Instituto Tecnológico Autónomo de México

Título

Tesis

que para obtener el título

Licenciado en Actuaría

presenta

Autor

Asesor: ...

México, D.F. 2016

Çon 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 Bailléres 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

Agradecimientos

Muchas gracias a todos!

Capítulo 1

Revisión sobre tarificación

1.1. Costos Directos de atención médica en pacientes con diabetes mellitus tipo 2 en México: análisis de microcosteo

La diabetes mellitus (DM) es un problema prioritario en el panorama de las instituciones de salud, pues por la naturaleza de la enfermedad es probable que se deriven complicaciones en los pacientes a las que estas instituciones deberán hacer frente. El tratamiento es extensivo y complejo por la gran cantidad de sistemas del cuerpo que el padecimiento ataca y el detrimento tan importante en la esperanza y calidad de vida del paciente. En México, esta enfermedad ha crecido su tasa de mortalidad en 3 % anual y mediante un censo realizado en 2004 por el Instituto Mexicano del Seguro Social (IMSS) se estiman que más de dos millones de personas son afectadas con este padecimiento, lo cual representa una gran carga financiera para estas instituciones.

Hay distintas maneras de calcular el costo de esta enfermedad para que

las instituciones puedan calcular anticipadamente la dimensión del gasto a efectuar. Para propósitos de estudio de este artículo se quiere estimar los costos directos de la atención médica a pacientes con DM tipo 2 atendidos específicamente en el IMSS.

El IMSS brinda sus servicios al 44 % de la población mexicana en cuatro regiones administrativas: Norte, Occidente, Centro y Sur. En total de las regiones existen 1340 unidades de atención médica (UAM) divididas en tres niveles de atención: primer nivel, donde se proporciona atención ambulatoria, segundo nivel, donde se brinda atención médica de las principales cuatro especialidades con apoyos técnicos y el tercer nivel, donde se proporciona atención médica de alta especialidad.

Para seleccionar una muestra representativa de las unidades médicas se tomaron en cuenta que la delegación donde hubiera una unidad de tercer nivel y al menos dos unidades representativas del segundo nivel de atención, además de criterios como el número de camas, población adscrita, etc. La muestra final quedó constituida de cuatro regiones, 12 delegaciones y 28 UAM. Después se procedió a seleccionar a los pacientes elegibles para participar en el estudio. Se considero el período comprendido entre junio 2002 a junio 2004. Se identificaron los casos de DM tipo 2 y sus complicaciones mediante la Clasificación Internacional de Enfermedades (CIE-10). Se redujo la muestra a aquellos pacientes cuyo expediente médico estuviera completo y que recibieran atención médica por primera vez en el período definido y que hubiera tendo mínimo un seguimiento de un año. La muestra final de pacientes consistió en 497 pacientes con diagnóstico confirmado de DM2 y proporcionalmente distribuidos entre la muestra de unidades médicas.

Una vez que se tienen las muestras definidas, se calculan los costos di-

rectos desde la perspectiva del proveedor de los servicios de salud. Se tomó el enfoque de metodología de costeo de enfermedad (CED) basada en la prevalencia que mide los costo actuales realcionados con la enfermedad. Aunque esta metodología puede realizarse bajo dos enfoques: arriba-abajo (top-down)y abajo-arriba (bottom-up), que es el utilizado en este análisis. En este enfoque se calculan los costos totales de la enfermedad a partir de los costos unitarios de manera que primero estima los costos de los tratamientos y procedimientos, para después agregarlos y obtener el costo total de la enfermedad.

Lo primero, se desarrollan modelos de costos unitarios considerando componentes fijos y variables asociados a las unidades médicas de segundo y tercer nivel. El modelo consiste, primero, en la información demográfica (número de pacientes), después los costos de capital fijos (inmuebles, mobilario, instrumentos), costos variables indirectos (recursos humanos no médicos, mantenimiento, consumibles), costos varibles directos (recursos humanos directos, medicamentos, material de curación, laboratorio) y por último, costos de eventos (cirugías, urgencias, cuidados intensivos, procedimientos ambulatorios); todo esto suma para el costo total de la enfermedad. Los costos de los materiales y medicamentos se obtuvieron con base en los precios promedios de 2008, el de los recursos humanos se calculó a partir del salario integrado neto anualizado y el costo de los activos fijos se depreciaron y prorratearon de acuerdo con la metodología de costo anual equivalente.

Los eventos médicos se dividen en cinco clasificaciones relevantes: ambulatorios, urgencias, hospitalarios, intervención quirúrgica, unidad de cuidados intensivos. El costo total anual por paciente considera los costos unitarios de cada evento y la frecuencia de utilización en el año de estudio, estos datos reportados en el expediente médico del paciente.

El costo total por paciente (CTP) se calcula,

$$CTP_{jkw} = \sum_{i=1}^{n_2} QR_{jkwi} \times PR_i$$

donde:

- CTP_{jkw}: Costo para el paciente k en evento médico j con DM2 en grado de severidad w.

- QR_{jkwi}: Utilización del recurso i durante un año para la atención para la atención de un evento médico j del paciente k con DM2 en grado de severidad w.

- PR_i: Precio o costo unitario del recurso i.

- w: Grado de severidad de la enfermedad. 1:sin complicaciones, 2:con complicaciones.

- k: Pacientes.

- j: Tipo de evento médico. 1:Atención ambulatoria, 2:Atención de urgencias, 3:Hospitalización, 4:Cirugías, 5:Cuidados intensivos.

- i: Recursos utilizados para la realización de la atención médica.

El costo total de la enfermedad es un promedio ponderado de los costos de atención médica por cada una de las complicaciones por el número de casos

registrados. La fórmula es:

$$CTE = CPA_{sincomplicaciones} \times N_{sincomplicaciones} + CPA_{concomplicaciones} \times N_{concomplicaciones}$$

donde:

- CTE: Costo total por enfermedad.
- CPA: Costo promedio anual por paciente en cada categoría.
- N: Casos incidentes de enfermedad en cada categoría.

El CPA se calcula con base en ls costos totales anuales de todos los pacientes para cada una de las complicaciones de DM2.

Tomando todas estas consideraciones, se procede a la ejecución del experimento. Primero se determinaron las características socioeconómicas de los pacientes, para después analizar y clasificar las complicaciones presentadas por los mismos y determinar los medicamentos administrados. Para efectos del estudio, se dividieron las conclusiones entre pacientes con y sin complicaciones, donde resultó que no había mucha deferencia entre estos dos grupos al observar la frecuencia y los costos de los eventos médicos. De este modo, el costo promedio por paciente con DM2 en México es de 3,193.8 USD, 2,740.34 USD para los pacientes sin complicaciones y 3,550.17 para los pacientes con complicaciones. Es importante mencionar qe estos son los costos promedios, pues como corresponde a cualquier padecimiento, el tratamiento es personalizado y no necesariamente estandarizado.

Ahora, para estimar los costos totales que los pacientes con DM2 de primer ingreso atendidos en durante 2008, que fueron 142,557 incurren hacia

el IMSS, se ponderan los costos totales según la proporcioón de pacientes con y sin complicaciones y sus costos estimados. El costo total resultante fue de 452,064,988 USD, esta cifra representa el 3.1 % del gasto total de la operación del IMSS.

Este análisis se considera más apegado a la realidad pues la metodología con la cual se calcularon los costos (CDE con enfoque bottom-up) permitió observar la demanda de servicios médicos, estudios y medicamentos de los pacientes diabéticos con más detalle. Aunque también se debe considerar que aún así puede existir un subreporte de algunos rubros, como el consumos de medicamentos por la misma forma en que se elaboran algunos reportes; a pesar de esta consideración, esta metodología se considera más verosímil que las anteriores con paneles de expertos o casos tipo. Otro mejora sustancial en la metodología fue la relevante a la estimación de costos unitarios acordes a las UAM de segundo y tercer nivel. Estos costos unitarios incluían costos de capital, variables, fijos y humanos con estrategias de amortización y prorrateo; los cuales fueron validados por las unidades contables de la misma institución asegurando así su precisión respecto a los costos reales.

