FACULTAD DE CIENCIAS ECONÓMICAS Y DE ADMINISTRACIÓN

UNIVERSIDAD DE LA REPÚBLICA URUGUAY

Curso Demografía - Licenciatura en Estadística

Docentes:

Daniel Ciganda

Facundo Morini

14ta Clase

23 de Octubre de 2025

Simulación de Procesos Demográficos

En la siguiente unidad temática vamos a trabajar sobre algunas técnicas para simular procesos demográficos y modelar los resultados de esas simulaciones. Nos vamos a enfocar en los eventos primer hijo y fallecimiento.

El primer objetivo es comprender como podemos generar tiempos de espera individuales a cada uno de estos eventos a partir de unas distribuciones de riesgo acumuladas.

Este ejercicio nos permitirá: introducir algunas ideas iniciales del análisis de supervivencia; entender mejor las tasas demográficas, su relación con la función de riesgo y la relación entre los niveles micro y macro de análisis.

Método de Inversión

El objetivo es generar realizaciones de X.

Para ello, utilizamos la **función inversa** de $F(x)$, denotada como F^{-1}

- La función de distribución acumulada $F(x)$ mapea valores de x a probabilidades y en el intervalo $[0, 1]$.
- Su inversa $F^{-1}(y)$ nos permite ir desde una probabilidad y de regreso al valor x correspondiente.

$$F^{-1}(F(x)) = x \qquad \ln(e^x) = x$$

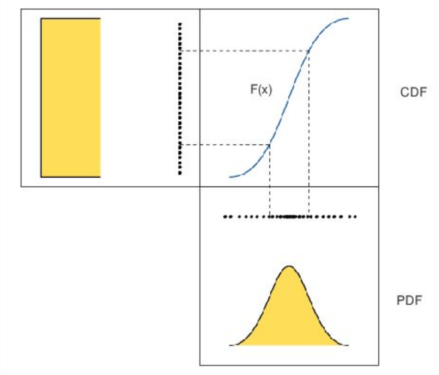


Figure 1: Método de Inversión

Método de Inversión

Para generar realizaciones de X:

1. Generamos valores aleatorios $u_1, \dots, u_n \sim U(0,1)$.
2. Aplicamos la transformación: $x_i = F^{-1}(u_i) \quad \text{para } i = 1, \dots, n$.

De esta manera, los x_i obtenidos siguen la distribución de probabilidad deseada.

Ejemplo:

Supongamos que queremos generar una variable aleatoria exponencial con parámetro λ:

- La función de distribución acumulada es $F(x) = 1 - e^{-\lambda x}$
- Su inversa es $F^{-1}(u) = -\frac{1}{\lambda} \ln(1-u)$

Entonces:

1. Generamos $u \sim U(0,1)$.
2. Calculamos $x = F^{-1}(u) = -\frac{1}{\lambda} \ln(1-u)$.

El método de inversión puede aplicarse tanto sobre la Función de Distribución como sobre la Función de Riesgo Acumulado cuando trabajamos con modelos de tiempo al evento.

Esto es relevante ya que las tasas específicas por edad de los eventos demográficos nos proporcionan la distribución de riesgo empírica bajo un supuesto de riesgo constante a intervalos.

Es decir, que podemos partir de unas tasas específicas de mortalidad para una cohorte, por ejemplo, y simular tiempos hasta la muerte usando el método de la transformación inversa.

Modelos de tiempo al evento

Análisis de sobrevida o análisis de historia de eventos

El tiempo al evento es entendido como una variable aleatoria caracterizada por una series de distribuciones:

- Función de distribución
- Densidad de probabilidad
- Función de supervivencia
- Función de riesgo

Realizaciones de la variable aleatoria “duración de la vida”

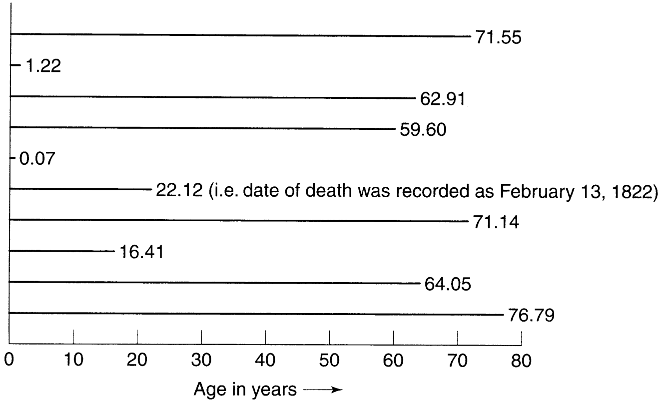


Figure 3.1 Age at death and life-lines of a hypothetical cohort of births (10 in all); date of birth: January 1, 1800

Variable Aleatoria $T > 0$, continua.

