

Análisis de Supervivencia

Introducción

Breve introducción histórica

Iniciemos con un breve bosquejo histórico sobre el desarrollo del análisis de supervivencia.

El análisis de supervivencia es una de las disciplinas estadísticas más antiguas y tiene sus raíces en la demografía y la ciencia actuarial del siglo XVII.

La metodología básica de *tabla de vida*, en la terminología moderna, equivale a la estimación de la *función de supervivencia* (uno menos la función de distribución) para *tiempos de vida* registrados en distintos momentos de ingreso a un estudio, o con *truncamiento por la izquierda*, y *censura por la derecha*. Esto era conocido antes 1700, además de modelos paramétricos explícitos, desde la aproximación lineal de De Moivre (1725), ejemplos posteriores debidos a Lambert y las influyentes propuestas del siglo XIX de Gompertz (1825) y Makeham (1837), que modelan la función de riesgo como bc^x y $a + bc^x$, respectivamente.

Motivado por la controversia sobre la inoculación de la viruela, Daniel Bernoulli (1766) sentó las bases de la teoría de los riesgos competitivos. Al realizar el cálculo del número esperado de muertes (cuántas muertes habría habido en una población de estudio, si se aplican un conjunto determinado de tasas de muerte o bajo la presencia de un conjunto de diversas causas de muerte) también se remonta al siglo XVIII.

Entre los importantes avances metodológicos del siglo XIX, además de los modelos paramétricos de análisis de supervivencia, mencionados anteriormente, fue la manipulación gráfica simultánea del tiempo calendario y la edad en el Diagrama de Lexis (1875).

Dos temas muy importantes en el análisis de supervivencia moderno, puede remontarse a principios del siglo XX de las matemáticas actuariales: *Los Modelos multi estado*, en el caso particular de los seguros por discapacidad, y la estimación no paramétrica en tiempo continuo de la función de supervivencia, considerando condiciones del tiempo de vida con truncamiento por la izquierda y censura por la derecha.

Hasta este momento, el análisis de supervivencia no había incorporado ningún componente

de teoría estadística a sus desarrollos. Un característico escepticismo sobre “el valor de las tablas de vida en investigación estadística” fue expresado por Greenwood (1922) en la revista de la Royal Statistical Society, y en una aparición especial de Westergaard (1925) en *Biometrika* sobre “los problemas modernos en las estadísticas vitales” argumentando ambos que no tenían referencias sobre la variabilidad de la muestra o variabilidad muestral (errores estándar). Esto debido al hecho de que ambos autores eran en realidad estadísticos pioneros en el análisis de supervivencia: Westergaard (1882) por haber derivado lo que llamaríamos el error estándar de la razón de mortalidad; y Greenwood (1926) por su famosa expresión de “los errores de muestreo de las tablas de supervivencia”.

La tabla Actuarial de Vida y el estimador Kaplan-Meier

En la mitad del siglo XX, se presentaron a la comunidad médica-estadística, estas dos técnicas demográficas y actuariales perfectamente establecidas, en encuestas influyentes como las de Berkson y Gage (1950) y Cutler y Ederer (1958). En este enfoque, el tiempo se discretiza (por ejemplo, intervalos de un año), y la probabilidad de supervivencia en un intervalo, es el producto de estas probabilidades en los intervalos de tiempo anteriores. El problema de esta aproximación radica en el hecho de realizar justamente este proceso de discretización del tiempo que es intrínsecamente continuo. El importantísimo resultado dado por Kaplan y Meier (El estimador Kaplan-Meier) (1958), la fascinante génesis de este estimador (del que Breslow realiza una crónica (1991)), en principio, eliminado la necesidad de estas aproximaciones en la situación de estadística médica usual, donde todos los tiempos de supervivencia y censura se conocen con precisión. El estimador desarrollado por Kaplan y Meier consistió en reducir los intervalos de observación (hacer que la longitud de ellos tendiera a cero) para incluir como máximo una observación por intervalo (En otras palabras, hacer un proceso límite con la tabla actuarial de vida). Aunque pasado por alto por muchos autores, Kaplan y Meier también formalizaron el milenar manejo de la entrada en distintos momentos al estudio, a través del ajuste necesario para los conjunto en riesgos, el conjunto de individuos vivos y bajo observación en un tiempo específico.

Entre las variaciones en el modelo actuarial, es conveniente mencionar dos.

Harris et al. (1950) anticiparon en su generalización mucho del trabajo reciente que se hace, por ejemplo, en los estudios de supervivencia de SIDA, en los que los estimadores de la tabla

de vida dependen de los tiempos de muertes y censuras registrados en grandes intervalos irregulares de tiempo.

Ederer et al. (1962) desarrolló una “tasa de supervivencia relativa” como la relación (cociente) de la tasa de supervivencia observada en un grupo de pacientes y la tasa de supervivencia esperada en un grupo similar a estos pacientes. “Haciendo un símil, de esta manera, con la larga tradición de la comparación de valores observados con valores esperados” . En la nomenclatura actual, la comparación de un grupo tratado vs. un grupo control.

Modelos paramétricos de supervivencia

Los modelos de supervivencia paramétricos estaban bien establecidos en la ciencia actuarial y la demografía, pero nunca han dominado dentro los usos médicos del análisis de supervivencia. Sin embargo, en los años 1950 y 1960 hubo importantes contribuciones a la teoría estadística de análisis de supervivencia basada en modelos paramétricos simples. Un ejemplo es el enfoque de máxima verosimilitud por Boag (1949) para *un modelo de cura* (cure rate model) asumiendo la vida eterna (supervivencia del grupo de cura) con probabilidad, c (proporción del grupo de cura), y tiempos de supervivencia con distribución lognormal.

La distribución exponencial fue utilizada por Littell (1952), cuando realizó la comparación de la tasa de supervivencia anual dada por el método actuarial y calculada a través de máxima verosimilitud. Armitage (1959) utilizó esta distribución en su estudio comparativo de pruebas de dos muestras para los ensayos clínicos con entrada escalonda, y también fue utilizada por Feigl y Zelen (1965) en su modelo para tiempos de vida (sin censura) cuya intención era que dependieran linealmente de covariables de los sujetos; generalizada, posteriormente, para datos censurados por Zippin y Armitage (1966).

Cox (1972) revolucionó el análisis de supervivencia a través de su modelo de regresión semi-paramétrico para la función de riesgo (modelo de riesgos proporcionales o Modelo de Cox), que depende arbitrariamente del tiempo y paramétricamente de las covariables.

Modelos multi estado

Las formas tradicionales en situaciones actuariales y demográficas de modelar varios tipos de falla o muerte al mismo tiempo, puede formalizarse a través de los Procesos de Markov a tiem-

po continuo con espacio de estados finito. Una importante documentación muy influyente en este sentido, fue propuesta por Fijar y Neyman (1951), quienes estudiaron la recuperación, recaída y la muerte (además de la censura) en lo que ahora se denomina comúnmente como un modelo de enfermedad o muerte con *riesgos competitivos*. Chiang (1968), por ejemplo, en su monografía de 1968, documenta ampliamente los modelos estocásticos subyacentes a esta metodología, y Sverdrup (1965), en un importante documento, realiza su estudio estadístico sistemático. Estos modelos tienen intensidades de transición constantes, aunque la subdivisión de tiempo en intervalos permite una metodología de tiempo agrupados del tipo de tabla de vida actuarial, tan cuidadosamente documentado por Hoem (1976).

