1. Modelos de Duración y Duración Marcada

Tenemos una muestra de microcostos de enfermedades crónicas de un cierto número de individuos, cada uno de los cuales tiene asociadas covariables sociodemográficas, socioeconómicas y médicas; así tenemos n individuos $(n_i)_{i=1}^{n}$ observados por un período de tiempo con costos asociados a su padecimiento. El objetivo es modelar y predecir la duración y el costo de las etapas de estos padecimientos por individuo.

Supongamos que empezamos el estudio de un individuo n_i en el tiempo $t_0 = 0$, es decir, este es el tiempo en el que el individuo entra al panel de estudio cuya duración es T; esto no quiere decir que no puedan ocurrir observaciones posteriores a T, a esto se le conoce como censuramiento de datos por la derecha. Sea $t_j \in (t_0, T]$ el momento en el que ocurre un cambio de tratamiento, de esta manera definimos la variable aleatoria $N(t)$ que cuenta el número de cortes o cambios en el intervalo.

A cada t_j se le asocia la variable precio; es decir, a cada momento en que ocurre un cambio de tratamiento le corresponde un nuevo precio p_j. De este modo, para cualquier individuo n_i tenemos una sucesión de variables asociadas $\{t_{i1}, p_{i1}\}, \{t_{i2}, p_{i2}\}, ..., \{t_{ik}, p_{ik}\}$. Dado que la sucesión de variables es una colección aleatoria de puntos en un espacio con una marca asociada a cada punto es un proceso puntual marcado.

Definicion 1. *Un proceso puntual $N(t)$ es una medida aleatoria en un espacio métrico separado S tomando valores en los enteros no negativos Z^+ (o infinito)[1].*

Definicion 2. *Un proceso puntual marcado localizado en un espacio métrico completamente separado χ y las marcas en otro espacio métrico completamente separado κ es un proceso puntual $\{(\chi_i, \kappa_i)\}$ en $\chi x \kappa$ con la propiedad adicional de que el proceso primario $N(t)$ es a su vez un proceso puntual[2].*

Lo que deseamos conocer es,

$$\mathbb{P}(t_{i1}, ..., t_{ik}, p_{i1}, ..., p_{ik}) = \mathbb{P}(t_{i1}, ..., t_{ik}, p_{i1}, ..., p_{ik} | N(t)) \tag{1}$$

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 $(t_0, T]$. Sin embargo, dado que al usar las variables en sus valores absolutos estas pueden dar saltos muy altos entre si debemos usar variables alternas.

Definimos las siguientes variables para un individuo n_i:

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

De este modo,

$$\mathbb{P}(t_{i1}, ..., t_{ik}, p_{i1}, ..., p_{ik}|N(t)) \cong \mathbb{P}(d_{i1}, ..., d_{ik}, c_{i1}, ..., c_{ik}|N(t)) \tag{2}$$

Referencias

[1] Paik Schoenberg,Frederic. Ïntroduction to Point Processes".

[2] Daley, D.J., Vere-Jones, D.J. .An Introduction to the Theory of Point Processes: Volume I: Elementary Theory and Methods", 2nd Edition.