn and Zhang, 2012] Hahn, M. G. and Zhang, G. (2012). Exchangeable random variables. *High Dimensional Probability*, 43:111.
- [Harrison and West, 1999] Harrison, J. and West, M. (1999). *Bayesian forecasting & dynamic models*. Springer New York.
- [Held and Sabanés Bové, 2014] Held, L. and Sabanés Bové, D. (2014). *Applied statistical inference*, volume 10. Springer.
- [Hex et al., 2012] Hex, N., Bartlett, C., Wright, D., Taylor, M., and Varley, D. (2012). Estimating the current and future costs of type 1 and type 2 diabetes in the uk, including direct health costs and indirect societal and productivity costs. *Diabetic Medicine*, 29(7):855–862.
- [Neal, 2003] Neal, R. M. (2003). Slice sampling. *Annals of statistics*, pages 705–741.

- [OrganizacionMundialdelaSalud, ] OrganizacionMundialdelaSalud. Diabetes. fact sheet. <http://www.who.int/mediacentre/factsheets/fs312/en/>. Accessed: 20-01-2017.
- [Paik Schoenberg, 2000] Paik Schoenberg, F. (2000). *Introduction to Point Processes*.
- [Pitt et al., 2002] Pitt, M. K., Chatfield, C., and Walker, S. G. (2002). Constructing first order stationary autoregressive models via latent processes. *Scandinavian Journal of Statistics*, 29(4):657–663.
- [Plummer et al., 2003] Plummer, M. et al. (2003). Jags: A program for analysis of bayesian graphical models using gibbs sampling. In *Proceedings of the 3rd international workshop on distributed statistical computing*, volume 124, page 125. Vienna.
- [Resnick, 1999] Resnick, S. I. (1999). *A Probability Path*. Birkhauser.
- [Smith, 2010] Smith, J. Q. (2010). *Bayesian decision analysis: principles and practice*. Cambridge University Press.
- [Zhuo et al., 2014] Zhuo, X., Zhang, P., Barker, L., Albright, A., Thompson, T. J., and Gregg, E. (2014). The lifetime cost of diabetes and its implications for diabetes prevention. *Diabetes care*, 37(9):2557–2564.

## Apéndice A

# Paradigma Bayesiano

El paradigma bayesiano se refiere a una manera de hacer inferencia basado en el trabajo del inglés Thomas Bayes. En este paradigma se establece que la hipótesis se va actualizando de acuerdo a la nueva información relevante. Según [Gelman et al., 2014], una de las principales razones para el pensamiento bayesiano es que facilita la interpretación basada en el sentido común de conclusiones estadísticas.

De acuerdo con [Gelman et al., 2014] la inferencia bayesiana se hace en base en una evaluación retrospectiva del procedimiento utilizado para estimar el parámetro sobre la distribución de todas las posibles observaciones. Es decir, que mediante la regla de Bayes se describe la relación entre la asignación previa de la probabilidad y la reasignación de esta misma condicionada a los datos observados. El paradigma está basado en la regla o teorema de Bayes.

Una definición de la probabilidad condicional de  $y$  dado  $x$  será la división de la función conjunta de probabilidad entre la función de probabilidad

de  $x$ . Es decir,

$$p(y|x) = \frac{p(x,y)}{p(x)}$$

O en otras palabras, la probabilidad de  $y$  *suceda* dado  $x$  es la probabilidad de que *sucedan* ambos eventos relativo a que  $x$  *suceda* en absoluto.