Función de Distribución:

$$F(t) = P(T \leq t)$$

Probabilidad de que el evento ocurra antes que t.

Función de Supervivencia:

$$S(t) = P(T > t) = 1 - F(t)$$

Probabilidad de que el evento no haya ocurrido hasta el tiempo t.

Función de Densidad:

$$f(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t)}{\Delta t}$$

Modelos de Tiempo al Evento (continuación)

Función de Riesgo (Hazard):

$$h(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t \mid T \geq t)}{\Delta t} = \frac{f(t)}{S(t)}$$

Es la tasa instantánea de ocurrencia del evento en el tiempo t, dado que no ha ocurrido antes de t.

Función de Riesgo Acumulado:

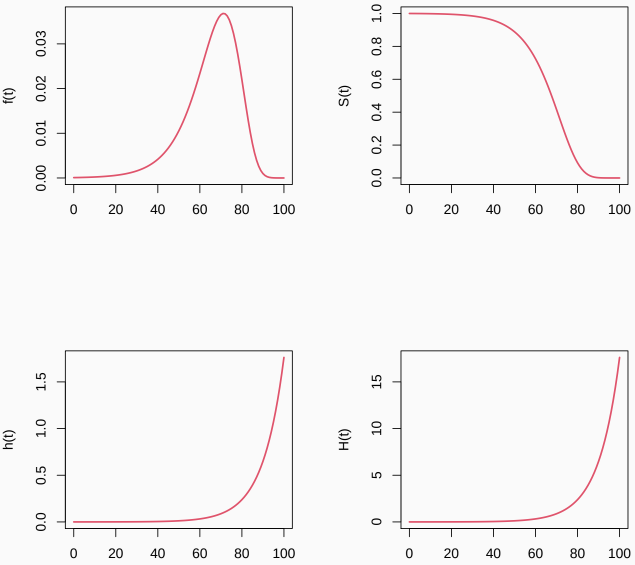
$$H(t) = \int_0^t h(s)\, ds$$

Representa la acumulación del riesgo hasta el tiempo t.

Además, podemos expresar la función de supervivencia en términos del riesgo acumulado:

$$S(t) = \exp\{-H(t)\}$$

Ej. Gompertz


f(t)
0.00
0.01
0.02
0.03
0
20
40
60
80
100
S(t)
0.0
0.2
0.4
0.6
0.8
1.0
0
20
40
60
80
100
h(t)
0.0
0.5
1.0
1.5
0
20
40
60
80
100
H(t)
0
5
10
15
0
20
40
60
80
100

$$m_x = \frac{D_x}{L_x}$$

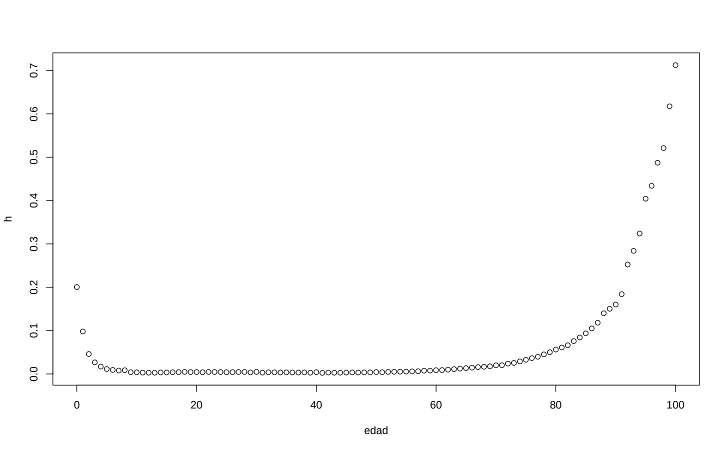


Figure 2: Tasas de Mortalidad por Edad de una Cohorte

Estimador de Máxima Verosimilitud de la Función de Riesgo

Cuando asumimos que el riesgo es constante en cada intervalo de edad, el vector de tasas de ocurrencia/exposición por edad observadas representa la función de riesgo que caracteriza a la variable aleatoria *tiempo al evento*.

Esto es así ya que, bajo el supuesto de riesgo constante, el Estimador de Máxima Verosimilitud (EMV) del riesgo de un evento es la cantidad de eventos observados sobre el tiempo de exposición al riesgo del evento, es decir la tasa de ocurrencia/exposición.