Por otro lado, debido a la poca integralidad de la información, este estudio es deficiente en cuánto a la estimación de los costos de atención de primer nivel y costos indirectos generados por invalidez, muerte o incapacidad, pues estos datos no se pudieron accesar de manera completa al momento del estudio. Es por eso quee este estudio está limitado a la atención directa en servicios médicos de segundo y tercer nivel.

Mediante estimación por microcosteo se presenta un beneficio agregado al IMSS, pues podrá determinar de manera más acertada la localización de sus recursos logrando mejor planeación financiera y mejor atención al

derechohabiente, garantizada en la regulación pertinente.

La problacón em México sigue en aumento, al igual que su esperanza de vida, la tasa de urbanización y las bajas tasas de actividad física por lo que se proyecta que para 2025 existirán 11.7 millones de diabéticos en el país. El análisis de microcostos implica entonces una gran herramienta para estimar la carga al gasto público que implicarán estos pacientes. Al conocer el costo de mayores niveles de atención médica, se puede tomar la decisión de enfocar más recursos a la prevención, diagnóstico temprano y atención eficaz en las primeras etapas del padeciemiento. El manejo de la epidemia de la DM2 es uno de los grandes retos futuros a enfrentar por el sector salud por lo se impone el uso de soluciones interdisciplicarias.

1.2. 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

Se distinguen dos tipos de diabetes: el Tipo 1 que es una enfermedad autoinmune causada por una ausencia de producción de insulina en el cuerpo y que representa del 10-15 % del total de pacientes diagnosticados. El Tipo 2 es una enfermedad causada si el cuerpo produce insulina pero las células son resistentes a la misma y el 85-90 % de los pacientes con diabetes presentan este tipo. En la mayoría de los estudios se mezclan estos dos tipos, al igual que en los estudios tradicionales sobre costos. La diferenciación es crucial, pues los dos tipos de diabetes tienen costos y causas muy distintas.

Siguiendo la tendencia mundial, en UK la diabetes es una de los pade-

cimientos crónicos degenerativos más comunes. Por este motivo, el artículo se concentra en cuantificar los costos directos e indirectos de la diabetes mellitus en UK para poder hacer una proyección realista sobre los costos futuros en los que incurrirá el Servicio Nacional de Salud (NHS) con una distinción en los dos tipos de padecimientos.

En este estudio se utilizará el análisis de costos top-down, a diferencia del análisis mexicano donde se utilizó bottom-up, que consiste en estimar los costos partiendo de datos agregados y utilizando fuentes de información secundarios. Una vez que se establecieron los costos directos (prevalencia, incidencia, mortalidad, atención médica, complicaciones, etc.). Todos los costos son estimados a precios de 2010-2011.

Los datos de prevalencia, que es la proporción de personas que sufre una enfermedad con respecto al total de la población, fueron obtenidos por el Modelo de Prevalencia de Diabetes elaborado por la Asociación de Observatorios de Salud Pública (APHO). Este modelo identifica la prevalencia de diabetes en distintos grupos de edad. De este modo, se estimó que la prevalencia de Tipo 1 y Tipo 2 es de 15 y 85 %, respectivamente para el total de la población y para 10 y 90 % para la población adulta.

La información relativa a los costos se obtuvo, primordialmente, de la base de datos de costos del NHS donde se utilizaron los datos del año base (2010) ajustados con la correspondiente inflación. Los costos para 2035/2036 se obtienen con el crecimiento de la población enferma estimada y los precios ajustados a la inflación. El análisis puede estar subestimado porque hay varios rubros de los cuales no se pudo obtener información pero los tratamientos que se incluyeron en este estudio fueron: estudios de diagnóstico, consultas de primer nivel, medicamentos específicos para diabetes, medicamentos no específicos a diabetes, bombas de insulina y equipo monitor de

glucosa, cursos de educación sobre el padecimiento.

Para hacer una estimación real del costo real de los dos tipos del padecimiento, se tiene que tomar en consideración los costos de las complicaciones derivadas. Se tomaron datos del Instituto Nacional de Salud y Excelencia Clínica (NICE) sobre complicaciones cardiacas, renales, neuropatías; además de diversos estudios de observación. Estos datos indican la cantidad de pacientes afectados por las complicaciones y se utiliza las bases de datos de costos del NHS.

El costo económico de la diabetes mellitus dene incluir también los costos sociales y de productividad. Aunque no existen estudios específicos relativo al aspecto económico de estos aspectos, se puede estimar. Por ejemplo, utilizando los datos de mortalidad por diabetes que provee la Auditoría Nacional de Diabetes se puede estimar el número de personas que mueren de manera prematura por diabetes y cuantificar los años sueldo que perdió esa persona. También se estima el número de personas que están incapacitadas y se calcula el costo de productividad perdida igualmente para la cifra de ausentismo.

Como resultado del análisis se tiene que la prevalencia en 2010/2011 de diabetes mellitus tipo 1 y tipo 2 es de 400,000 y 3,400,000; respectivamente. Mientras que la proyección para 2035/2036 indica que habrán 650,000 personas con diabetes tipo 1 y 5,600,000 personas con diabetes tipo 2. En lo referente a costos obtenidos mediante la metodología top-bottom se hicieron mediante estimaciones de la incidencia y prevalencia de la enfermedad para después agregarlos usando costos unitarios, tenemos para los años base un costo directo de £719 millones para diabetes Tipo 1 y £7000 millones para diabetes Tipo 2 con la proyección hacia 2035/2036 £1,238 millones y £12,224 millones, respectivamente, estas cifras es más del 10 % del gastos

asignado a la NHS. Esto son solamente los costos directos, pues los indirectos reportan cifras aún mayores. Como se puede ver, la diferencia entre los tipos de diabetes es notable.

El análisis realizado demuestra que la carga económica de este padecimiento es muy grande y crecerá con el tiempo, sobre todo que al ser una enfermedad crónico degenerativa, menos de un cuarto de los costos estimados son relacionados con el tratamiento específico de diabetes mientras que lo restante son los costos relativos a las complicaciones. Al igual que los costos directos son mucho menores que los indirectos que se refieren a los costos de oportunidad de ausentismo y productividad. Mediante este análisis se puede identificar como cambios en el tratamiento de cualquier tipo de este padeciemiento influiría en el costo total pues es un modelo dinámico.

1.3. The lifetime cost of diabetes and its implications for diabetes prevention

Este artículo se refiere al panorama económico concerniente al diagnóstico de diabetes mellitus en USA. Se estima que si los recursos fueran empleados en prevención en vez de tratamiento, el ahorro no sería necesariamente mayor dado que sin el padecimiento, la esperanza de vida podría aumentar y los recursos serían empleados en los padecimientos relacionados a la vejez. De este supuesto nace la necesidad de explorar los costos médicos si los recursos se usaran en la prevención o en el tratamiento.

El costo individual de un paciente con diabetes se determina mediante tres pasos, primero se estima el gasto médico del tratamiento en el que incurren los pacientes, después se ajusta este gasto con la tasa de mortalidad por diabetes por sexo para poder y, por último, agregamos estos costos por el

remanente del tiempo de vida estimado. Este resultado se comparará con aquel relacionado al costo por supervivencia de una persona sin diabetes.

Los datos para estimar el costo anual de un paciente diagnosticado con dabetes se obtuvieron en la Encuesta del Gasto del Panel Médico (MEPS) que están a su vez ligados con los datos de la Encuesta Nacional de Entrevistas de Salud (NHIS) de 2005-2008. La MEPS es una encuesta representativa de toda la población estadounidense de donde se obtiene una muestra de hogares que participaron en la NHIS, donde cada año se toma una submuestra para darle seguimiento en el año siguiente, teniendo así una misma muestra dos año. Dado que la NHIS contiene información sobre el diagnóstico de los pacientes, incluyendo diabetes, por lo que también se obtiene la edad de diagnóstico, duración del padecimiento y gasto anual en el mismo.

Para delimitar la muestra del estudio se excluyen aquellos pacientes que reportaron tener diabetes solamente en la MEPS pero no el NHIS y aquellos que empezaron terapia de insulina antes de los 30 ños, dado que estos son pacientes de diabetes tipo 1 para la cual no hay mecanismos de prevención. Al final, la muestra quedó en 2,827 pacientes diagnosticados con diabetes tipo 2 y 29,413 que no.

