Tarea 4

Leobardo Enriquez Hernández & Saul Tlahuiz Tenorio

2024-04-28

4.5 Ejercicios

1. Se tiene una variable aleatoria X con función de densidad $f(x) = (1 + 2x^2)e^{-2x}$ $x > 0$

a) Determine la función de supervivencia $S(x)$

La función de supervivencia $S(t)$, tanto en el caso continuo como en el discreto, se define como la probabilidad de que un individuo sobreviva más allá del tiempo t. Para el caso continuo:

$$S(t) = \mathbb{P}(T > t) = 1 - F_T(t) = \int_t^\infty f_T(u)du$$

Para la variable aleatoria X dada, planteamos:

$$S(x) = \int_x^\infty f_X(u)du = \int_x^\infty (1 + 2u^2)e^{-2u}du = \int_x^\infty e^{-2u}du + \int_x^\infty 2u^2e^{-2u}du =$$

$$= -\frac{e^{-2u}}{2}|_x^\infty + 2[-\frac{1}{2}e^{-2u}u^2|_x^\infty - \int_x^\infty -e^{-2u}udu] = -\frac{e^{-2u}}{2}|_x^\infty - e^{-2u}u^2|_x^\infty - 2\int_x^\infty -e^{-2u}udu =$$

$$= -\frac{e^{-2u}}{2}|_x^\infty - e^{-2u}u^2|_x^\infty - 2[-\frac{1}{4}(-2e^{-2u}u|_x^\infty - \int_x^\infty -2e^{-2u}(1))]$$

$$= -\frac{e^{-2u}}{2}|_x^\infty - e^{-2u}u^2|_x^\infty - 2[-\frac{1}{4}(-2e^{-2u}u|_x^\infty - (e^{-2u}|_x^\infty)] =$$

$$= -\frac{e^{-2u}}{2}|_x^\infty - e^{-2u}u^2|_x^\infty - e^{-2u}u|_x^\infty - \frac{1}{2}e^{-2u}|_x^\infty = -0 + \frac{e^{-2x}}{2} - 0 + e^{-2x}x^2 - 0 + e^{-2x}x - 0 + \frac{1}{2}e^{-2x} =$$

$$= \frac{e^{-2x}}{2} + e^{-2x}x^2 + e^{-2x}x + \frac{1}{2}e^{-2x} = e^{-2x}(x^2 + x + 1)$$

b) Determine la función de riesgo $h(x)$

La función de riesgo $h(t)$ (hazard function), también llamada **tasa de falla condicional** (en el análisis de confiabilidad) o *tasa de mortalidad* (en demografía), se define como la probabilidad de falla durante un intervalo de tiempo muy pequeño suponiendo que el individuo ha sobrevivido hasta el inicio del intervalo (aunque en la definición de $h(t)$ se tenga explícitamente la palabra "probabilidad", hay que tener en claro que esta función no es una función de probabilidad, si no tal cual una **tasa**, ya que la acumulación de esta puede dar valores superiores a 1).

$$h(t) = \frac{f(t)}{S(t)} = \frac{(1 + 2x^2)e^{-2x}}{e^{-2x}(x^2 + x + 1)} = \frac{1 + 2x^2}{x^2 + x + 1}$$

c) Determine la función de media residual $mr(x)$

Para individuos de edad x este parámetro, denotado por $mr(x)$, mide la esperanza de vida residual; esto es, "la esperanza de vida que les queda a partir de la edad x".

Se define el caso contínuo como:

$$mr(x) = \frac{\int_x^\infty S(t)dt}{S(x)} = \frac{\int_x^\infty e^{-2t}(t^2+t+1)dt}{e^{-2x}(x^2+x+1)} = \frac{\int_x^\infty t^2e^{-2t}dt + \int_x^\infty te^{-2t}dt + \int_x^\infty e^{-2t}dt}{e^{-2x}(x^2+x+1)} =$$

$$\frac{(-\frac{1}{2}e^{-2t}t^2 - \frac{1}{2}e^{-2t}t - \frac{1}{4}e^{-2t})|_x^\infty + (\frac{1}{4}(-2e^{-2t}t - e^{-2t})|_x^\infty) - (\frac{1}{2}e^{-2t}|_x^\infty)}{e^{-2x}(x^2+x+1)} =$$