Características particulares del análisis de supervivencia

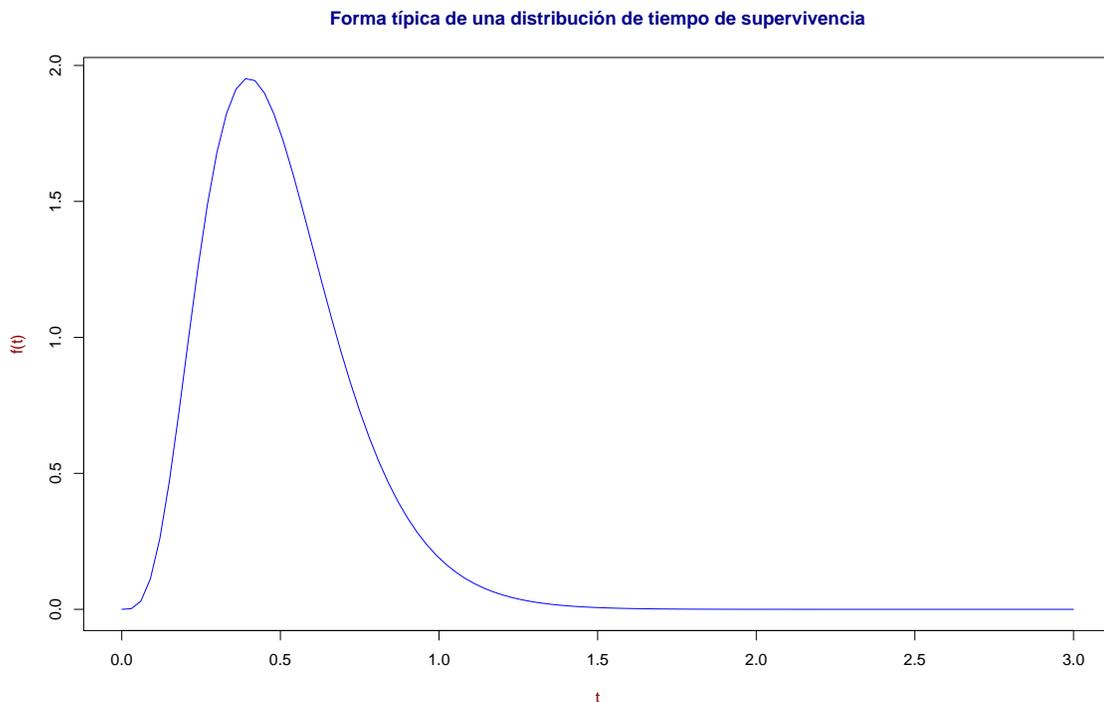
- **Análisis de Supervivencia** (conocido como Análisis de Sobrevida, en medicina) es el conjunto de métodos estadísticos para analizar datos de supervivencia.
- **Datos de supervivencia.** Este tipo de datos se generan cuando se tiene interés en estudiar el tiempo que transcurre entre un *evento inicial* (que determina la inclusión en el estudio de un individuo) y un *evento final* (genéricamente denominado como falla), que define el término del estudio para cada individuo. Al tiempo transcurrido entre estos dos eventos se le denomina *tiempo de falla, tiempo de supervivencia o tiempo de muerte*.
- Los estudios de supervivencia pertenecen a los llamados *estudios longitudinales*, en los que los individuos se siguen a través del tiempo; en este caso, el seguimiento se hace desde la entrada al estudio de un individuo hasta la ocurrencia de su falla.
- Los estudios de supervivencia son muy comunes en el área *biomédica*, en los que la falla puede ser la muerte de un paciente o la recuperación de una enfermedad. No obstante, pueden presentarse en múltiples áreas del conocimiento, como:
 - *Industria*: Tiempo de falla de un componente de alguna máquina o aparato eléctrico (Confiabilidad).
 - *Economía*: Tiempo de desempleo o de empleo de una persona económicamente activa.
 - *Demografía*: Tiempo de duración del matrimonio o tiempo para embarazarse.
 - *Psicología*: Tiempo para realizar alguna tarea en una evaluación psicológica
 - *Educación*: Tiempo para concluir un proceso educativo. Tiempo para titularse de una licenciatura.

- *Biología*: Estudios de supervivencia de especies.
- *Actuaría*: Tiempo que permanece vivo un individuo asegurado con una póliza de vida, etc., etc.

Características especiales de los datos de supervivencia

Una pregunta que debemos hacernos es ¿qué hace diferente al análisis de supervivencia? En otras palabras, ¿por qué constituye un área particular de la estadística?.

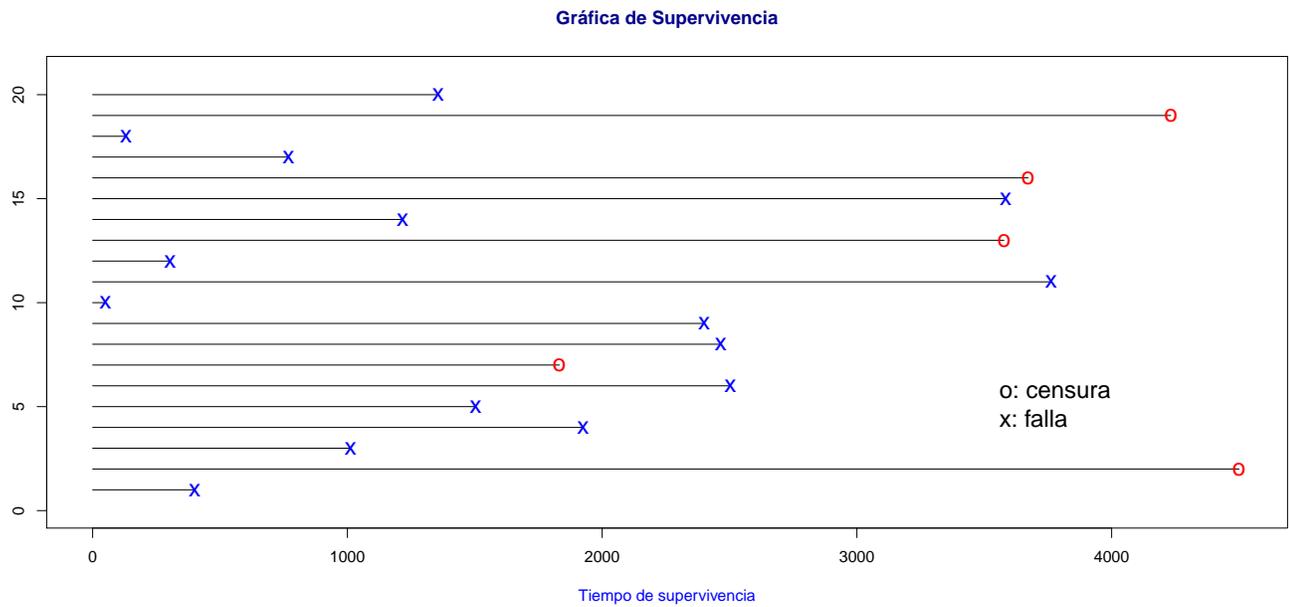
- El tiempo de supervivencia es una variable aleatoria positiva, generalmente con sesgo positivo (cola derecha larga), por lo que, por ejemplo, está descartado el uso de la *distribución normal* como un modelo plausible. No obstante, esta característica de positividad del tiempo de supervivencia, podría eliminarse si utilizamos una transformación de esta variable, e.g., una transformación logarítmica; sin embargo, es más conveniente trabajar con la variable en su escala natural, y utilizar modelos que consideren esta característica positiva de la misma.