Para estimar el gasto de ambas clases de pacientes se utilizó un modelo de dos partes, que considera las condiciones del padecimiento según cada pacientes (duración del mismo, edad de diagnóstico, etc.), sexo, raza, lugar de residencia, estado civil, entre otras cosas. Primero se hizo una regresión logística para estimar la probabilidad de que un individuo tenga gasto médico diferentes de cero y después, con un modelo lineal generalizado, se modeló el gasto médico anualizado dado que el paciente tenía un gasto distinto de cero. Estos modelos se ajustaron según padecimientos simultáneos reportados como: hipertensión, asma, artritis, colesterol alto, etc. que se

pudieran ver afectadas en su costo por el diagnóstico de diabetes, por lo que estas condiciones son incluidas en las regresiones.

Con los resultados de la regresión se tiene la media del gasto médico anual desagregado por las características del paciente con diabetes (sexo, edad, duración del padecimiento) y el gasto anual del paciente sin diabetes se calcula de la misma manera pero poniendo cero en la variable de estatus del padecimiento. Este método nos asegura que los gastos médicos de las poblaciones con y sin diabetes sean comparables entre si.

Las personas con diabetes tienen una mayor probabilidad de muerte prematura por lo que se deben ajustar el gasto a la supervivencia de los pacientes con y sin diabetes. Esto se calcula multiplicando el estimado gasto anual por la probabilidad del paciente a sobrevivir una edad determinada, una vez que obtenemos este gasto agregamos estas cantidades por el resto estimado del tiempo de vida de los pacientes de la muestra. Así podemos determinar y comparar los gastos médicos totales de los pacientes diagnosticados con diabetes y sin ese diagnóstico. Estas tasas de mortalidad se obtuvieron mediante el Buró de Censo de EUA, desagregadas por sexo y edad.

Se estimó el tiempo remanente de vida y el gasto correspondiente a las edades 40, 50, 60 y 65 años. La edad inferior es de 40 años es porque los caso de diabetes tipo 2 antes de esa edad son muy poco frecuentes y el límite superior de 65 años porque después de esa edad, cualquier ciudadano es acreedor al servicio de Medicare que ya no entra en el alcance de este estudio; aunque se siga calculando el gasto médico después de los 65 años hasta la muerte. Estos gastos están basados en precios de 2012 y se ajustaron con la inflación.

En este estudio, la prevalencia de diabetes fue de 7.4 %, de los cuales el

54 % fue diagnosticado entre 45 y 64 años de edad, la media de edad fue de 55 años con duración media de 9.4 *años*. Los adultos con diabetes fueron en comparación con aquellos sin diabetes, en promedio, 11 años más viejos, con menor ingreso, no caucásicos, menos educado y con mayor probabilidad a no tener seguro. Los pacientes con diabetes gastan al año USD 13,966, más del doble que los pacientes sin diabetes.

El gasto médico ajustado por la probabilidad de supervivencia disminuye después de los 60 años de edad para los diabetes con y sin diabetes por el decrecimiento en las tasas de supervivencia. Sin embargo, el gasto médico disminuye dramaticamente en los pacientes con diabetes pues, al ser más baja su supervivencia, el gasto se cancela al morir el paciente. Por ejemplo, un hombre de 40 años, con diabetes tiene 34 % de probabilidad de llegar a la edad 80; por el otro lado, este mismo hombre, sin diabetes, tiene 55 % de probabilidad de llegar a 80 años. El gasto médico anual de un hombre con diabetes es de USD 8,500 a la edad 40 y decrece a USD 3,400 a los 80 años, reduciendose a menos de la mitad; este decrecimiento en un hombre sin diabetes en el mismo período de tiempo es de USD 700 al gastar USD 3,900 a los 40 años y USD 3,200 a los 80. La brecha entre gastos médicos anuales se a haciendo menor entre más alta se vuelva la edad de diagnóstico inicial. Los gastos de las mujeres con diabetes son consistentemente más altos que aquellos de los hombre.

La edad del diagnóstico influye en saber el gasto médico por el tiempo remanente de vida. Es decir, en tre más tarde la edad de diagnóstico el paciente vive menos años, se pierden menos años de vida. Aún así, a cualquier edad analizada con cualquier edad de diagnóstico, el paciente con diabetes tendrá un gasto médico mucho mayor que uno sin ese padecimiento, donde los gastos más grandes se registran en medicamentos y cuidados del paciente.

En conclusión, los pacientes con diabetes a pesar de tener una esperanza de vida menor que sus contrapartes, acumulan un gasto médico mucho mayor en el tiempo de vida remanente. Lógicamente, entre mayor sea la edad de diagnóstico, menor es el gasto médico acumulado por su menor esperanza de vida. En el estudio comparativo de diabetes con y sin diabetes, se pudo hacer un análisis más profundo sobre lo que eleva el costo del paciente diabético es el costo mismo del tratamiento y la atención a las complicaciones mismas de la enfermedad o enfermedades simultáneas. A pesar de estas consideraciones, el costo médico resultado de este análisis es menor que el obtenido mediante otros estudios por que en los costos fueron ajustados por la tasa de supervivencia.

Conocer los costos de la diabetes toma un papel preponderante en la planeación financiera de las instituciones de salud actuales, pues en las últimas tres décadas la población diagnósticada con diabetes mellitus tipo 2 se ha triplicado. Este aumento desmedido en la prevalencia de DM2 implica una gran carga en el presupuesto nacional dedicado a la salud, lo que pone en relevancia el impulso a prácticas de prevención.

Al realizar un análisis de costos, sabemos que para realizar eficientemente uso de los mismos, se debe gastar menos en prevención efectiva y en los gastos médicos de las personas sin diabetes que en los costos médicos de los pacientes con diabetes. En diversos estudios se ha demostrado la causalidad entre diabetes y menor esperanza de vida, por lo que queda demostrado cómo un cambio en el estilo de vida puede reducir el riego de ser diagnósticado con diabetes en 50-58 %. Estas estrategias de prevención pueden ser logradas a muy bajo costo e incluso tener un retorno a largo plazo, haciendo estas estrategias financieramente eficientes.

Este estudio, al usar datos reales de pacientes a lo largo de USA tiene limitaciones inherentes a la información, pues el reporte del diagnóstico, al ser generado por el mismo paciente, puede ser que esté subestimado, resultando en un costo médico subestimado. También puede ser que la información de costos se vea alterado pues en el estudio se utilizan precios y tratamientos de 2012, y los precios de los tratamientos pueden cambiar dependiendo de desarrollos tecnológicos o cambios en los mismos. Por último, los costos pueden cambiar dado que hubo datos que no se pudieron conseguir, las tasas de supervivencia no son reflejo de toda la población pues solo se tomaron en cuenta a los pacientes no institucionalizados, las tasas pueden estar sesgadas; tampoco se obtuvieron los datos de costos médicos de los pacientes a la edad de muerte, por lo que no se puede comparar entonces entre pacientes con y sin diabetes.

Como resultado general, se puede determinar que el uso más eficiente de los recursos es en la promoción de estrategias de prevención, pues aún con una menor esperanza de vida en pacientes con diabetes, el gasto generado por ellos es mayor a los pacientes sin diabetes. Es por esto que se deben buscar modos de prevenir diabetes eficientes para lograr disminuir el costo médico a largo plazo.

1.4. Estimation of medical costs by copula models with dynamic change of health status

El cálculo de los costos médicos es de suma importancia para el análisis de riesgos, conocer la efectividad de los mismos y, para la aseguradoras, saber si las primas y reservas están calculadas correctamente. Aunque al inicio pudiera parecer que el estudio estadístico es muy simple, en realidad

hay varias razones por las cuales los datos de costos no se pueden modelar tan fácilmente como el sesgo inherente a los datos, el costo acumulado está correlacionado con el tiempo de supervivencia, tanto los datos de costos como los de tiempo de supervivencia están censurados por la derecha, muchas veces en algunos rubros que no aparecen se obtienen muchos ceros, etc. Para poder sortear este problema se han propuesto varias soluciones: modelos paramétricos, modelos con funciones hazard proporcionales, modelos de regresión, etc.