Vamos a ilustrar esta idea a través de un ejemplo.

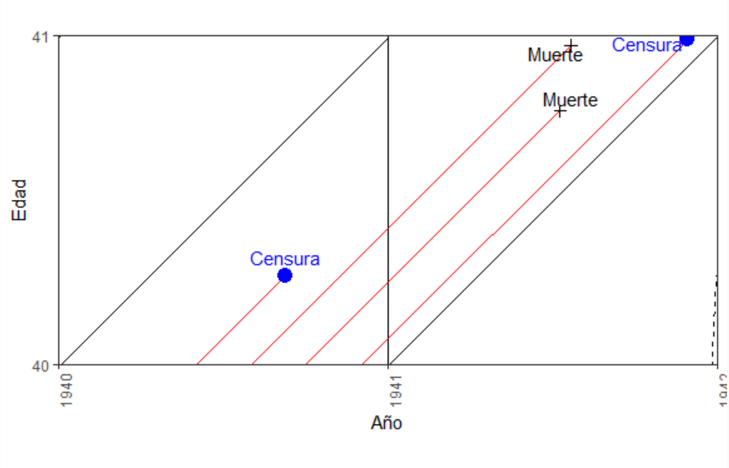


Figure 3: Seguimiento de 4 individuos en un estudio longitudinal. La primer persona abandona el estudio antes de morir o cumplir 41 años. La segunda y tercera persona fallecen antes de cumplir 41 años de edad. La cuarta observación también es censurada, pero en este caso porque la persona alcanza su 41 cumpleaños.

Ejemplo: Mortalidad de una cohorte en edades 40-41

En un estudio, se observó el tiempo hasta la muerte o censura de individuos en el intervalo de edad de 40 a 41 años:

- $x_i =$ tiempo observado para el individuo i, desde los 40 años hasta su muerte o censura en el intervalo $[40, 41)$.
- d_i es un indicador: $d_i = 1$ si el individuo murió en ese intervalo, $d_i = 0$ si el individuo fue censurado (no tenemos información sobre su muerte) en el intervalo.

Nuestro objetivo es derivar el EMV para el riesgo de muerte bajo la distribución exponencial en este contexto específico.

Supuestos del Modelo Exponencial

Bajo el supuesto de que nuestros tiempos hasta la muerte se distribuyen de acuerdo a la distribución exponencial, tenemos:

- **Función de riesgo:** $h(x) = \lambda$
- **Función de supervivencia:** $S(x) = e^{-\lambda x}$
- **Función de densidad:** $f(x) = \lambda e^{-\lambda x}$

Esto implica un riesgo constante a lo largo del tiempo en el intervalo de estudio.

Función de Verosimilitud

Para cada individuo:

- Si $d_i = 1$ (muerte): Contribución a la verosimilitud $= f(x_i) = \lambda e^{-\lambda x_i}$
- Si $d_i = 0$ (censura): Contribución a la verosimilitud $= S(x_i) = e^{-\lambda x_i}$

Por lo tanto, la función de verosimilitud para toda la muestra es:

$$L(\lambda) = \prod_{i=1}^{n} \left(\lambda e^{-\lambda x_i}\right)^{d_i} \left(e^{-\lambda x_i}\right)^{1-d_i}$$

Log-verosimilitud y EMV

Tomando el logaritmo de la función de verosimilitud:

$$\ell(\lambda) = \sum_{i=1}^{n} [d_i \log(\lambda) - \lambda x_i]$$

Derivando respecto a λ e igualando a cero:

$$\frac{d\ell}{d\lambda} = \frac{\sum d_i}{\lambda} - \sum x_i = 0$$

Despejando λ:

$$\hat{\lambda} = \frac{\sum d_i}{\sum x_i}$$

Este es el EMV del riesgo de muerte, que corresponde a:

$$\hat{\lambda} = \frac{\text{Número total de muertes}}{\text{Tiempo total de exposición al riesgo}} = \frac{D_x}{L_x} = m_x$$

Ejemplo Numérico

Supongamos que tenemos los siguientes datos para cuatro individuos entre las edades de 40 y 41 años:

Individuo	x_i (años)	d_i
1	0.3	0
2	0.9	1
3	0.7	1
4	1.0	0

Calculamos:

- **Número total de muertes:** $\sum d_i = 0 + 1 + 1 + 0 = 2$
- **Tiempo total de exposición:** $\sum x_i = 0.3 + 0.9 + 0.7 + 1.0 = 2.9$ años-persona

Aplicando el estimador de máxima verosimilitud:

$$\hat{\lambda} = \frac{\sum d_i}{\sum x_i} = \frac{2}{2.9} \approx 0.69 \text{ muertes por año-persona}$$

Esto significa que el riesgo estimado de muerte en el intervalo de 40 a 41 años es aproximadamente 0.69 por año-persona.

