Confiabilidad y Análisis de supervivencia

Tarea 3

Fecha de entrega: Martes 3 de octubre, 2017

Instrucciones: Para los ejercicios 4, 5, 6, 7a envie su código a mi correo en UN SOLO archivo con nombre "tarea3.*apellido materno*.R". Asegúrese de divir la respuesta a cada pregunta, comentar con cuidado lo que cada función hace, sus entradas y salidas y pasos importantes. Incluya un EJEMPLO de como se corre (o el código usado para obtener la solución particular).

1. Obtenga el estimador máximo verosímil de la media del tiempo de falla T con los siguientes datos

   3.60, 4.01, 4.28, 4.29, 4.86, 5.01, 5.08, 5.37, 6.27, 6.27, 6.72, 8.00+, 8.00+, 8.00+, 8.00+

   que son datos censurados a la derecha y truncados a la izquierda en tres. Considere que la distribución de los tiempos de falla es Gama($\alpha = 2, \beta$) con fdp

$$f(x) = \frac{1}{\beta^a \Gamma(a)} x^{a-1} \exp\left[-\frac{x}{\beta}\right].$$

2. Suponga que T es el tiempo continuo de vida (o falla) con funciones de supervivencia $S(t)$, riesgo $h(t)$, riesgo acumulado $H(t)$ y vida media residual $m(t)$. Demuestre

   a. Demuestre que "$h(t)$ no decreciente $\forall t \geq 0$ (IFR)" implica "$H(t)/t$ no decreciente $\forall t \geq 0$" implica "$m(t) \leq m(0), \forall t \geq 0$"
   b. Demuestre que "$h(t)$ no decreciente $\forall t \geq 0$" implica "$m(t)$ es decreciente en $t \geq 0$" implica "$m(t) \leq m(0), \forall t \geq 0$"
   c. Mostrar que $h'(t) < 0 \leftrightarrow f'(t) < 0$
   d. Pruebe entonces que una mezcla discreta de tiempos de falla con tasa de falla decreciente (DFR) es una DFR.
   e. Muestre que una mezcla discreta de distribuciones exponenciales entonces tiene DFR.
   f. Muestre que una mezcla discreta de IFR no necesariamente es una IFR.

3. Consideremos el estimador Nelson-Åalen en una situación especial. Supongamos que no tods los individuos comparten la misma tasa de riesgo $h(t)$ sino que para el i-ésimo individuo, ésta es de forma multiplicativa e igual a $h_i(t) = \alpha(t)\mu_i(t)$, donde $\alpha(t)$ es la mortalidad o falla común a todas las peronas, y $\mu_i(t)$ es la tasa de riesgo propia para el i-ésimo individuo y es conocida. Esta tasa μ_i puede describir la diferente exposición al riesgo del evento y puede ser conocida individualmente o por grupos de individuos. Por ejemplo, la diferencia en exposición entre los grupos de hombres/mujeres, trabajadores en mina/no en minas, carro funcionando en tierra adentro/carro funcionando en la costa, etc. Bajo este esquema, ¿Cuál es el estimador Nelson-Åalen para $H(t)$? Comente su resultado.

4. Por ser el estimador Kaplan-Meier un EMV, es consistente y converge a una distribución gaussiana. Usando la aproximación gaussiana. Obtenga intervalos de confianza para $S(t)$ para los siguientes datos:

   10.5+ 8.5+ 3.5+ 16.5+ 18.5 14.5+ 5.5 1.5+ 12.5+ 2.5+ 14.5+ 9.5+ 8.5 15.5 11.5+ 12.5+ 4.5+ 3.5+ 26.5+ 4.5+

a. Sin usar ninguna transformación
b. Usando la trasformación logit
c. Usando la transformación log-log.

Grafique todos los intervalos en una misma gráfica para compararlos. Comente.

5. Construya su propia función en R para obtener el estimador Kaplan-Meier y usando la base de datos "azt" (del paquete KMsurv) compárela contra el estimador Kaplan-Meier que se obtiene con la función survfit del paquete "survival".

6. Construya la función en R para obtener intervalos de confianza (a nivel α) para $S(t)$.

7. Obtenga

   a. Una función (o funciones) en R con la que se pueda hacer lo siguiente

      i. Simule n datos de una distribución de vida exponencial (con media λ) o Gamma(α, β) o Gumbel(α, β) o Log Normal(μ, σ^2). Cualquiera de éstas a escoger.
      ii. Usando esos datos obtenga el estimador Nelson-Åalen y Kaplan-Meier (puede auxiliarse de las funciones en el paquete "survival").
      iii. Que repita los dos pasos anteriores m veces y grafique las estimaciones obtenidas para $H(t)$ en una gráfica y $S(t)$ en otra.
      iv. Que a las gráficas anteriore sobre ponga las verdadera (aproximada tal vez con métodos númericos) funcionones de riesgo acumulada y de supervivencia.

   b. Con la función (o funciones) anterior obtenga $m = 100$ simulaciones de datos de una población de tamaño 100 donde los individuos tienen distribución de falla exponencial con media 1.5. Considere que los datos son censurados todos al tiempo 40 y usando las m simulaciones obtenga las m estimaciones N-A y K-M. Presente las gráficas de éstos comparando con los verdaderos $H(t)$ y $S(t)$.