Dentro de las cosas a tomar en cuenta para hacer el mejor estudio posible se propone usar los .estados de salud"dinámicos, es decir, que el paciente entre estados con probabilidades de transición siguiendo un modelo de Markov. El tiempo que el paciente permanece en un estado antes de trancisionar al siguiente se le denominará "sojournz al costo asociado a ese tiempo. Además, este modelo tendrá las características de que los datos son a nivel individual, los costos y los tiempos de espera son dependientes entre estados, esto se modela con una cópula, los valores en cero se acomodan fácilmente. Esto tiene como ventaja que es mucho más preciso que el análisis tradicional.

La relación entre costos y la etapa del padecimiento a veces puede ser un poco confusa, pues el costo acumulado también depende de la tasa de supervivencia. Es decir, a veces puede ser más costosa una enfermedad crónico degenerativa que una terminal, pues el primer paciente tiende a sobrevivir muchos más años. Es por esto que se utiliza una cópula, para tomar en cuenta las correlaciones entre costo y tiempo entre los estados.

Para poder estudiar el cambio dinámico entre los estados de salud, se define primero un conjunto finito de los mismos entre los que se mueve el paciente en un intervalo determinado. Sea $E = \{1, 2, ..., D\}$ el conjunto de estados

de salud para todos los pacientes y $X(t)$ un proceso de conteo que modela el estado de salud de un paciente al tiempo t. Este proceso es homogéneo en el tiempo, con probabilidades de trancisión:

$$P(X(s+t) = h|X(s) = k, X(u) : u < s) = P(X(s+t) = h|X(s) = k) = p_{kh}(t)$$

Se denomina T_j el tiempo o sojourn entre la $j-1$ y j trancisión o salto. Entonces la cadena embebida $\{\hat{X}_0, \hat{X}_1, ...\}$ definida por $\hat{X}_0 = X(0)$ y $\hat{X}_j = X(T_1 + ... + T_j)$ es una cadena de Markov homogénea.

Se asume que el sojourn T_j se distribuyen exponencialmente con media $1/\nu_k$ y con la función de supervivencia,

$$S_k(t) = P(T_j > t) = \exp\{-\nu_k t\} \qquad j = 1, 2, ..., \nu_k > 0$$

Este modelo es muy conveniente para la progresión clínica de un padecimiento pues puede incorporar fácilmente distintas covariables.

Suponiendo que la tasa de acumulación de costos de un paciente i al tiempo t en el estado h es denotado por $B_h^i(t)$, tal que en el intervalo $[t, t+dt]$ el costo se denotaría $B_h^i(t)dt$. Los costos del sojourn se consideran eventos recurrentes condicionados al estado de salud y al tiempo de trancisión, variables de las cuales los costos pueden ser dependientes.

Sea t_{ij} el punto en el tiempo donde el paciente i hace la j-ésima trancisión de un estado a otro. Para la j-ésima entrada suponemos que el paciente i entra al estado h al tiempo t_{ij} con probabilidad $\pi_h(t_{ij})$ y ocupa ese estado entre t_{ij} a $t_{i(j+1)}$. El sojourn Δ_{ijh} se distribuye de manera exponencial y se

calcula el costo de esta sojourn,

$$m_{ijh} = E[M_{ijh}] = E[\int_{t_{ij}}^{t_{ij}+\Delta_{ijh}} b_h^i(s|Z_{ih})ds$$

$$b_h^i(s|Z_{ih}) = E[B_h^i(s)|Z_{ih}] = b_{0h}(s)\exp(Z_{ih}\top\beta_h)$$

Se puede asumir que los costos se distribuyen de manera log-normal.

El modelo a desarrollar consiste en un modelo de cópulas. Una cópula es una función bivariada que especifica la distribución conjunta de dos variables aleatorias dependientes (X, Y) dadas sus distribuciones marginales. Para la correlación entre el sojourn Δ_{ijh} y su costo M_{ijh} de las observaciones dependientes, se propone un modelo de cópula con parámetro γ,

$$C_\gamma(S^1_{\Delta_{ijh}}(\cdot), S^2_{M_{ijh}}(\cdot))$$

donde $C_\gamma(\cdot,\cdot)$ es la función cópula con indexada por un valor real y $S^1_{\Delta_{ijh}}(\cdot)$ y $S^2_{M_{ijh}}(\cdot)$ son las funciones marginales de supervivencia. Este modelo se puede utilizar para predecir el costo total de los sojourns de cualquier conjunto de variables observadas que se comporten de la manera establecida.

Dado que este modelo de cópulas describe la dependencia entre las trancisiones de los estados de salud y los costos totales de cada sojourns, ha probado ser muy útil para el cálculo particular de costos de salud. Las ventajas de este modelo es la flexibilidad para incorporar los cambios entre los estados de salud y el ajuste que tiene a las estructuras de dependencia entre sojourns y costos.

Sobre la elección de la cópula, se eligió una cópula arquímedeana particularmente una Clayton. Las cópulas arquímedeanas C_φ es una función de supervivencia con dos variables y densidad en $[0,1]^2$. Sea W_1 el tiempo de

falla en un evento terminal y W_2 el costo total en el intervalo $[0, \tau]$. El modelo de cópula especifica la función conjunta de supervivencia $S(\cdot, \cdot)$, densidad conjunta $f(\cdot, \cdot)$ de (W_1, W_2) mediante la cópula C_φ

$$S(x_1, x_2) = C_\varphi(S_{W_1}(x_1), S_{W_2}(x_2)) \quad x_1, x_2 \geq 0$$
$$f(x_1, x_2) = C_\varphi(S_{W_1}(x_1), S_{W_2}(x_2)) f_{W_1}(x_1) f_{W_2}(x_2)$$

Si φ es una función dos veces diferenciable, estrictamente decreciente y convexa entonces esta función tiene una inversa φ^{-1} análoga. Cada función φ genera la siguiente cópula,

$$C_\varphi(u, v) = \varphi\{\varphi^{-1}(u) + \varphi^{-1}(v)\} \quad si 0 \leq u, v \leq 1$$

Por lo que la cópula Clayton se define como,

$$C_\gamma(u, v) = (u^{-\gamma} + v^{-\gamma} - 1)^{-1/\gamma} \quad \gamma > 0$$

esta es una cópula arquímedeana con generador $\varphi(t) = (1 + t)^{-1/\gamma}$.

Para construir las funciones de verosimilitud, se asume que las distribuciones de sojourns y sus costos se distribuyen Gamma.

Capítulo 2

Modelos de Duración y 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, estacionareidad que son vitales para realizar inferencia y predicción de los datos.

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. Definición del proceso de Duración y Duración Marcada

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, socioeconómicas y médicas. De este modo, podríamos decir que 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 $t_{i0} = 0$, es decir, este es el tiempo en el que el individuo 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^+ (o infinito) 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 un cambio de tratamiento, 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, decimos que a cada t_j se le asocia la variable costo de tratamiento; es decir, a cada momento en que ocurre un cambio de tratamiento le corresponde un

nuevo costo p_j. De este modo, para cualquier individuo n_i tenemos una sucesión de variables asociadas $\{t_{i1}, p_{i1}\}, \{t_{i2}, p_{i2}\}, ..., \{t_{ik}, p_{ik}\}$. 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 χ y las marcas en otro espacio métrico completamente separado κ, 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,

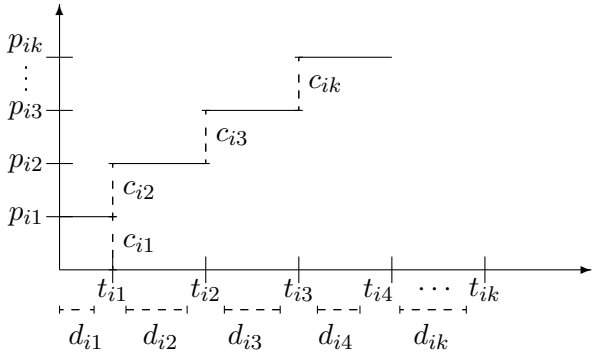
$$P(t_{i1}, ..., t_{ik}, p_{i1}, ..., p_{ik}) = P(t_{i1}, ..., t_{ik}, p_{i1}, ..., p_{ik}|N(t)) \qquad (2.1)$$

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, por lo que debemos usar variables alternas.