Simulación de Procesos Demográficos

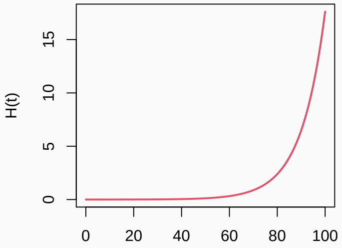
Pasos a seguir:

- Obtener una distribución de tasas específicas por edad *condicionales* del evento en cuestión
- Utilizar estas tasas como un modelo de riesgo constante a intervalos (edades) del evento
- Invertir la función de riesgo acumulado para obtener realizaciónes (números aleatorios) de nuestra distribución de partida

Método de Inversión - Función de Riesgo Acumulado

Podemos obtener realizaciones de T a partir de la función de riesgo acumulado $H(t)$ con:

$$T = H^{-1}(-\log U)$$



Pasos:

- Generando valores aleatorios $u_1, ..., u_n \sim U(0,1)$
- Transformarlos $H^{-1}(-\log u_i) = t_i$

Podemos invertir $H(t)$

$$H(t_i) = -\log u_i \qquad \Rightarrow \qquad H(t_i) + \log u_i = 0$$

En R: `uniroot()`

Aplicación del Método de Inversión en Variable de Tiempo al Evento

- Dado que $S(t) = \exp(-H(t))$, entonces:

$$U = S(t) = \exp(-H(t))$$

- Tomando logaritmos naturales:

$$\ln(U) = -H(t)$$

- Reordenando:

$$H(t) = -\ln(U)$$

- Por lo tanto:

$$T = H^{-1}(-\ln(U))$$

$$m_x = \frac{D_x}{L_x}$$

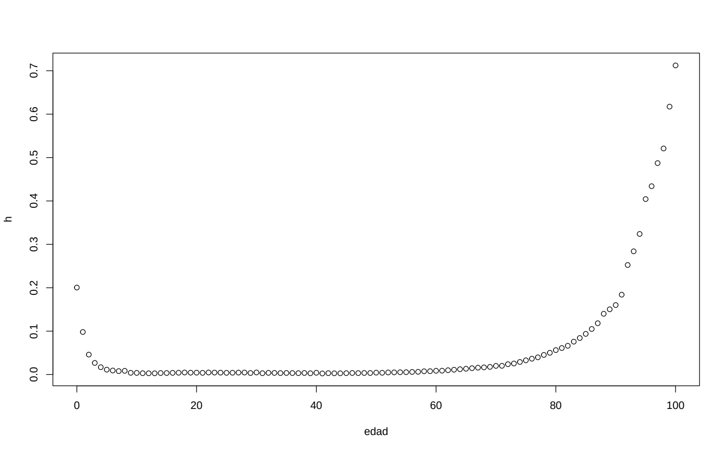


Figure 4: Tasas de Mortalidad por Edad de una Cohorte

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

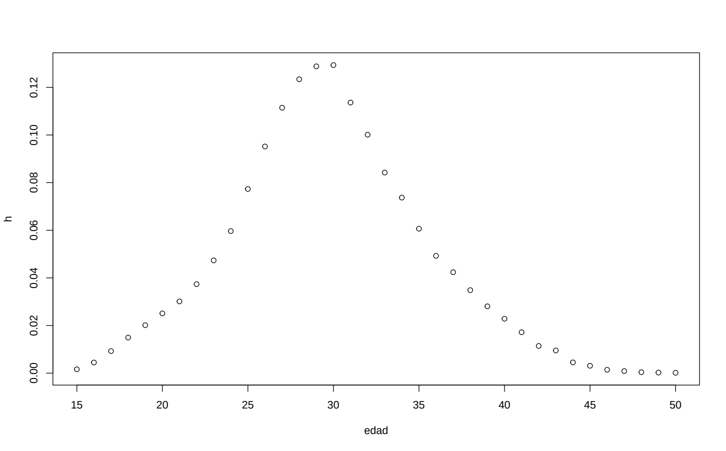


Figure 5: Tasas Condicionales de Fecundidad por Edad de una Cohorte - Primer Hijo

Labratorio en R: Generar realizaciones de el tiempo a la muerte y la transición a la maternidad a partir de distribuciones empíricas.