- *Observaciones censuradas*. Indudablemente una de las características más importantes de los datos de supervivencia.

En estudios longitudinales puede ocurrir que algún individuo lo abandone antes de que le haya ocurrido la falla, por lo que sólo se tendrá información parcial (*observación censurada*) sobre su tiempo de falla. En otras áreas de la estadística, estos datos cen-

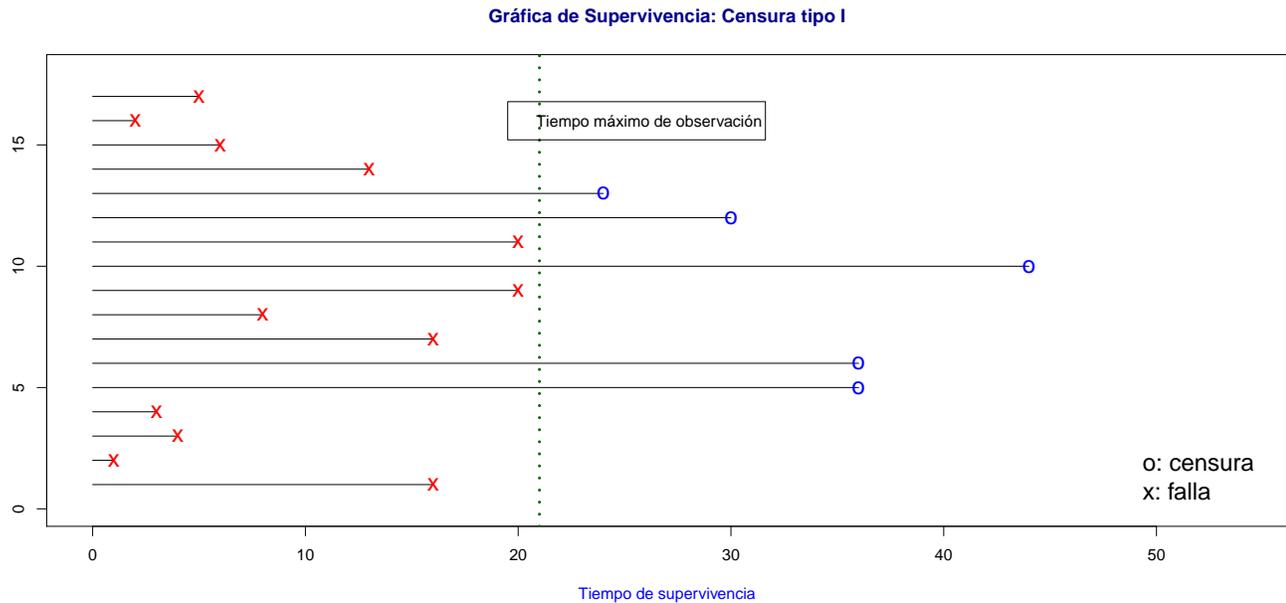
surados suelen considerarse como *datos faltantes*. La presencia de datos censurados es el mayor problema técnico en el desarrollo de métodos estadísticos dentro del análisis de supervivencia, tanto que, incluso métodos tan simples como el análisis descriptivo, requieren procesos más complejos que los habituales.



Tipos de censura

En los estudios de supervivencia, la censura puede presentarse de diversas formas, por lo que recibe distintas clasificaciones.

- *Censura tipo I.* En muchos estudios, el investigador se ve en la necesidad de fijar un tiempo máximo de observación de los individuos para que presenten la falla, usualmente, debido a razones de presupuesto o de diseño del estudio. Los individuos que al término de este periodo no hayan presentado la falla, constituyen las observaciones censuradas.



- *Censura tipo II.* En este caso, el investigador decide prolongar la observación de los individuos en el estudio hasta que ocurran k fallas de n posibles ($k \leq n$). Una razón común para determinar el número de fallas que deben observarse, es la potencia requerida en el estudio. Los individuos que no experimentaron la falla al completarse estas primeras k , representan observaciones censuradas. En estas dos situaciones, la censura está controlada por el investigador.
- *Censura aleatoria.* Este tipo de censura ocurre sin ningún control del investigador. Las censuras pueden presentarse por abandono del estudio de un individuo, con la

correspondiente pérdida de su seguimiento, o muerte por alguna causa que no está relacionada con el evento de interés.

Otro tipo de clasificación de las censuras

En los tres casos anteriores, lo que se sabe es que, de haber ocurrido la falla (*no observada*), ésta se presentaría después del tiempo de *censura observado*, por lo que se conoce como *censura por la derecha*. Este tipo es el más común en estudios biomédicos.

Poco común en supervivencia, la *censura por la izquierda* ocurre cuando antes de ingresar al estudio, el individuo ha presentado la falla, por lo que se sabe que su tiempo de falla *no observado* es menor que el tiempo de censura observado.

En algunos estudios, la observación de los sujetos no se realiza de manera continua y los periodos de observación pueden ser muy largos entre dos observaciones consecutivas. En esta situación, es común que, respecto al tiempo de falla *no observado* del sujeto, sólo se sabe que ocurrió en el intervalo determinado por estos dos periodos de revisión. Esta observación se dice que presenta *censura por intervalo*.

Actualmente, los métodos de supervivencia contemplan, además de la posibilidad de observaciones censuradas, también observaciones truncadas por la derecha o por la izquierda.

Funciones involucradas en el análisis de supervivencia

Recordemos que intentaremos modelar una variable aleatoria \mathbf{T} no negativa, que representa el tiempo de falla de los individuos en el estudio. Para analizar esta variable se cuenta con cuatro funciones básicas, a saber

- **La función de supervivencia**

Esta función representa la probabilidad de que un individuo sobreviva (no experimente la falla) a un tiempo determinado, t . En símbolos se escribe