Definimos las siguientes variables para un individuo n_i:

- $d_{ij} = t_{ij} - t_{ij-1}$, donde d_{ij} es la duración entre los tiempos de ocurrencia de cada individuo.

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



.

De este modo,

$$P(t_{i1}, ..., t_{ik}, p_{i1}, ..., p_{ik}|N(t)) \cong P(d_{i1}, ..., d_{ik}, c_{i1}, ..., c_{ik}|N(t)) \qquad (2.2)$$

Esto quiere decir que calcular la función de distribución conjunta de los tiempos de ocurrencia y los precios asociados a éstos es análogo a 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.

2.3. Propiedades del Proceso 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 estacionareidad.

Independencia

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

Definicion 1. *Los eventos* A_1, ... , A_n *(n≥ 2) son independientes si*

$$P(\bigcap_{i\in I} A_i) = \prod_{i\in I} P(A_i), \qquad I \subset \{1, ..., n\}$$

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:

$$P(d_1, c_1, ..., d_k, c_k | N(t) = k) = \prod_{j=1}^{N(t)} P(d_j, c_j) \tag{2.3}$$

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:

Definicion 2. *Una sucesión de variables* $X = (X_1, X_2, ..., X_n)$ *es intercambiable si para cada* n

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

para cualquier permutación σ *de* $1, 2, ..., n$.

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* $\bar{X} = (X_1, X_2, ...)$ *es una mezcla de varibles independientes e idénticamente distribuidas (i.i.d). Esto es, que existe un espacio de probabilidad* (U, Θ) *tal que*

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

donde $\bar{X}(u) = (X_1(u), X_2(u), ...)$ *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

$$P(d_1, c_1, ..., d_k, c_k | N(t) = k) = \int_\Theta \prod_{j=1}^{N(t)} P(d_j, c_j|\theta)\pi(\theta)d(\theta) \tag{2.4}$$

donde θ 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 desarrollará a profundidad en el siguiente capítulo.

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* $r = 1, 2, ...$ *y todos los enteros* $i_i, ..., i_r$ *, la distribución conjunta de* $\{\tau_{i_{1+k}}, ..., \tau_{i_{r+k}}\}$ *no depende de* k *(*$k = 0, \pm 1, ...$*).*

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 $\bar{X} = (X_1, ..., X_n)$ tendríamos que,

$$P(X_n, ..., X_1) = P(X_n|X_{n-1}, ..., X_1) * P(X_{n-1}|X_{n-2}, ..., X_1) * \cdots * P(X_2|X_1) * P(X_1)$$

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

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

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

De este modo, para el proceso de duración marcada la estacionareidad se podría plantear como,

$$P(d_1, c_1, ..., d_k, c_k|N(t) = k) = \prod_{j=2}^{N(t)} P(d_j, c_j|d_{j-1}, c_{j-1}) * P(d_1, c_1)$$

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

Una vez que nuestro modelo de duración marcada cumple las propiedades descritas en este capítulo podemos empezar a hacer inferencia sobre las variables y predecir las observaciones futuras. En el siguiente capítulo, 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.

Capítulo 3

Revisión de Literatura

3.1. Full backward non-homogeneous semi-Markov processes for disability insurance models: A Catalunya real data application

Los procesos semi-Markov han sido utilizados en contextos financieros, actuariales y de demografía. Éstos procesos se refieren a aquellos procesos aleatorios que evolucionan con el tiempo y cuyas realizaciones en cualquier momento dado del tiempo tiene un estado definido. Por lo que, la generalización de las probabilidades de trancisión de los procesos semi-markovianos no homogéneos se obtiene introduciendo la reversibilidad, pues en este caso las probabilidades de trancisión dependen del tiempo en el que el proceso entró en un cierto estado, no como en un proceso semi-markoviano homogéneo donde se entra al sistema en el estado inicial al tiempo inicial. La recurrencia en el tiempo en los procesos reversibles se pueden considerar al inicio o al final del horizonte de tiempo considerado.

Suponemos las siguientes variables aleatorias (J_n, T_n) como un proceso de

renovación de Markov no homogéneo, J_n representa el estado a la n-ésima trancisión y T_n el tiempo a la n-ésima trancisión, de este modo se define $X_n = T_{n+1} - T_n$ como el proceso de tiempo de llegada. Con esta información, se puede definir lo siguiente:

- $Q_{ij}(s,t) = P(J_{n+1} = j, T_{n+1} \leq t | J_n = i, T_n = s)$.
  La probabilidad de que en la $n+1$-ésima realización el proceso esté en el j-ésimo estado en un tiempo menor o igual a t si en la realización anterior estaba en el i-ésimo estado en el tiempo s. ij son los estados en los que está el proceso y (s,t) los tiempos del mismo.

- $H_i(s,t) = P(T_{n+1} \leq t | J_n = i, T_n = s) \Rightarrow H_i(s,t) = \sum_{j=1}^{m} Q_{ij}(s,t)$.
  Probabilidad que el proceso salga del estado i en el intervalo del tiempo s a t en una sola realización.

- $F_{ij}(s,t) = P(T_{n+1} \leq t | J_n = i, J_{n+1} = j, T_n = s)$.
  La función de distribución del tiempo de espera en cada estado i dado que el estado en la realización siguiente es conocido.

La mayor diferencia entre los procesos de Markov no-homogéneos discretos y los procesos semi-Markov reside en las funciones $F_{ij}(s,t)$. En los primeros ésta tendría que comportarse como una función de distribución geométrica, mientras que con el proceso semi-Markov no-homogéneos discretos ésta puede ser de cualquier tipo.

Ahora definimos el proceso de conteo de realizaciones como,

$$N(t) = sup\{n \in \mathbb{N} : T_n \leq t\}$$

Ya que tenemos establecido el concepto de proceso de conteo se puede definir el proceso semi-Markov no-homogéneo discreto, $Z(t) = J_{N(t)}$ como el

estado ocupado por el proceso a cada momento del mismo. Por lo que las probabilidades de trancisión serían,

$$\phi_{ij}(s,t) = d_{ij}(s,t) + \sum_{\beta=1}^{m} \sum_{\vartheta=s+1}^{t} b_{i\beta}(s,\vartheta)\varphi_{\beta,j}(\vartheta,t)$$

La primera parte de la fórmula $d_{ij}(s,t)$ se refiere a la probabilidad de que el proceso no transicione al tiempo t dado que entró en el estado i al tiempo s, esto ocurre solo cuando $i = j$, es decir, es referente al tiempo de transición no al estado. La segunda parte $(b_{i\beta}(s,\vartheta))$, representa la probabilidad de que el sistema entre al estado β justo en el tiempo ϑ dado que entró al estado i en el tiempo s. Después de la trancisión, el sistema llegará al estado j en el tiempo t siguiendo cualquiera de las posibles trayectorias que van del estado β al tiempo ϑ.

Â¿Se podría interpretar las variables β, ϑ como variables latentes que nos podrían dar información sobre las variables que sí observamos? Eso podría dar a entender que en la segunda sumatoria donde la variable ϑ empieza el conteo al tiempo $s+1$ y termina al tiempo t, es decir recorre la trayectoria entre los tiempos que conocemos.

Ahora bien, definimos $B(t) = t - T_{N(t)}$ como el proceso reversible que se denota como,

$${}^{b}\phi_{ij}(l,s;t) = P(Z(t) = j | Z(s) = i, B(s) = s - l)$$

$$\phi_{ij}^{b}(s;l',t) = P(Z(t) = j, B(t) = t - l' | Z(s) = i)$$

Estos son las probabilidades de transición del proceso semi-Markoviano con el tiempo de recurrencia reversible al inicio y al final, respectivamente.

En la primera ecuación sabemos que el sistema está en el estado i al tiempo s, sabemos también que entró a ese estado en el tiempo l por lo que $s-l$ representa el tiempo reversible inicial; así que lo que se busca la probabilidad de estar en el estado j al tiempo t. En la segunda ecuación sabemos que el sistema entró al estado i al tiempo s, en este caso el objeto de interés es saber la probabilidad de estar en el estado j al tiempo t entrando a este estado en el tiempo l'; el tiempo reversible final es $t-l'$.

