Simulación de Tiempos al Evento mediante el Método de Inversión

Daniel Ciganda - Curso Demografía, IESTA

Introducción

La simulación de procesos demográficos, como el tiempo hasta el primer hijo, se basa en la capacidad de generar realizaciones de una variable aleatoria de tiempo al evento a partir de información empírica (por ejemplo, tasas específicas por edad). Bajo ciertos supuestos, estas tasas condicionales por edad pueden interpretarse como *funciones de riesgo constantes por tramos* (piecewise-constant hazard), lo que permite emplear métodos de inversión para generar realizaciones (tiempos al evento) de la distribución deseada.

Conceptos Básicos de Tiempo al Evento

Sea $T > 0$ una variable aleatoria continua que representa el tiempo hasta la ocurrencia de un evento (por ejemplo, el nacimiento del primer hijo). Esta variable puede caracterizarse mediante:

- **Función de Distribución:** $F(t) = P(T \leq t)$.
- **Función de Supervivencia:** $S(t) = P(T > t) = 1 - F(t)$.
- **Función de Densidad:** $f(t) = F'(t)$.
- **Función de Riesgo (hazard):** $h(t) = \frac{f(t)}{S(t)}$, la tasa instantánea a la cual ocurre el evento en t, dado que no ha ocurrido antes.
- **Riesgo Acumulado:** $H(t) = \int_0^t h(s)\, ds$, con la relación $S(t) = \exp(-H(t))$.

Ver las diapositivas del curso para definiciones más detalladas y ejemplos de algunas de estas funciones.

Interpretación de Tasas Condicionales como Función de Riesgo

Para interpretar un calendario de tasas por edad como una función de riesgo, necesitamos que las tasas reflejen la intensidad del evento en cada intervalo dado que el evento no ha ocurrido con anterioridad. Esto implica:

- **Condicionalidad en el estatus "no ocurrido":** Las tasas por edad deben calcularse sobre la población que aún no ha experimentado el evento. Por ejemplo, las tasas específicas de fecundidad por edad para el primer hijo deben construirse considerando únicamente a las mujeres que llegan a esa edad sin haber tenido un hijo antes. De esta manera, la tasa a esa edad se interpreta como la intensidad (tasa instantánea) con la que ocurre el primer nacimiento entre quienes alcanzaron esa edad sin el evento.
- **Riesgo constante en intervalos:** Si se asume que dentro de cada intervalo de edad el riesgo es constante, entonces la tasa observada en ese intervalo puede interpretarse directamente como el hazard correspondiente a ese tramo temporal.

Bajo estos supuestos, el conjunto de tasas por edad se convierte en una función de riesgo constante por tramos. La relación detallada entre las tasas observadas y la función de riesgo se explicará más adelante.

Del Riesgo Acumulado a la Generación de Tiempos

Una vez que se interpretan las tasas como función de riesgo, es posible construir el riesgo acumulado $H(t)$. Este se obtiene sumando la contribución al riesgo de cada intervalo hasta el momento t. A medida que avanzamos en el tiempo, el riesgo acumulado crece, y a partir de él podemos recuperar la función de supervivencia $S(t) = \exp(-H(t))$.

El método de inversión se basa en utilizar una variable uniforme $U \sim U(0,1)$ y relacionarla con la supervivencia:

$$U = S(T) = \exp(-H(T)) \implies H(T) = -\ln(U).$$

Podemos invertir $H(t)$ a partir de un valor u_i generado de una distribución uniforme $U(0,1)$:

$$H(t_i) = -\log(u_i) \implies H(t_i) + \log(u_i) = 0.$$

Dado que u_i es conocido, resolver esta ecuación para t_i (encontrar la raíz) nos permite obtener el tiempo al evento correspondiente a ese valor u_i.

Ejemplo: Primer Hijo

Supongamos que contamos con un calendario de tasas condicionales de fecundidad por edad para el primer hijo:

$$f_{1,x} = \frac{B_{1,x}}{L_{1,x}},$$

donde $B_{1,x}$ es el número de primeros nacimientos observados a la edad x entre quienes no tenían hijos antes, y $L_{1,x}$ es el tiempo total de exposición (en años-persona) de las mujeres que llegaron a la edad x sin el evento.
Asumiendo riesgo constante en el intervalo $[x, x+1)$, $f_{1,x}$ puede interpretarse como el hazard λ_x en ese intervalo. A partir de estas tasas (y por ende de la función de riesgo), se puede construir el riesgo acumulado y aplicar el método de inversión para simular tiempos hasta el primer hijo coherentes con las tasas observadas.

Relación con el EMV del Riesgo

Bajo el supuesto de riesgo constante, el Estimador de Máxima Verosimilitud (EMV) del hazard en un intervalo es la tasa de ocurrencia/exposición observada. Esto nos dice que:

$$\hat{\lambda} = \frac{D_x}{L_x},$$

donde D_x es el número de eventos observados y L_x el tiempo total en riesgo. El hecho de que la tasa observada sea el EMV del hazard respalda la idea de usar estas tasas como estimaciones directas del riesgo constante por tramos. Esta coherencia estadística permite asumir que la distribución del tiempo al evento subyacente puede reconstruirse a partir de las tasas, empleando el método de inversión para generar realizaciones individuales del tiempo al evento.
Ver derivación y ejemplo en las diapositivas del curso.

Procedimiento Paso a Paso para Simular Tiempos al Evento

A continuación se presenta un resumen del procedimiento para simular tiempos al evento a partir de tasas condicionales por edad:

1. **Recolección de datos:** Obtenga el número de eventos (por ejemplo, nacimientos) $B_{1,x}$ y el tiempo de exposición $L_{1,x}$ para cada intervalo de edad x en el que no haya ocurrido el evento previamente.

2. **Cálculo de tasas condicionales:** Calcule $f_{1,x} = \frac{B_{1,x}}{L_{1,x}}$.

3. **Interpretación de las tasas como hazard:** Asuma que dentro de cada intervalo el riesgo es constante. Entonces, $f_{1,x} = \lambda_x$.

4. **Construcción del riesgo acumulado:** Acumule las tasas a lo largo de los intervalos para conocer cómo crece el riesgo con la edad.

5. **Generación de números uniformes:** Genere $U_i \sim U(0,1)$ independientes para cada individuo que se desee simular.

6. **Transformación mediante el riesgo acumulado:** Para cada realización u_i de la variable uniforme $U(0,1)$, resuelva numéricamente $H(t_i) = -\ln(u_i)$ para obtener t_i. Aplicando, por ejemplo, un método numérico de búsqueda de raíces.

Siguiendo estos pasos, es posible generar realizaciones del tiempo al evento que sean coherentes con las tasas observadas y el supuesto de riesgo constante en cada tramo.