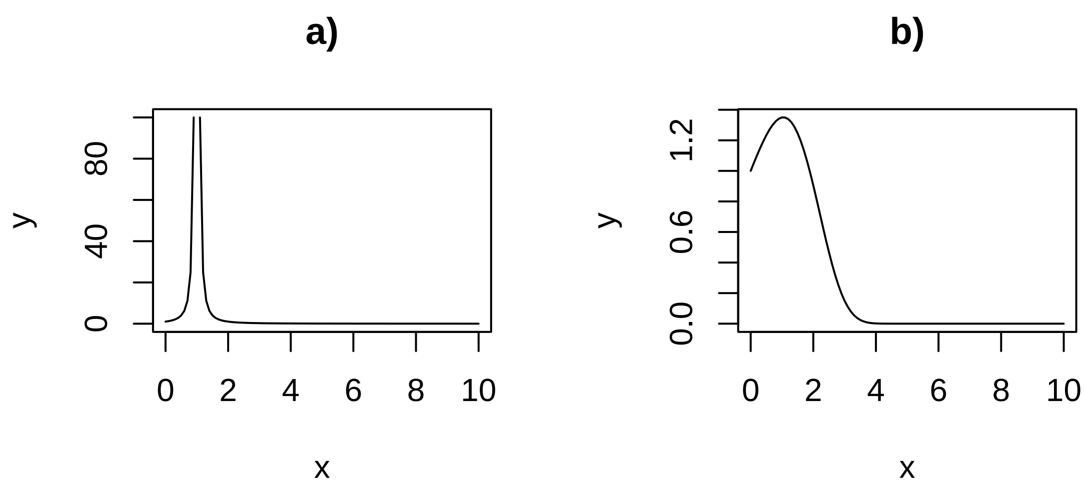
$$= \frac{(-\frac{1}{2}e^{-2t}t^2 - \frac{1}{2}e^{-2t}t - \frac{1}{4}e^{-2t} - \frac{1}{2}e^{-2t}t - \frac{1}{4}e^{-2t} - \frac{1}{2}e^{-2t})|_x^\infty}{e^{-2x}(x^2+x+1)} = \frac{(-\frac{1}{2}e^{-2t}t^2 - e^{-2t}t - e^{-2t})|_x^\infty}{e^{-2x}(x^2+x+1)} =$$

$$= \frac{(-0-0-0) - (-\frac{1}{2}e^{-2x}x^2 - e^{-2x}x - e^{-2x})}{e^{-2x}(x^2+x+1)} = \frac{\frac{1}{2}e^{-2x}x^2 + e^{-2x}x + e^{-2x}}{e^{-2x}(x^2+x+1)} = \frac{1}{2}$$