Si definimos un proceso reversible en el tiempo inicial y final se tiene que,

$$ {}^{b}\phi_{ij}^{b}(l,s;l',t) = P(Z(t)=j, B(t)=t-l' | Z(s)=i, B(s)=s-l) $$

De igual modo que con el proceso sin la reversibilidad, se definen las siguientes probabilidades de transición:

$$ {}^{b}\phi_{ij}(l,s;t) = d_{ij}(l,s;t) + \sum_{\beta=1}^{m}\sum_{\vartheta=s+1}^{t} b_{i\beta}(l,s;\vartheta)\varphi_{\beta j}(\vartheta,t) $$

$$ \phi_{ij}^{b}(s;l',t) = d_{ij}(s,t)\mathbf{1}_{\{l'=s\}} + \sum_{\beta=1}^{m}\sum_{\vartheta=s+1}^{l'} b_{i\beta}(s,\vartheta)\varphi_{\beta j}^{b}(\vartheta;l',t) $$

$$ {}^{b}\phi_{ij}^{b}(l,s;l',t) = d_{ij}(l,s;t)\mathbf{1}_{\{l'=s\}} + \sum_{\beta=1}^{m}\sum_{\vartheta=s+1}^{l'} b_{i\beta}(l,s;\vartheta)\varphi_{\beta j}^{b}(\vartheta;l',t) $$

En las últimas dos ecuaciones donde se tiene el término $\mathbf{1}_{\{l'=s\}}$, se refiere a que la expresión es igual a 1 si y solo si $\{l'=s\}$, sino es igual a 0.

La primera ecuación se refiere a la probabilidad de que el sistema esté en el estado j en el tiempo t dado que estaba en el estado i en el tiempo s entrando a ese estado al tiempo l, si $l=s$ entonces tenemos el proceso sin reversibilidad.

La segunda ecuación da como resultado la probabilidad de que el sistema llegue al estado j al tiempo l' y permanecerá allí hasta el tiempo t, dado que entró al estado i al tiempo s . La primera parte $d_{ij}(s,t)\mathbf{1}_{\{l'=s\}}$ significa la probabilidad de que no exista trancisión de estados entre el tiempo s,t por lo que el tiempo reversible final $t-l'$ debe ser exactamente igual que $t-s$ y solo tiene sentido cuando $i=j$. La segunda parte se refiere significa que el sistema no se mueve del tiempo s al tiempo ϑ y que, justo en este tiempo, salta al estado β; después siguiendo cualquiera de las trayectorias posibles, el sistema llega al estado j en el tiempo l' y se queda allí hasta el tiempo t.

Es importante mencionar que teniendo todos los valores del proceso reversible final tenemos las posibles probabilidades de transición del proceso sin reversibilidad, es decir,

$$\phi_{ij}(s,t)=\sum_{l'=s}^{t}\phi_{ij}^{b}(s;l',t)$$

Por último, la tercera ecuación expresa la probabilidad de que el sistema entre al estado j al tiempo al tiempo l' y se queda sin transicionar hasta el tiempo t, dado que entró al estado i al tiempo l y se quedó allí hasta el tiempo s. El primer término $d_{ij}(l,s;t)\mathbf{1}_{\{l'=s\}}$ es, de manera análoga, la probabilidad de no tener trancisiones de estados entre los tiempos l a t, es decir, quedarse en el estado i dado que no ocurrieron transiciones entre los tiempos l a s. Esta probabilidad es distinta a 0 si y solo si $i=j$ y $l'=s$. La segunda parte de la ecuación representa la probabilidad de hacer la siguiente transición al tiempo l del estado i a cualquier estado β al cualquier tiempo ϑ, para después seguir cualquier trayectoria para llegar al estado j al tiempo l' sin volver a moverse hasta el tiempo t.

Este modelo es el que se utiliza en algunos cálculos referentes a modelos de incapacidad que consideran tiempos reversibles iniciales y finales. Estos modelos tienen los siguientes estados, con sus respectivas transiciones:

- Activo
    - Activo
    - Pensionado
    - Incapacitado
    - Muerte
- Pensionado:
    - Pensionado
    - Incapacitado
    - Muerte
- Incapacitado:
    - Incapacitado
    - Muerte
- Muerte

Es decir, que la muerte es un estado absorbente.

El experimento a desarrollar en el artículo es sobre una población de 150,000 asegurados de la cobertura de invalidez en la región de Cataluña,España durante 30 años. La condición de invalidez es verificada por un perito médico y corresponde a los padecimientos establecidos en la póliza.

Se aplica el modelo de procesos semi-Markov discretos no-homogéneos con

reversibilidad incial y final a la muestra. Dado que el número de transiciones no era suficiente para intervalos de un año de edad, se construyeron grupos de cinco años de edad con los cuatro estados descritos. Los resultados mostraron diferencias según el tiempo de reversibilidad aplicado, inicial o final. Este modelo sirve exclusivamente para ver las transiciones entre estados, por lo que el siguiente paso sería incluir la modelación de costos asociados a cada estado.

3.2. Parametric Modelling of cost data, some simulation evidence

Los estudios concernientes al análisis de costos medios de algún padeciemiento se complican cuando se llevan a cabo con datos observados pues éstos se pueden ver muy sesgados con datos de unos pocos pacientes con costos muy altos.Para lidiar con estas complicaciones se han propuestos varias soluciones como usar métodos no paramétricos, transformar los datos, tomar la media muestral, etc. Cada una de estas soluciones conlleva sus respectivas críticas, por lo que no hay consenso sobre el mejor camino a seguir.

Es por esto que el propósito de este artículo es explorar dos particulares opciones para calcular estimadores de la población de la media de los costos de tratamientos hospitalarios. La primera opción consiste en observar el comportamiento de los datos cuando se someten a restricciones de parámetros supuestos. Para la segunda opción, se repite la comparación utilizando tres muestras con datos de costos hospitalarios y sacando los estimadores empíricos. El objetivo es evidenciar el beneficio en eficiencia que se obtiene utilizando los estimadores adecuados y también lo costoso que

sería lo contrario.

Para el primer acercamiento al problema, usaremos dos distribuciones: Gamma y log-normal, pues ambas se utilizan para modelar datos sesgados positivamente. Al utilizar estimadores de máxima verosimilitud (EMV) tenemos, para la distribución Gamma que su EMV es la media muestral y para la distribución log-normal es $exp(lm + lv/2)$ donde lm y lv son la media y varianza en escala logarítmica, respectivamente.

En el experimento, para ambas distribuciones, la media de la población fue designada de 1000 con cinco opciones de coeficientes de variación (CoV= 0.20,0.50,1.0,1.5,2.0) para definir los parámetros de la distribución. A su vez, se hicieron experimentos con cinco distintos tamaños de muestra (n= 20,50,200,500,2000) para cada CoV, lo cual resulta en 50 experimentos y para cada uno de ellos se realizaron 10,000 realizaciones.

Para observar el sesgo y la precisión de los estimadores se calcula la Raíz del Error Mínimo Cuadrado (REMC), como esperado este coeficiente decrece entre es más bajo es el coeficiente de variación y mayor es la muestra sin importar el EMV que se utilice.

Cuando los datos son log-normales el mejor estimador es su propio EMV y exhiben menor REMC que con el estimador de la media muestral; en cambio, cuando el estimador log-normal es aplicado a datos que se distribuyen Gamma los resultados son terribles, sobre todo cuando el CoV es más grande. Esto se debe a que el estimador log-normal está mucho más sesgado a cambios en el coeficiente de variación, mientras que la media muestral no. Estos resultados se repiten cuando medimos el intervalo de confianza, entre menos coeficiente de variación y mayor número de muestra, mayor es el intervalo de confianza.

El segundo enfoque es usar tres bases de datos observados de costos hospitalarios y sacar estimadores empíricos, pues dificilmente se obtienen datos que se comporten como una distribución. Estas bases de datos a nivel costo-paciente son,

- Datos CPOU: Datos obtenidos de la Unidad de Observación para Dolores de Pecho en un hospital escuela. Se reclutaron 972 pacientes con costos a precios de 2001-2002 de las primeras 6 horas de hospitalización, duración de la misma, medicinas, estudios, procedimientos, etc.