$$S(t) = \mathbb{P}(T > t)$$

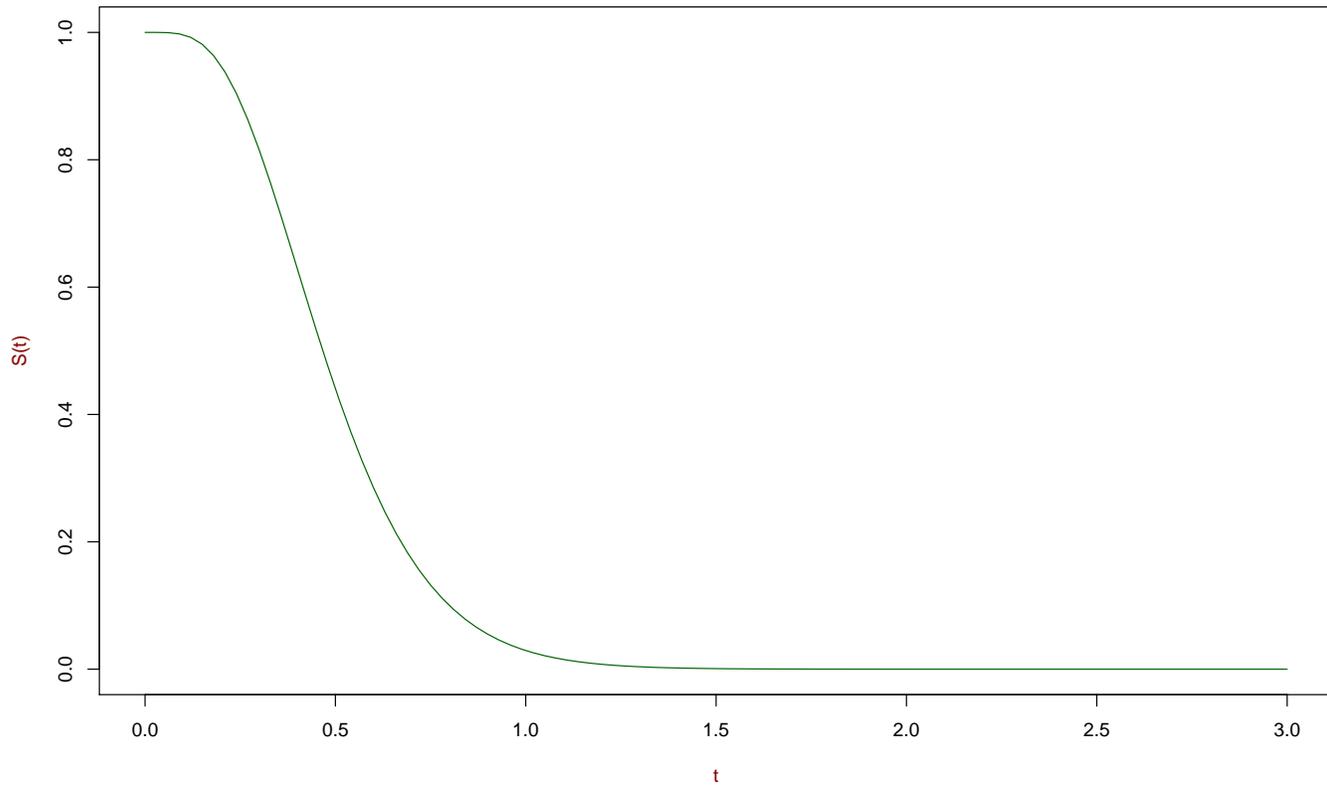
De su definición es claro que $S(t) = 1 - F(t)$, con $F(\cdot)$ la función de distribución acumulada. Además se tiene

$$S(0) = 1, \quad S(t) = 0 \text{ si } t \rightarrow \infty \quad S(t_1) \geq S(t_2) \text{ si } t_1 < t_2$$

Si \mathbf{T} es continua, con función de densidad $f(t)$, entonces

$$S(t) = \mathbb{P}(T > t) = \int_t^{\infty} f(u) du$$

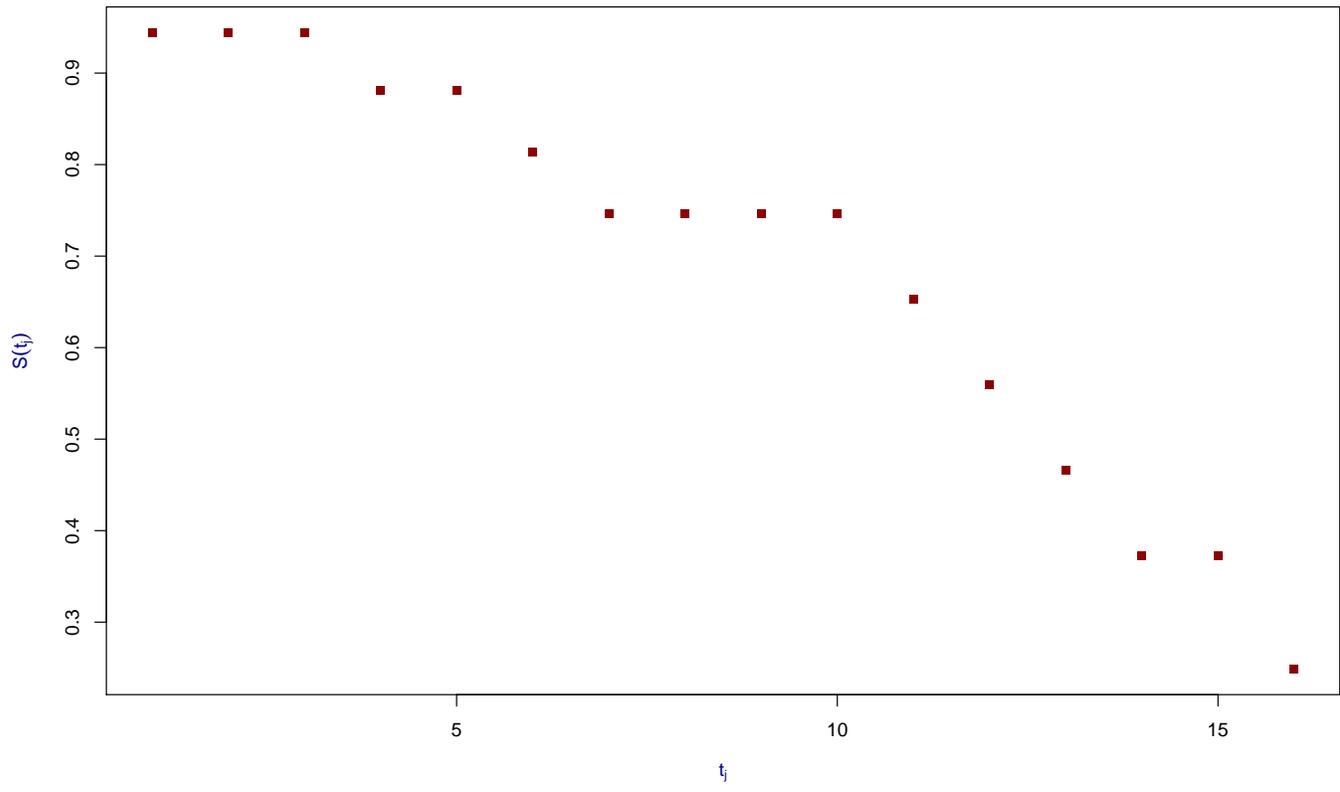
Forma típica de una función de supervivencia



Si \mathbf{T} es discreta con función de masa de probabilidad $f(t_j) = \mathbb{P}(T = t_j)$, $j = 1, 2, \dots$ y $t_1 < t_2 < \dots$. Entonces, su función de supervivencia es

$$S(t) = \mathbb{P}(T \geq t) = \sum_{t_j \geq t} f(t_j)$$

Supervivencia discreta



■ Función de riesgo

Esta función determina la tasa instantánea de falla al tiempo $\mathbf{T} = t$, dado que el individuo ha sobrevivido un instante antes de t , y se define como