2. De los siguientes ¿cuál(es) no podría servir como un modelo de supervivencia?

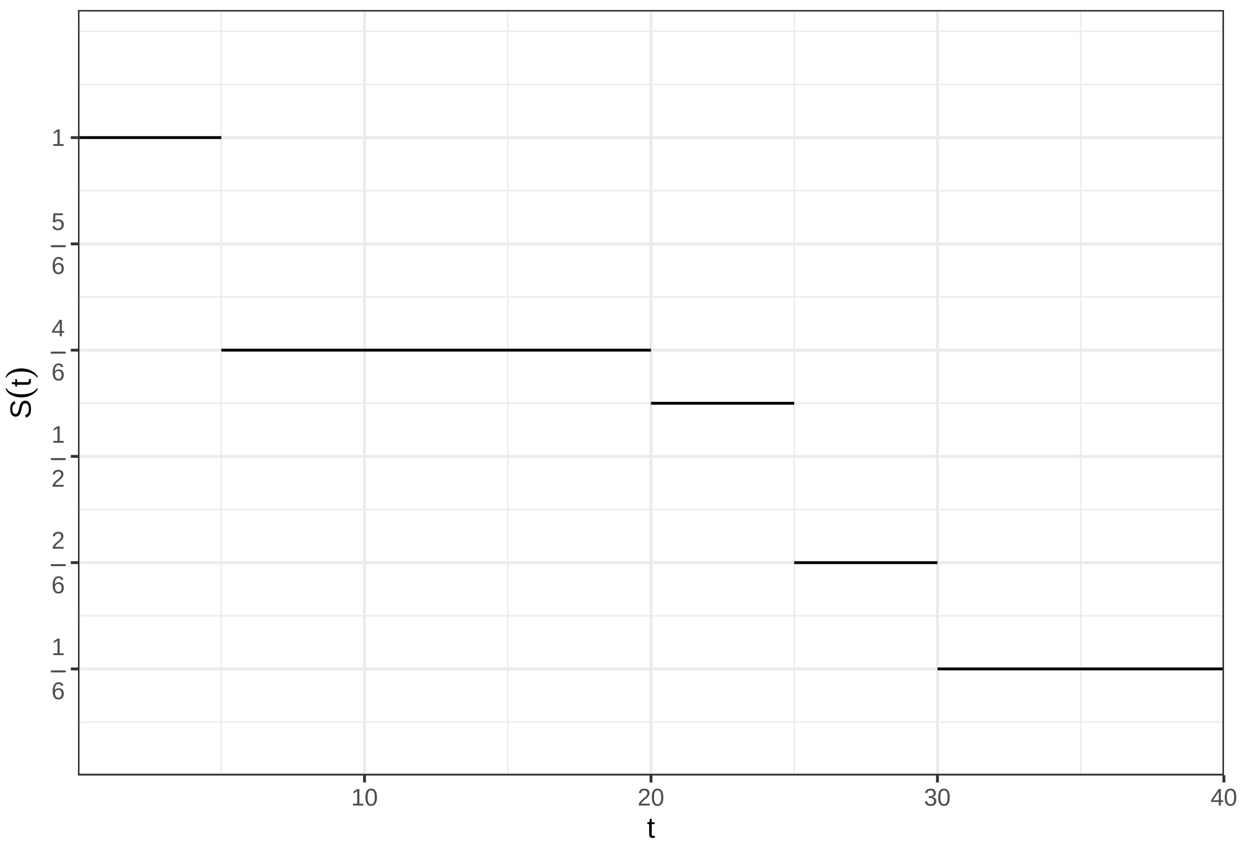
a) $S(t) = e^{\left[t-0.7\left(2^t-1\right)\right]}$

b) $S(t) = (1-t)^{-2}$



Observamos que ninguno de los dos representa un modelo de supervivencia.

3. Suponga que la curva de supervivencia estimada de un conjunto de datos es la siguiente:



a) Calcule la esperanza de vida residual para $t = 15$.

$$mr(15) = \frac{\sum_{k=15}^{\infty} S(u_k)}{S(15)} = \frac{4/6 + 7/12 + 2/6 + 1/6}{4/6} = \frac{21/12}{4/6} = \frac{21}{8} = 2.625$$

b) Calcule el cuantíl correspondiente a una supervivencia del 80%

$$t_p = inf\{t : S(t) \leq 1 - p\}$$
$$t_{0.8} = inf\{t : S(t) \leq 0.2\} = 30$$

4. Suponga que Z tiene la siguiente función de riesgo $h(z) = \frac{1}{100-z}$, $0 < z < 100$

a) Determine la función de supervivencia $S(x)$

La función de riesgo acumulado:

$$H(t) = \int_0^t h(u)du = \int_0^t \frac{1}{100-u}du = (-ln|u - 100|)|_0^t = -ln|t - 100| + ln| - 100| = ln(100) - ln|t - 100|$$

$$S(t) = e^{-H(t)} = e^{ln(100) - ln|t-100|} = e^{ln(100)}e^{-ln|t-100|} = 100(-|t - 100|) = -100|t - 100|$$

b) Determine la función de distribución $F(x)$

$$F(t) = 1 - S(t) = 1 - (-100|t - 100|) = 1 + 100|t - 100| = 1 - 100(t - 100) = -100t + 1001$$

c) Determine la función de probabilidad $f(x)$

$$f(t) = h(t)S(t) = \frac{1}{100-t}(-100|t-100|) = \frac{-100|t-100|}{100-t} = \frac{-100|100-t|}{100-t} = -100$$

$$f(t) = -\frac{dS(t)}{dt} = -[\frac{d}{dt}(-100(-(t-100)))] = -[\frac{d}{dt}(100t-10000)] = -100$$