- Datos de Fluidos IV: Estos datos fueron obtenidos mediante dos protocolos de atención con fluidos intravenosos aplicado por paramédicos en pacientes con traumas severos antes de llegar al hospital. Se obtuvieron datos de costos hasta 6 meses después del trauma, costos de ambulancia, de fluidos, cuidado ambulatorio para 1191 pacientes a costos de 1997-1998.

- Datos de Paramédicos: Estos datos se consiguieron a traveés de un estudio controlado de paramédicos y técnicos de ambulancia para pacientes con traumas. La muestra de pacientes es de 1852 con datos de 1996-1997 con costos hasta de 6 meses después del trauma inicial incluyendo costos de ambulancia y tratamientos, hospitalización y cuidados ambulatorios.

En un primer análisis exploratorio de los datos vemos que la curtosis y el sesgo de los datos es muy grande comparado con los valores de la normal, es también interesante que la desviación estándar es lo doble a la media para las tres bases de datos. Dado estos resultados se transforman los datos mediante el logaritmo natural, lo que hace que los resultados de media,

varianza, sesgo y curtosis se vean más normales.

Tomando otra vez el experimento de simulación, de nuevo con 10,000 realizaciones, que consiste en la extracción de datos de manera aleatoria sin reemplazo de las bases de datos, modificando el tamaño de la muestra, es decir, (n=20,50,300,500) se vuelven a usar los estimadores anteriores: media muestral y $exp(lm + lv/2)$ con sus recpectivos intervalos de confianza. Una vez realizado este experimento se analizan los resultados con muchas similitudes a los resultados anteriores.

La REMC, como esperado, disminuye entre mayor en el tamaño de la muestra. Es importante mencionar que cuando el tamaño de la muestra es más chico para los dos estimadores en cada una de las bases las REMC's son bastante cercanas; sin embargo, cuando el tamaño aumenta el Teorema del Límite Central empieza a ser relevante y la media muestral se vuelve más certera. Sin embargo, lo contrario le ocurre al estimador log-normal, pues sus intervalos de confianza se deterioran rápidamente mientras más grande es la muestra. Esto es porque el Teorema del Límite Central es crucial en la validación de estimadores como la media muestra, el incremento en el tamañono es garantía para otros estimadores paramétricos cuyos supuestos no aguanten este incremento.

Normalmente se diría que el Teorema del Límite Central aplica para cualquier muestra mayor a 30 observaciones, independientemente de su distribución original; sin embargo, esta regla de la práctica no aplica con distribuciones asimétricas como podría ser este el caso.

Como se puede ver cuando se tienen datos de costos hospitalarios es poco probable que estos se comporten como una distribución paramétrica. La simple construcción de estos datos nos da la pista, pues la mayoría de ellos

son la suma de costos más pequeños (costos de ambulancia, tratamientos, medicinas, etc.) lo cual significa que es una suma de distintas distribuciones. Es por esto que no podemos confiar que los datos solos nos den mucha información sobre la forma de la distribución.

Con los experimentos realizados a partir de los datos observados confirman que cuando se conoce la forma de la distribución de los datos de costos, el uso del estimador correcto es una gran ganancia en eficiencia; al igual que el uso del estimador incorrecto supone resultados totalmente engañoso. La literatura sobre riesgos recomiendan una cuidadosa modelación paramétrica de los datos para escoger el mejor estimador, aunque se recomienda un número mayor de muestra (entre 10,000-50,000 observaciones), este número es difćil de obtener en cualquier protocolo de estudios.

Por la dificultad de conseguir una muestra de datos de costos hospitalarios suficientemente grande para una modelación paramétrica adecuada, la manera de una correcta estimación con datos basados en la experiencia siguie siendo un reto.

3.3. Development and applications of a three-parameter Weibull distribution with load-dependent location and scale parameters

La distribución más utilizada en el campo de la confiabilidad es la distribución Weibull, dado que es la más apropiada para modelar datos de falla. El propósito de este artículo es evidenciar el incremento en confiabilidad al utilizar la distribución Weibull con los parámetros de escala y locacón esti-

mados por el nivel de la carga, en contraste al análisis tradicional cuando se analizan datos dependientes de cantidades de carga.
Este experimento consiste en datos sobre cinco conjuntos de cadenas industriales, que son probadas con cinco niveles de carga distintos hasta que todos sus componentes fallen. El tiempo de vida será definido como el número de ciclos completados por la cadena antes de que la falla de todos sus componentes. Los factores de interés serían el tiempo de vida, las curvas del Número Confiable de Carga (Reliability-load-number; R-L-N) y la distribución de la fuerza.

Para ambos análisis, definimos N como la variable aleatoria del tiempo de vida, por lo que la distribución de $F(N)$ es

$$F(N) = 1 - R(N) = 1 - \exp(-(\frac{N-\gamma}{\eta})^{\beta})$$

donde $R(N)$ es la función de confiabilidad, por lo que se concluye que $F(N)$ es la probabilidad de falla de los componentes. η, β y γ son los parámetros de escala, forma y localización, respectivamente.

Con el análisis tradicional de los datos, se puede concluir que para cada nivel de carga la distribución de vida se asume como una distribución de Weibull tradicional donde se ajusta la distribución teórica a la empírica. De este modo, se fija un "nivel de confiabilidad"R y se calcula una "vida confiable"N_i tal que corresponda a la distribución de vida al nivel de carga L_i (i=1,...,5), la relación entre estas variables se puede modelar con la siguiente fórmula, donde m y C son parámetros que se determinaran según los valores del experimento:

$$L^m N = C$$

Cuando cambia esta relación también cambia R, generando una serie de ecuaciones R-L-N. A su vez, estas ecuaciones se utilizaran para calcular la probabilidad de falla correspondiente a cada N_i a nivel de carga L_i; dado que es igual a la probabilidad de falla del nivel de carga S_i en el componente de vida N_i, es decir,

$$F_{Li}(N_i) = F_{Ni}(L_i)$$

donde F_{Li} es la distribución de vida a nivel de carga L_i y F_{Ni} es la distribución de fuerza al componente de vida N_i.

Sin embargo, con el análisis tradicional se encuentras varios problemas por ejemplo, la dificultad de analizar todos los datos de manera simultánea, los estimadores son menos realistas y se pueden comportar de maneras contraintuitivas, debido a que el tamaño de las muestras es pequeño y las distribuciones de vida son fijas; y las ecuaciones R-L-N y las distribuciones de fuerza no se pueden derivar directamente de las distribuciones de vida.

Dado estos problemas se busca desarrollar un modelo dentro del marco de la distribución de Weibull de tres parámetros que permita el análisis de todos los datos de falla de los componentes. Para este propósito se propone el siguiente modelo,

$$F(L, N) = 1 - \exp(-(\frac{N - \gamma(L)}{\eta(L)})^{\beta})$$

Retomando la ecuación que relaciona el componente de vida y el nivel de carga, podemos asumir que,

$$\gamma(L) = aL^{-b} \quad a > 0, b > 0$$
$$\eta(L) = cL^{-d} \quad c > 0, d > 0$$

El parámetro de forma (β) se asume que es el mismo para todos los niveles de carga.

Para estimar los parámetros de la distribución Weibull suponemos n_i componentes que se ponen a prueba con un nivel de carga L_i y para cada nivel de carga L_i tenemos componentes de vida N_{ij}. Si se eligen m niveles de carga, el número total de componentes probados es,

$$n = \sum_{i=1}^{m} n_i$$

Sea $X = (X_1, X_2, X_3, X_4, X_5) = (a, b, c, d, \beta)$ el conjunto de parámetros desconocidos, podemos estimar X minimizando la desviación máxima absoluta entre las distribuciones teóricas y empíricas.

$$f(X) = min_x(max_i, D_{n_i})$$

donde

$$D_{n_i} = max_j |F(L_i, N_{ij}) - F_n(L_i, N_{ij})|$$
$$F_n(L_i, N_{ij}) = \frac{j - 0{,}3}{n_i + 0{,}4} \quad i = i, 2, ..., m; j = 1, 2, ..., n_i$$

y $F(L_i, N_{ij})$ se calcula conforme a la distribución Weibull definida. Una vez que X ha sido estimada, las distribuciones de vida a cualquier nivel de carga, las ecuaciones R-L-N y las distribuciones de fuerza pueden ser calculadas de manera directa. Es importante notar que esta estimación fue hecha así por el tamaño de la muestra, pues si esta fuera más grande, la estimación se pudiera hacer mediante máxima verosimilitud, mínimos cuadrados, etc.