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{\mathbb{P}(t < T \leq t + \Delta t | T > t)}{\Delta t}$$

en el caso discreto

$$h(t_j) = \mathbb{P}(T = t_j | T \geq t_j), \quad j = 1, 2, \dots$$

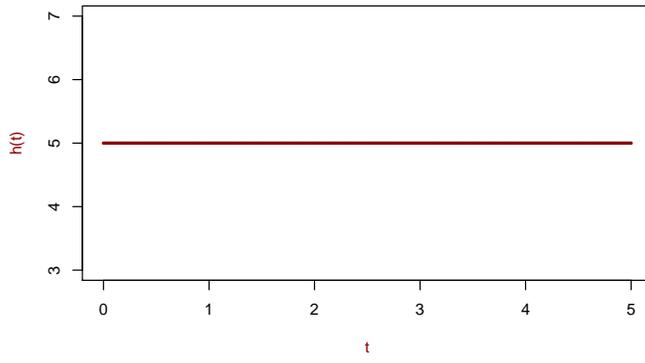
En el caso continuo, podemos observar que si $\Delta t \approx 0$, entonces

$$h(t) * \Delta t \approx \mathbb{P}(t < T \leq t + \Delta t | T \geq t)$$

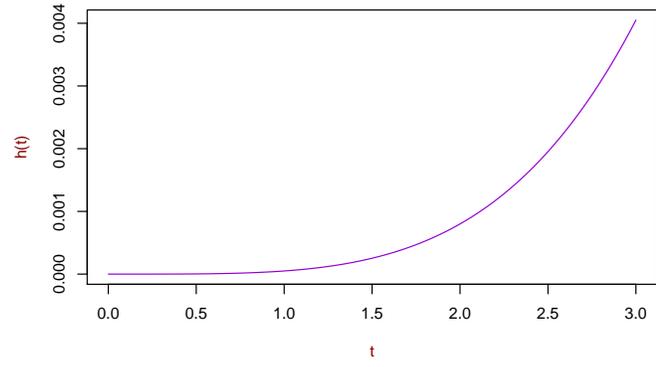
que es, aproximadamente, la probabilidad de fallar en el intervalo $(t, t + \Delta t]$, dado que se está vivo al tiempo t . Podemos interpretarla como una medida de intensidad de falla al tiempo t , o una medida de falla potencial al tiempo t .

Una característica importante de esta función, es que puede servir de guía para proponer un modelo paramétrico para el tiempo de supervivencia. Lo común en muchas áreas, es que el investigador tenga alguna idea de cómo se comporta el riesgo de falla en la población bajo estudio. Este conocimiento se puede intentar asociar con la forma de alguna función de riesgo paramétrica.

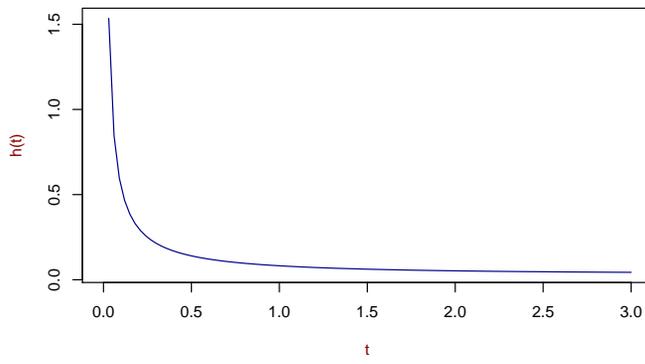
Riesgo exponencial



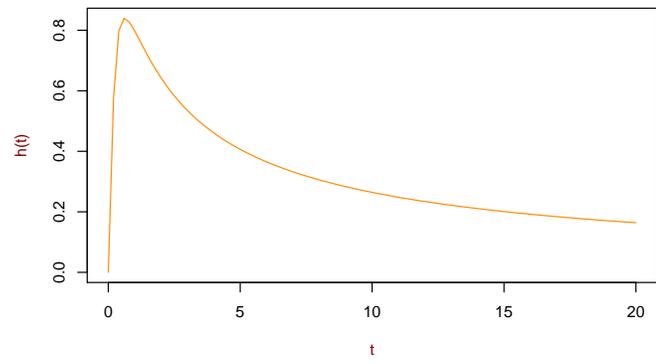
Riesgo Weibull



Riesgo Weibull



Riesgo Log-normal



- **Función de riesgo acumulado**

En el caso de que \mathbf{T} sea continua, esta función se define como

$$H(t) = \int_0^t h(u)du$$

y en el caso discreto

$$H(t) = \sum_{t_j \leq t} h(t_j)$$

que se prefiere calcular de forma alternativa como

$$H(t) \approx - \sum_{t_j \leq t} \log [1 - h(t_j)]$$

con

$$H(0) = 0, \quad \text{si } t_1 > t_2 \Rightarrow H(t_1) \geq H(t_2) \quad \text{es no decreciente}$$

- **Función de vida media residual**

La última de las funciones para realizar el análisis de supervivencia es *la función de vida media residual*, denotada como $mrl(t)$ (por sus siglas en inglés: *mean residual life*). Para los sujetos de edad t , esta función mide la esperanza de vida restante, o el tiempo esperado de vida, después de t , hasta la ocurrencia de la falla. Por lo que se define como

$$mrl(t) = \mathbb{E}(T - t | T > t)$$

que, para el caso continuo, se puede escribir como

$$\begin{aligned} \mathbb{E}(T - t | T > t) &= \int_0^{\infty} (u - t) f(u | u > t) du \\ &= \int_0^{\infty} (u - t) \frac{f(u)}{S(t)} I_{(t, \infty)}(u) du \\ &= \int_t^{\infty} (u - t) \frac{f(u)}{S(t)} du \end{aligned}$$

Tarea 1

Demostrar que

- $H(t) \approx - \sum_{t_j \leq t} \log [1 - h(t_j)]$

- $mrl(t) = \int_t^\infty \frac{S(u)}{S(t)} du$

Cómo interpretan esta última expresión? Qué es $mrl(0)$?

Algunas relaciones entre las funciones para el análisis de supervivencia

Las relaciones de $S(t)$ con el resto de funciones, en el caso continuo

$$S(t) = \int_t^{\infty} f(u)du \Rightarrow f(t) = -\frac{dS(t)}{dt}$$

Ahora, iniciemos con la definición de $h(t)$

$$\begin{aligned} h(t) &= \lim_{\Delta t \rightarrow 0} \frac{\mathbb{P}(t < T \leq t + \Delta t | T > t)}{\Delta t} \\ &= \lim_{\Delta t \rightarrow 0} \frac{\mathbb{P}(t < T \leq t + \Delta t)}{\Delta t * \mathbb{P}(T > t)} \\ &= \lim_{\Delta t \rightarrow 0} \frac{F(t + \Delta t) - F(t)}{\Delta t} \frac{1}{S(t)} \\ &= - \lim_{\Delta t \rightarrow 0} \frac{1 - F(t + \Delta t) - (1 - F(t))}{\Delta t} \frac{1}{S(t)} \\ &= - \lim_{\Delta t \rightarrow 0} \frac{S(t + \Delta t) - S(t)}{\Delta t} \frac{1}{S(t)} \\ &= -\frac{S'(t)}{S(t)} \\ &= -\frac{d \log S(t)}{dt} \end{aligned}$$

La relación entre $mrl(t)$ y $S(t)$, fue propuesta como tarea.