6.8 Ejercicios

1. El tiempo de vida en años de una batería de marca paso sigue una distribución *Pareto* centrada con parámetros $\theta = 4$, $\lambda = 5$

Recordemos que la función de densidad de una variable aleatoria X que se distribuye Pareto(θ, λ), con $\theta > 0$, $\lambda > 0$ y $x > 0$ es: $f(x;\theta,\lambda) = \frac{\theta\lambda^{\theta}}{x^{\theta+1}}$

Considerando $\theta = 4$, $\lambda = 5$, se tiene: $f(x) = \frac{(4)(5)^4}{(x)^{4+1}} = \frac{3125}{x^5}$

Por otro lado, para esta distribución, la función de riesgo es $h(x) = \frac{\theta}{x} = \frac{4}{x}$ la cual es una tasa de falla; la función de supervivencia es $S(x) = \frac{\lambda^{\theta}}{x^{\theta}} = \frac{5^4}{x^4} = \frac{625}{x^4}$ la cual es la probabilidad de sobrevivencia más allá de un tiempo t; y la media $E(X) = \frac{\theta\lambda}{\theta-1}$ para $\theta > 1$.

a) ¿Cuál es la probabilidad de que la batería funcione por lo menos 10 años?

$$S(10) = P(T > 10) = 1 - F_T(10) = \int_{10}^{\infty} f(x)dx = \int_{10}^{\infty} \frac{3125}{x^5}dx = -\frac{3125}{4x^4}|_{10}^{\infty} = -0 + \frac{3125}{40000} = 0.078125$$

La probabilidad de que la batería funcione por lo menos 10 años es de aproximadamente 7.8%.

Notemos que la esperanza de vida residual (la que queda después de 10 años), es de $mr(10) = \frac{\int_{10}^{\infty} S(t)dt}{S(10)} = \frac{\int_{10}^{\infty} \frac{625}{t^4}dt}{0.078125} = \frac{\frac{-625}{3t^3}|_{10}^{\infty}}{0.078125} = \frac{(-0+0.20833)}{0.078125} = 2.66$

b) ¿Cuál es el tiempo de vida medio de la batería?

$$E(X) = \frac{\theta\lambda}{\theta-1} = \frac{4*5}{4-1} = \frac{20}{3} = 6.66$$

c) Encontrar el tiempo de reemplazo t_r para el cual 90% de las baterías no hayan fallado.

Notemos que esto es el tiempo de reemplazo t_r para el cual 10% de las baterías hayan fallado.

$$S(t_{0.10}) = 1 - 0.10 = 0.90$$
$$S(t) = \frac{625}{t^4} = 0.9 \longrightarrow t = \sqrt[4]{\frac{625}{9}} = 8.33$$

Por lo tanto el tiempo de reemplazo buscado es de 8.33

2. El tiempo de vida (en años) X de una bacteria que se encuentra en el lago "Gran Ojo" sigue una distribución "mezcla". Calcule la probabilidad de que el tiempo de vida de la bacteria sea menor a 200 años.

a) Con probabilidad 0.8, X tiene una distribución *Pareto* centrada con parámetros $\alpha = 2$ y $\theta = 100$.

$$S(x) = \frac{\theta^{\alpha}}{x^{\alpha}} = \frac{100^2}{x^2} = \frac{1000}{x^2}, \qquad f(x) = \frac{\alpha\theta^{\alpha}}{x^{\alpha+1}} = \frac{(2)100^2}{x^3} = \frac{20000}{x^3}$$

$$P(T < 200) = 1 - S(200) = 1 - \int_{200}^{\infty} f_T(u)du = 1 - \int_{200}^{\infty} \frac{20000}{u^3}du$$

$$= 1 - 20000[-\frac{1}{2u^2}]|_{200}^{\infty} = 1 - 20000[-0 + \frac{1}{80000}] = 0.75$$

De donde tenemos que $0.8 * 0.75 = 0.6$

b) Con probabilidad 0.2, X tiene una distribución $Pareto$ centrada con parámetros $\alpha = 4$ y $\theta = 3000$.