Tomando en cuenta esta definición de la probabilidad condicional, el teorema de Bayes, se define como

$$p(y|x) = \frac{p(x|y)p(y)}{p(x)}$$

El Teorema de Bayes resulta muy útil cuando el modelo de probabilidad se basa en variables observables y parámetros,  $D$  y  $\theta$  respectivamente. De este modo, el modelo se escribe como

$$p(\theta|D) = \frac{p(D|\theta)p(\theta)}{p(D)}$$

Donde los elementos de esta ecuación significan

- $p(\theta|D)$  = la distribución posterior, es decir, la probabilidad de  $\theta$  tomando en cuenta las observaciones.
- $p(D|\theta)$  = la verosimilitud, es decir, la probabilidad de los datos generados por el modelo con el parámetro  $\theta$ .
- $p(\theta)$  = la distribución previa, es decir, la probabilidad de  $\theta$  sin tomar en cuenta las observaciones  $D$ .
- $p(D)$  = la distribución de las observaciones, es decir, la probabilidad total de las observaciones, ponderadas por todos los valores que puede tomar el parámetro de acuerdo al peso que se le asigna.

Esta misma ecuación se puede reescribir como,

$$\begin{aligned}
p(\theta|D) &= \frac{p(D|\theta)p(\theta)}{p(D)} \\
&\propto p(D|\theta)p(\theta) \\
&\propto \textit{verosimilitud} \times \textit{inicial}
\end{aligned}$$

La distribución posterior está en función del parámetro, por lo que la distribución posterior es proporcional a la multiplicación de la función de verosimilitud por la distribución inicial del parámetro. En otras palabras, la distribución previa del parámetro es la información a priori del mismo, que se va actualizando con las observaciones, tomando la información relevante al parámetro. Este mismo principio puede extenderse para varios parámetros.

Como describe [Gelman et al., 2014], esta lógica es similar para hacer inferencias sobre futuras observaciones. Una vez que se tienen todas las observaciones  $D = (d_1, \dots, d_n)$  se quiere inferir la siguiente observación  $d_{n+1}$ . La distribución de esta observación se llama la distribución llamada distribución posterior predictiva, posterior porque toma la información de las observaciones pasadas y predictiva porque predice la siguiente observación,

$$\begin{aligned}
p(d_{n+1}|D) &= \int p(d_{n+1}, \theta|D) d\theta \\
&= \int p(d_{n+1}|\theta, D) p(\theta|D) d\theta \\
&= \int p(d_{n+1}|\theta) p(\theta|D) d\theta
\end{aligned}$$

En la segunda y tercera línea de la ecuación se muestra la distribución posterior predictiva como un promedio de las distribuciones predictivas

condicionales de la distribución posterior del parámetro  $\theta$ . En la última línea se asume la independencia condicional de  $D$  y  $d_{n+1}$  dado  $\theta$ .

Como enunciado en [Smith, 2010], en el modelo bayesiano de inferencia, como descrito anteriormente; la evidencia, las observaciones y los argumentos científicos se utilizan para soportar la distribución de probabilidad del modelo propuesto. Por otro lado, en el análisis bayesiano de decisión el Tomador de Decisiones debe tomar este sustento evidencial y científico para resolver el problema específico que se encuentre.

Es decir, en un ambiente de incertidumbre tenemos el espacio de decisión  $A$  donde cualquier decisión puede ser tomada por el Tomador de Decisión y  $\Phi$  es el espacio de posibles resultados  $\phi$ . El Tomador de Decisiones debe cuantificar las consecuencias de elegir cada decisión  $a \in A$  para todos los posibles resultados  $\phi \in \Phi$ . Con esta información se deben especificar la función de pérdida  $L(a, \phi)$ , en la cual se mide la pérdida de tomar la decisión  $a$  con el resultado  $\phi$ . Otra función a especificar es la función de probabilidad  $p(\phi)$  que da las probabilidades de los posibles resultados  $\phi$  antes de tomar la decisión  $a$ ; esta función representa la incertidumbre que enfrenta el Tomador de Decisión. De este modo, la mejor decisión que se puede tomar es aquella que minimice la pérdida esperada,

$$\bar{L}(a) = \int_{\phi \in \Phi} L(a, \phi) p(\phi)$$

Con el enfoque del análisis bayesiano de decisión los problemas clásicos de inferencia como la estimación puntual, estimación por regiones y contraste de hipótesis pueden resolverse de esta manera. Además, los estimadores obtenidos no solo suelen coincidir con los estimadores clásicos en algunos casos, sino en otros casos de hecho los mejoran.