De $h(t)$ con el resto. Partiendo de

$$h(t) = -\frac{d \log S(t)}{dt} \Rightarrow S(t) = \exp \left\{ - \int_0^t h(u)du \right\}$$

Para $f(t)$

$$h(t) = \frac{f(t)}{S(t)} \Rightarrow f(t) = h(t)S(t) \Rightarrow f(t) = h(t)\exp\left\{-\int_0^t h(u)du\right\}$$

De $H(t)$ con otras funciones

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

Caso discreto

De $S(t)$ con el resto.

$$\begin{aligned} S(t) = \mathbb{P}(T \geq t) &= \sum_{t_j \geq t} f(t_j) \Rightarrow S(t_j) = \sum_{t_i \geq t_j} f(t_i) \text{ y } S(t_{j-1}) = \sum_{t_i \geq t_{j-1}} f(t_i) \\ &\Rightarrow f(t_j) = -\{S(t_j) - S(t_{j-1})\} = S(t_{j-1}) - S(t_j) \end{aligned}$$

Para la relación con la función de riesgo discreta, tenemos

$$h(t_j) = \mathbb{P}(T = t_j | T \geq t_j) = \frac{\mathbb{P}(T = t_j)}{\mathbb{P}(T \geq t_j)} = \frac{f(t_j)}{S(t_{j-1})} = \frac{S(t_{j-1}) - S(t_j)}{S(t_{j-1})} = 1 - \frac{S(t_j)}{S(t_{j-1})}$$

De $h(t)$ con las demás

De la igualdad anterior, tenemos que $S(t_j) = \{1 - h(t_j)\} S(t_{j-1})$. Desarrollando esta expresión para $j = 1, 2, \dots$, tenemos

$$\begin{aligned} S(t_1) &= [1 - h(t_1)] S(t_0) = 1 - h(t_1) \\ S(t_2) &= [1 - h(t_2)] S(t_1) = [1 - h(t_2)] [1 - h(t_1)] \\ S(t_3) &= [1 - h(t_3)] S(t_2) = [1 - h(t_3)] [1 - h(t_2)] [1 - h(t_1)] \\ &\vdots \end{aligned}$$

Finalmente obtenemos

$$S(t) = \prod_{t_j \leq t} [1 - h(t_j)]$$

Para $f(t)$

$$\begin{aligned} h(t_j) &= \frac{f(t_j)}{S(t_{j-1})} \Rightarrow f(t_j) = h(t_j) S(t_{j-1}) \\ \Rightarrow f(t_j) &= h(t_j) \prod_{k=1}^{j-1} [1 - h(t_k)] = \frac{h(t_j)}{1 - h(t_j)} \prod_{k=1}^j [1 - h(t_k)] \end{aligned}$$

Análisis de supervivencia: Resumen

- Estudios longitudinales
- **T**: Tiempo de supervivencia entre el inicio del estudio y el momento de la falla. ($T \geq 0$).
- Característica particular, *Censura*: Información parcial del tiempo de falla
- Por la derecha: $T_i \in [C_i, \infty)$ T_i : No observado C_i : Observado
- Por la izquierda: $T_i \in (0, C_i]$ T_i : No observado C_i : Observado
- Por intervalo: $T_i \in [L_i, R_i]$ T_i : No observado (L: Left, R: Right)
- **Controlada por el investigador**
- *Tipo I*: Tiempo máximo de observación t_c , fijo

Muestra observada: (Y_1, Y_2, \dots, Y_n) con

$$Y_i = \begin{cases} T_i, & \text{si } T_i \leq t_c \\ t_c, & \text{si } T_i > t_c \end{cases}$$

Aquí el número de fallas es aleatorio.

- *Tipo II*: Se observa hasta que ocurran k fallas de n posibles. $k \leq n$

Muestra observada: $(Y_{(1)}, Y_{(2)}, \dots, Y_{(n)})$ con

$$Y_{(i)} = \begin{cases} T_{(i)}, & \text{si } i \leq k \\ T_{(k)}, & \text{si } i > k \end{cases}$$

Número de fallas fijo, tiempo de duración del estudio aleatorio.

- **No controlada por el investigador**

- *Censura aleatoria*: Los individuos pueden abandonar el estudio, o fallar por alguna causa que no es de interés

T_i : Tiempo de falla C_i : Tiempo de censura.

Muestra observada: (Y_1, Y_2, \dots, Y_n) con $Y_i = \min(T_i, C_i)$

Indicadora de falla

$$\delta_i = \begin{cases} 1, & \text{si } T_i \leq C_i \text{ (Falla)} \\ 0, & \text{si } C_i < T_i \text{ (Censura)} \end{cases}$$

La información la proporciona la pareja (Y_i, δ_i) .

- **Funciones para el análisis de supervivencia**

- *Función de supervivencia*

$$S(t) = \mathbb{P}(T > t)$$

- *Función de riesgo*

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{\mathbb{P}(t < T \leq t + \Delta t | T > t)}{\Delta t}$$

- *Función de densidad*

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

- *Función de riesgo acumulado*

$$H(t) = \int_0^t h(u)du$$

- *Función de vida media residual*

$$mrl(t) = \int_t^\infty (u - t) \frac{f(u)}{S(t)} du$$

Modelos paramétricos para el tiempo de supervivencia

En esencia, cualquier distribución de una variable aleatoria con soporte en los reales no negativos, puede ser un modelo para el tiempo de supervivencia. No obstante, existen un número reducido de distribuciones de uso común para este fin.

Distribución exponencial

En este caso, la función de densidad es bastante conocida

$$f(t, \theta) = \theta e^{-\theta t} \quad \theta > 0 \quad t \geq 0 \quad o$$
$$f(t, \theta) = \frac{1}{\theta} e^{-\frac{t}{\theta}}$$

Función de supervivencia

$$S(t) = \int_t^{\infty} \theta e^{-\theta u} du = -e^{-\theta u} \Big|_t^{\infty} = e^{-\theta t}$$

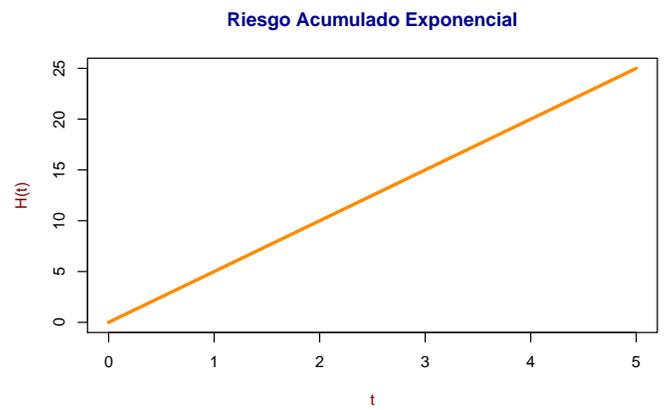
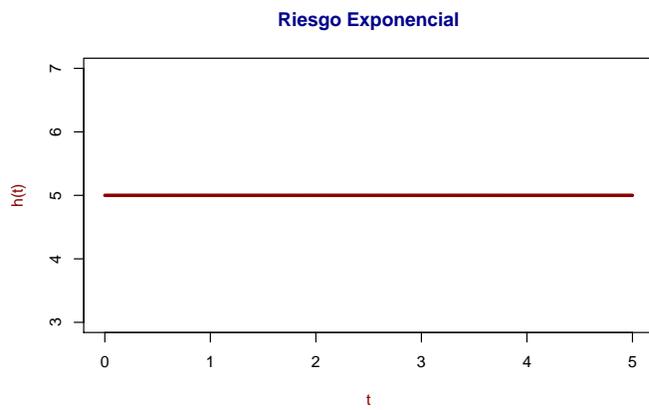
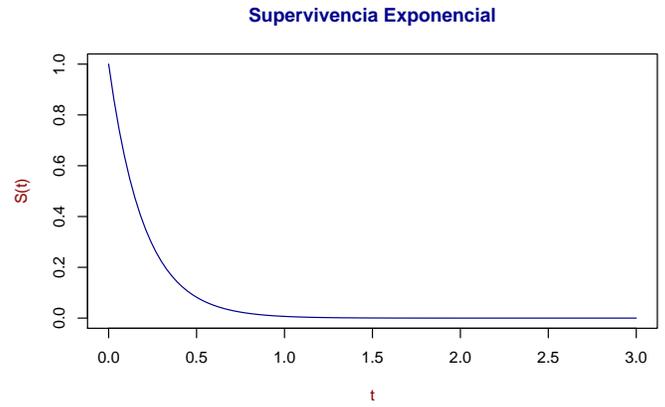
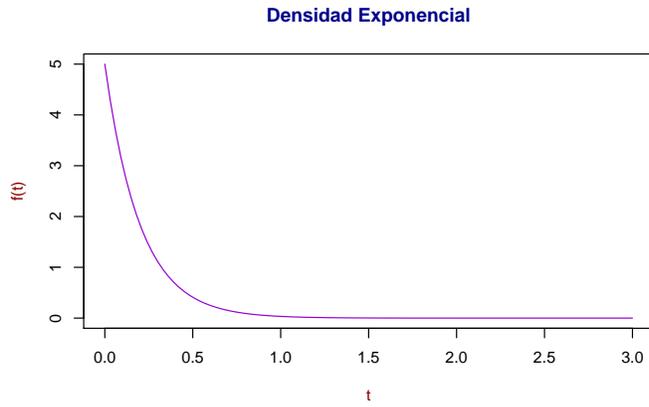
y *función de riesgo*