$$S(x) = \frac{\theta^\alpha}{x^\alpha} = \frac{3000^4}{x^4}, \qquad f(x) = \frac{\alpha\theta^\alpha}{x^{\alpha+1}} = \frac{(4)3000^4}{x^5}$$

$$P(T < 200) = 1 - S(200) = 1 - \int_{200}^{\infty} f_T(u)du = 1 - \int_{200}^{\infty} \frac{(4)3000^4}{u^5}du$$

$$= 1 - (4)(3000)^4[-\frac{1}{4u^4}]|_{200}^{\infty} = 1 - (4)(3000)^4[-0 + \frac{1}{(4)(200)^4}]$$

3. Considere una distribución Weibull cuya función de de supervivencia está dada por $S(x) = exp(-\lambda x^\alpha)$ con $\lambda > 0$ y $\alpha > 0$. Encuentre la función de verosimilitud bajo los siguientes casos:

a) Datos truncados por la izquierda (y_{l_i}, x_i), $y_{l_i} \leq x_i$, $i = 1, \ldots, n$ donde y_{l_i} es un tiempo de truncamiento por la izquierda (simplifique la expresión).

Recordemos que si se tienen los tiempos de fallo T_i con censura, provenientes de una distribución Weibull, tenemos que $f(t) = \lambda\gamma(\lambda t)^{\gamma-1}e^{-(\lambda t)^\gamma}$, $S(t) = e^{-(\lambda t)^\gamma}$ y $h(t) = \lambda\gamma(\lambda t)^{\gamma-1}$, y que para el truncamiento por la izquierda $T_i|T_i > u_i$, tenemos la siguiente contribución a la función de verosimilitud: $P(T_i|T_i > u_i) = \frac{f(t_i)}{S(u_i)}$

Recordando que la función de verosimilitud en general se puede componer como:

$$\prod_{i\in D} f(t_i) \prod_{i\in R} S(C_i) \prod_{i\in L}(1 - S(C_i)) \prod_{i\in I}[S(L_i) - S(R_i)]$$

donde

- D: Conjunto de tiempos de fallo.
- R: Conjunto de observaciones censuradas por la derecha.
- L: Conjunto de observaciones censuradas por la izquierda.
- I: Conjunto de observaciones censuradas por intervalo.

Donde, si hay datos truncados, se sustituye $f(t_i)$ por $\frac{f(t_i)}{S(u_i)}$ y $S(C_i)$ por $\frac{S(C_i)}{S(u_i)}$.

Entonces tenemos en el caso particular de tiempos de fallo con truncamiento por la izquierda $T_i|T_i \geq u_i$, con u_i valor de truncamiento, las observaciones serán (u_i, t_i, δ_i) con $t_i \geq u_i$ tiempo de fallo y δ_i indicador de censura. Por lo que la función de verosimilitud es:

$$\prod_{i=1}^{n} \left\{\frac{f(t_i)}{S(u_i)}\right\}^{\delta_i} \left\{\frac{S(t_i)}{S(u_i)}\right\}^{1-\delta_i} = \prod_{i=1}^{n} \{h(t_i)\}^{\delta_i} \left\{\frac{S(t_i)}{S(u_i)}\right\}$$

Sustituyendo, tenemos la función de verosimilitud:

$$\prod_{i=1}^{n} \left[\lambda\gamma(\lambda t_i)^{\gamma-1}\right]^{\delta_i} \left[\frac{e^{-(\lambda t_i)^\gamma}}{e^{-(\lambda u_i)^\gamma}}\right]$$

b) Datos censurados por intervalos (L_i, R_i), $i = 1, \ldots, n$.

Recordemos que la contribución a la función de verosimilitud para observaciones con censura por intervalo $Li < T_i < R_i$ es $P(Li < T_i < R_i) = S(Li) - S(R_i)$.