Como mencionado, las ecuaciones R-L-N se derivan de manera inmediata.

Así, tomando la distribución Weibull con tres parámetros, y las relaciones entre vida y nivel de carga tenemos,

$$(\frac{NL^d - aL^{d-b}}{c}) = -\ln R$$

Que es equivalente a

$$L^d(N - aL^{-b}) = C$$

donde

$$C = c(-\ln R)^{1/\beta}$$

esta ecuación R-N-L es distinta de la convencional pues presenta una modificación realista a la última.

Cuando N es igual al número de ciclos básicos N_b, tenemos la siguiente función de distribución

$$F(L, N_b) = 1 - R(L, N_b) = 1 - \exp(-(\frac{N_b - aL^{-b}}{cL^d})^{\beta})$$

La función de densidad de la fuerza es la primer derivada parcial de esta función de distribución con respecto a L. El resultado es igualmente distinto al tradicional, como pasó con las ecuaciones R-L-N, pues cuando el componente de vida a cada nivel de carga conforma una distribución Weibull de tres paraámetros, su distriución de fuerza no es una distribución normal, ni Gamma o cualquier otra distribución conocida.

El paraámetro de locación de la distribución de fuerza es dependiente del componente de vida. Este parámetro en la distribución de fuerza significa la mínima fuerza en la que la distribución es igual a cero,

$$F(L_0, N_b) = 0$$

En este caso, el parámetro L_0 se determina como,

$$L_0 = (\frac{a}{N_b})^{1/b}$$

Esta fórmula indica como el parámetro de locación y la distribución de fuerza están relacionadas a N_b, a y b de $\gamma(L)$. Con a y b conocidas, si N_b es más grande entonces L_0 se hace más pequeño; lo cual es un resultado intuitivo.

Una vez definido el modelo Weibull con tres parámetros, éste se utiliza para analizar los datos que se tienen. A diferencia del análisis convencional en el que se obtuvieron algunos resultados contraintuitivos, ahora el parámetro de locación decrece en tanto el nivel de carga se hace mayor, sin resultados anormales. También las curvas R-L-N se comportan mejor, garantizando que entre mayor sea el componente de confiabilidad del componente menor será el componente de fuerza correspondiente.

Después del análisis con el modelo Weibull se puede concluir que el análisis de todos los datos a diferentes niveles de carga se puede realizar de manera simultánea, las ecuaciones R-L-N y las distrubiciones de fuerza se pueden derivar directamente de la función de distribución además de resultar diferentes a aquellas obtenidas mediante métodos empíricos, lo cual sugiere que determinar las escuaciones R-L-N y las distribuciones de fuerza pueden llevar a resultados que no corresponden a la realidad.

Igualmente, dado que los parámetros de locación y escala son dependientes del nivel de carga, esto asegura que el modelo no produzca ningún resultado poco realista y más útil en la práctica dada su simplicidad.

3.4. Explaining the Gibb Sampler

Entre los métodos computacionales que han ayudado al desarrollo de la estadística tenemos al Muestreador de Gibb, que es una técnica que genera variables aleatorias indirectamente de distribuciones marginales sin tener que calcular la densidad. Este algoritmo se basa en las propiedades principales de las Cadenas de Markov. Aunque normalmente realcionado con la estadística Bayesiana, el Muestreador de Gibb también es útil en la visión clásica de la estadística.

Supongamos que tenemos una distribución conjunta $f(x, y_1, y_2, ..., y_p)$

$$f(x) = \int \cdots \int f(x, y_1, y_2, ..., y_p) dy_1, dy_2, ..., dy_p$$

Y nos interesan las características de la densidad marginal como la media o la varianza de x, con el Muestreador de Gibb podemos generar una muestra $X_1, ..., X_m \sim f(x)$ sin requerir calcular $f(x)$ directamente y obteniendo la media o la varianza con suficiente precisión.

Para explorar con detalle como funciona el Muestreador de Gibb, se toman dos variables aleatorias (X, Y) y el Muestreador de Gibb genera una muestra de $f(x)$ muestreando las distribuciones condicionales $f(x|y)$ y $f(y|x)$ que normalmente son conocidas en los modelos estadísticos. Esto se logra generando una "secuencia de Gibb"de variables aleatorias donde los valores iniciales son especificados y el resto se obtiene de manera iterativa generando así valores para

$$X'_j \sim f(x|Y'_j = y'_j)$$
$$Y'_{j+1} \sim f(y|X'_j = x'_j)$$

Esto es lo que se llama muestreo de Gibb, si $k \to \infty$ la distribución de X'_k convergerá con la verdadera distribución marginal de X ($f(x)$).

El Muestreador de Gibb se puede pensar como una implementación práctica del conocimiento de que el conocimiento de las distribuciones marginales es suficiente para conocer la distribición conjunta y aunque esto parezca claro para casos bivariados no es tan directo para los casos multivariados.

Suponemos dos variables aleatorias X, Y, de las cuales sabemos sus distribuciones condicionales $f_{X|Y}(x|y)$ y $f_{Y|X}(y|x)$. A partir de estas podríamos calcular la función marginal de X y la distribución conjunta de ambas variables, mediante el siguiente argumento:

$$f_X(x) = \int f_{XY}(x, y) \quad dy$$

donde $f_{XY}(x, y)$ aún es desconocida. Si usamos el hecho que $f_{XY}(x, y) = f_{X|Y}(x|y) f_Y(y)$ tendríamos que,

$$f_X(x) = \int f_{X|Y}(x|y) f_Y(y) \quad dy$$

asimismo, si sustituimos $f_Y(y)$,

$$\begin{aligned} f_X(x) &= \int f_{X|Y}(x|y) f_{Y|X}(y|t) f_X(t) \quad dtdy \\ &= \int [\int f_{X|Y}(x|y) f_{Y|X}(y|t) dy] f_X(t) dt \\ &= \int h(x, t) f_X(t) dt \end{aligned}$$

Esto se llama una ecuación integral con un punto fijo que tiene como solución $f_X(x)$. 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 X, Y determinan las condicionales y las marginales, no siempre las condicionales determinen de manera tan directa la distribución marginal.

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 sontiene para todas las condicionales y marginales. Pero se pueden hacer varios conjuntos de variables para construir las ecuaciones integrales con un punto fijo para calcular la distribución marginal de interés.

Supongamos que tenemos las variables aleatorias X, Y, Z y queremos la distribución $f_X(x)$, la ecuación integral de punto fijo si tomamos (Y, Z) como una sola variable, lo que resultaría en,

$$f_X(x) = \int [\int \int f_{X|YX}(x|y,z) f_{YZ|X}(y,z|t) dydz] f_X(t) dt$$

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

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

Este esquema de iteraciones nos produce una secuencia de Gibbs,

$$Y'_0, Z'_0, X'_0, Y'_1, Z'_1, X'_1, ...$$

con la propiedad de que ente más grande es la k, $X'_k = x'_k$ es un punto de la distribución marginal $f(x)$ y resolverá la ecuación integral con punto fijo.

En la estadística bayesiana, el Muestreador de Gibbs se utiliza para calcular la distribución posterior mientras que en la estadística clásica se utiliza para calcular la función de verosimilitud. Es importante mencionar que tanto el Muestreador de Gibbsy el algoritmo EM tienen en común el uso de una estructura subyacente, o variables no observables.

La utilidad del Muestreador de Gibbs es más evidente con problemas de mayor complejidad pues ahorra muchos cálculos engorrosos de una manera más elegante y con igual de precisión; además de su potencial práctico.

Bibliografía

Daley, D. and Vere-Jones, D. (2003). *An Introduction to the Theory of Point Processes: Volume I: Elementary Theory and Methods*. 2nd edition edition.

Hahn, M. G. and Zhang, G. (2012). Exchangeable random variables. *High Dimensional Probability*, 43:111.

Paik Schoenberg, F. (2000). *Introduction to Point Processes*.

Resnick, S. I. (1999). *A Probability Path*. Birkhauser.