$$h(t) = \frac{f(t)}{S(t)} = \frac{\theta e^{-\theta t}}{e^{-\theta t}} = \theta \quad ! \text{ constante !}$$

El hecho de que la función de riesgo sea constante, implica que el riesgo de presentar la falla no cambia a lo largo del tiempo, i.e., se tiene igual riesgo de fallar en cualquier momento del estudio. Esta característica de la función de riesgo de la exponencial, está relacionada con la propiedad de *pérdida de memoria* de esta distribución, ya que la probabilidad de fallar en un intervalo $[t, t + \Delta t]$, no depende de la supervivencia previa a este intervalo. Esta función es de uso común en *confiabilidad*.

Función de riesgo acumulado

$$H(t) = \int_0^t h(u) du = \int_0^t \theta du = \theta t$$



Características numéricas

Media y varianza

$$\mathbb{E}(T) = \frac{1}{\theta}, \quad \mathbb{V}(T) = \frac{1}{\theta^2}$$

Como la exponencial es una función sesgada a la derecha, la media no es una buena medida de tendencia central, en su lugar es mejor utilizar la mediana, definida como

$$t = \frac{\log(2)}{\theta}$$

Distribución Weibull

La distribución Weibull es, probablemente, el modelo paramétrico más común para modelar datos de supervivencia. Es ampliamente utilizado en aplicaciones biomédicas.

Existen diversas parametrizaciones de la distribución Weibull

$$f(t; \gamma, \lambda) = \begin{cases} \gamma \lambda t^{\gamma-1} \exp\{-\lambda t^\gamma\} & \gamma > 0, \lambda > 0, t \geq 0 \\ \frac{\gamma}{\lambda} \left(\frac{t}{\lambda}\right)^{\gamma-1} \exp\left\{-\left(\frac{t}{\lambda}\right)^\gamma\right\} \\ \gamma \lambda (\lambda t)^{\gamma-1} \exp\{-(\lambda t)^\gamma\} \end{cases}$$

con γ parámetro de forma y λ parámetro de escala.

Función de supervivencia

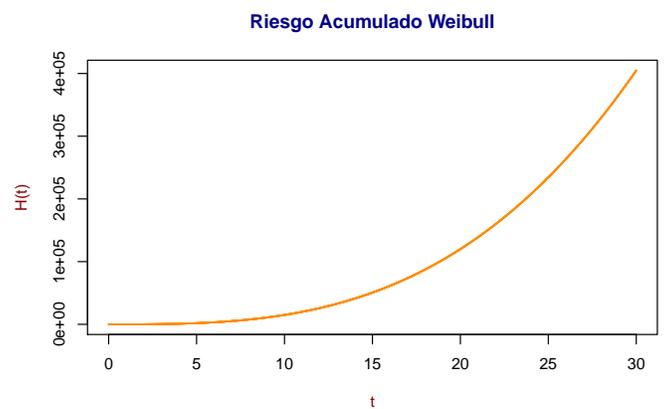
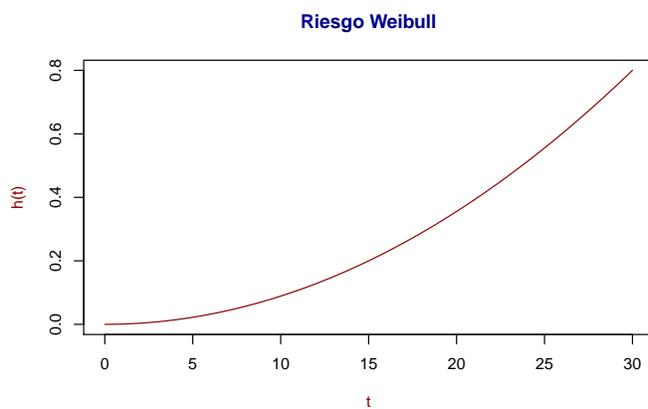
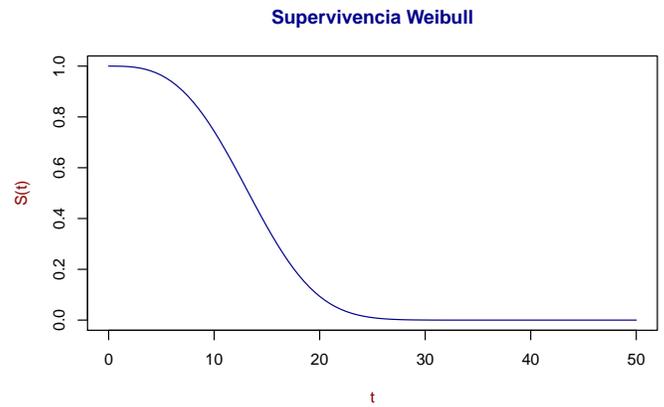
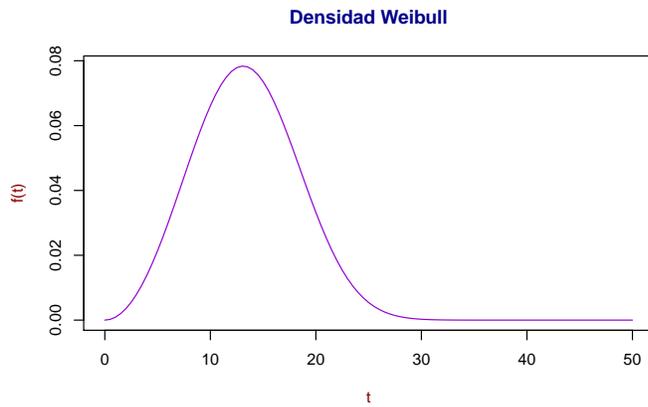
$$S(t) = \int_t^\infty \gamma \lambda u^{\gamma-1} \exp\{-\lambda u^\gamma\} du = -\exp\{-\lambda u^\gamma\} \Big|_t^\infty = \exp\{-\lambda t^\gamma\}$$

Función de riesgo

$$h(t) = \frac{f(t)}{S(t)} = \gamma \lambda t^{\gamma-1}$$

Función de riesgo acumulado

$$H(t) = \int_0^t h(u) du = \int_0^t \gamma \lambda u^{\gamma-1} du = \lambda t^\gamma$$



¿Qué tipo de riesgos modela?