Entonces, la función de verosimilitud queda como:

$$\prod_{i \in D} f(t_i) \prod_{i \in I} [S(L_i) - S(R_i)]$$

$$\prod_{i=1}^{n} \left[\lambda\gamma(\lambda t_i)^{\gamma-1} e^{-(\lambda t_i)^{\gamma}}\right]^{\delta_i} \left[e^{-(\lambda l_i)^{\gamma}} - e^{-(\lambda u_i)^{\gamma}}\right]^{1-\delta_i}$$

c) Datos doblemente truncados (y_{l_i}, x_i, y_{r_i}), donde $y_{l_i} \leq x_i \leq y_{r_i}$, $i = 1, \ldots, n$ (simplifique la expresión).

Recordemos que las contribuciones a la verosimilitud de los datos truncados por la derecha $T_i|T_i \leq v_i$ es $P(T_i|T_i \leq v_i) = \frac{f(t_i)}{1-S(v_i)}$, y para los datos truncados por la izquierda $T_i|T_i > u_i$ es $P(T_i|T_i > u_i) = \frac{f(t_i)}{S(u_i)}$. De donde:

Para el caso del truncamiento a la derecha, consideramos tiempos de fallo T_i tal que $T_i \leq v_i$ para que sea observado, las observaciones serán (t_i, v_i), $\forall i = 1, ..., n$, entonces:

$$\prod_{i=1}^{n} \left\{\frac{f(t_i)}{1 - S(v_i)}\right\} = \prod_{i=1}^{n} \mathbb{P}(T_i|T_i < v_i)$$

Para el caso del truncamiento a la izquierda, consideramos tiempos de fallo T_i tal que $T_i \geq u_i$ para ser observado, las observaciones serán (u_i, t_i, δ_i), $\forall i = 1, ..., n$, con $t_i \geq u_i$ tiempo de fallo y δ_i indicador de censura, entonces:

$$\prod_{i=1}^{n} \left\{\frac{f(t_i)}{S(u_i)}\right\}^{\delta_i} \left\{\frac{S(t_i)}{S(u_i)}\right\}^{1-\delta_i} = \prod_{i=1}^{n} \{h(t_i)\}^{\delta_i} \left\{\frac{S(t_i)}{S(u_i)}\right\}$$

De donde tenemos la siguiente expresión para la verosimilitud doble truncada.

$$\prod_{i=1}^{n} \left\{\frac{f(t_i)}{1 - S(v_i)}\right\} \cdot \left\{\frac{f(t_i)}{S(u_i)}\right\}^{\delta_i} \left\{\frac{S(t_i)}{S(u_i)}\right\}^{1-\delta_i}$$

$$\prod_{i=1}^{n} \left\{\frac{\lambda\gamma(\lambda t_i)^{\gamma-1} e^{-(\lambda t_i)^{\gamma}}}{1 - e^{-(\lambda v_i)^{\gamma}}}\right\} \cdot \left\{\frac{\lambda\gamma(\lambda t_i)^{\gamma-1} e^{-(\lambda t_i)^{\gamma}}}{e^{-(\lambda u_i)^{\gamma}}}\right\}^{\delta_i} \left\{\frac{e^{-(\lambda t_i)^{\gamma}}}{e^{-(\lambda u_i)^{\gamma}}}\right\}^{1-\delta_i}$$

$$\prod_{i=1}^{n} \left\{\frac{[e^{-2(\lambda t_i)^{\gamma}}]}{1 - e^{-(\lambda v_i)^{\gamma}}}\right\} \cdot \left\{\frac{[\lambda\gamma(\lambda t_i)^{\gamma-1}]}{e^{-(\lambda u_i)^{\gamma}}}\right\}^{\delta_i+1} \left\{\frac{1}{1}\right\} \left\{\frac{e^{-(\lambda u_i)^{\gamma}}}{1}\right\}^{\delta_i}$$

$$\prod_{i=1}^{n} \left\{\frac{[e^{-2(\lambda t_i)^{\gamma}}]}{1 - e^{-(\lambda v_i)^{\gamma}}}\right\} \cdot \left\{\frac{[\lambda\gamma(\lambda t_i)^{\gamma-1}]^{\delta_i+1}}{e^{-(\lambda u_i)^{\gamma}}}\right\}$$

d) n = 4 pacientes cuya edad de ocurrencia del evento son los siguientes intervalos $(90, 120]$, $(110, 115]$, $(80, 100]$, $(70, 75]$, sujetos a la condición de entrada *edad* ≥ 50.