$h'(t) = (\gamma - 1)\gamma\lambda t^{\gamma-2}$, entonces

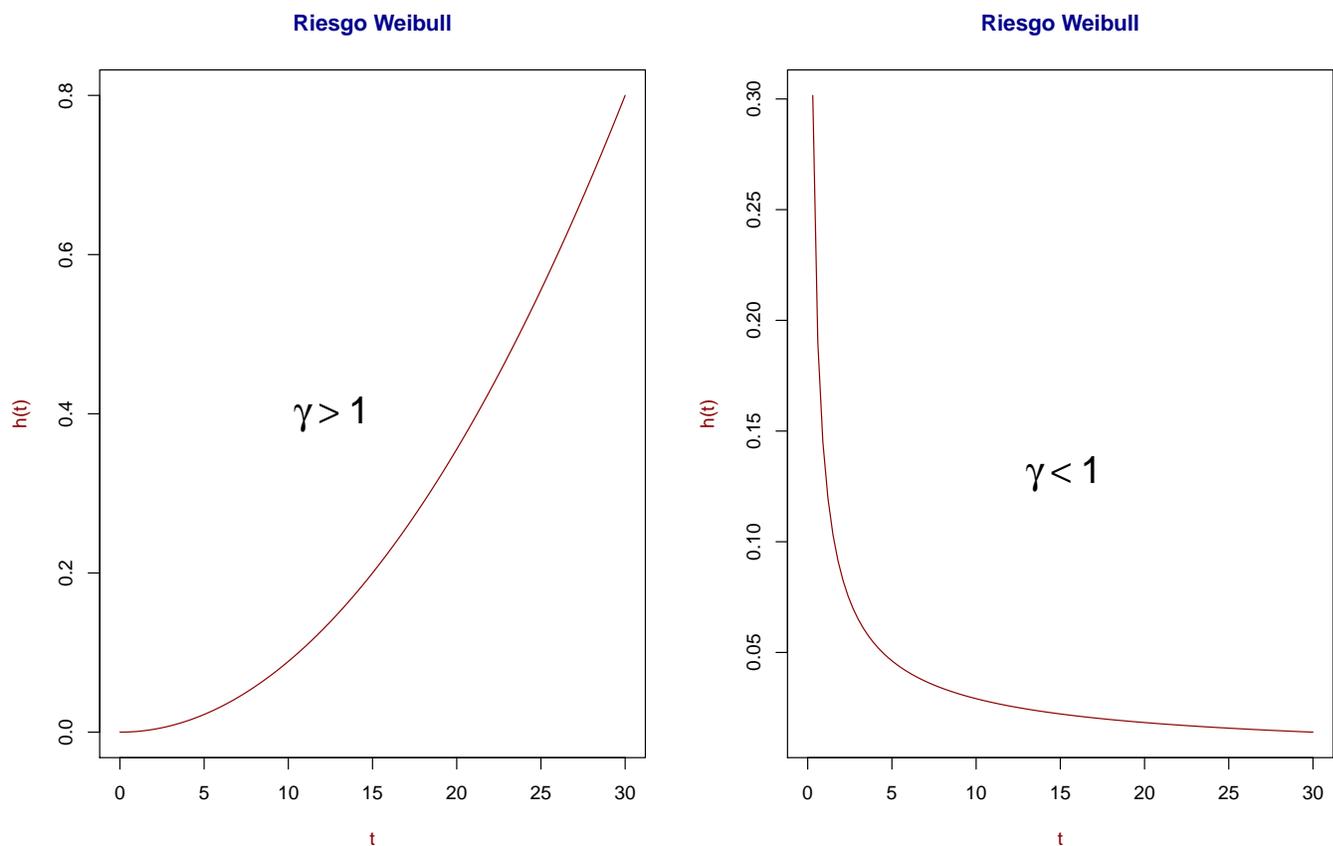
$$h'(t) \begin{cases} > 0, & \text{si } \gamma > 1 \\ < 0, & \text{si } \gamma < 1 \\ 0, & \text{si } \gamma = 1 \end{cases}$$

Por lo que concluimos que:

Si $\gamma > 1$, modela riesgos monótonos crecientes

Si $\gamma < 1$, modela riesgos monótonos decrecientes

Si $\gamma = 1$, $h(t) = \lambda$ y la Weibull se transforma en la exponencial.



Características numéricas

Media y varianza

$$\mathbb{E}(T) = \lambda^{-1/\gamma} \Gamma(1 + 1/\gamma), \quad \mathbb{V}(T) = \lambda^{-2/\gamma} [\Gamma(1 + 2/\gamma) - \Gamma(1 + 1/\gamma)^2]$$

Donde $\Gamma(t)$ es la función gamma, definida como

$$\Gamma(t) = \int_0^{\infty} y^{t-1} e^{-y} dy$$

Usual como medida de tendencia central es la mediana, porque la Weibull es sesgada a la derecha

$$t = \left(\frac{\log(2)}{\lambda} \right)^{1/\gamma}$$

Distribución Log normal

Se dice que $T \sim \text{Log-normal}(\mu, \sigma^2)$ si $\log(T) \sim N(\mu, \sigma^2)$

Función de densidad

$$f(t) = \frac{1}{\sqrt{2\pi\sigma^2 t}} \exp\left\{-\frac{(\log(t) - \mu)^2}{2\sigma^2}\right\}, \quad \mu \in \mathbb{R}, \sigma^2 > 0, t \geq 0$$

Función de supervivencia

$$S(t) = \int_t^\infty \frac{1}{\sqrt{2\pi\sigma^2 u}} \exp\left\{-\frac{(\log(u) - \mu)^2}{2\sigma^2}\right\} du = 1 - \Phi\left(\frac{\log(t) - \mu}{\sigma}\right)$$

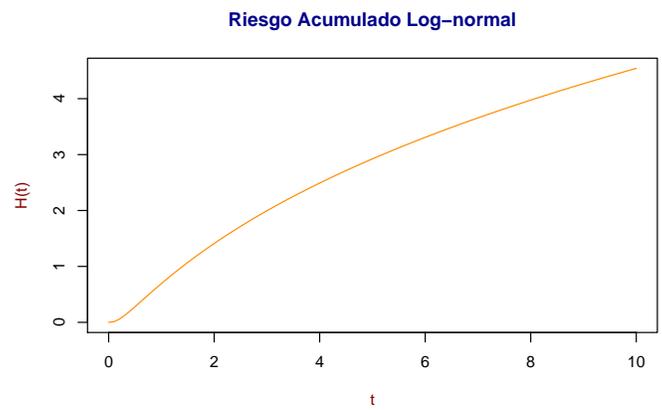
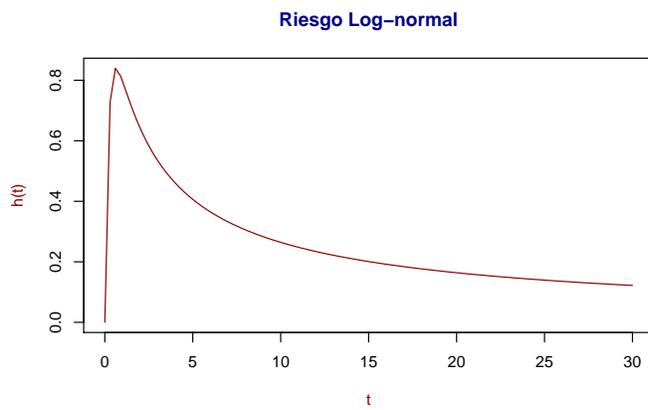
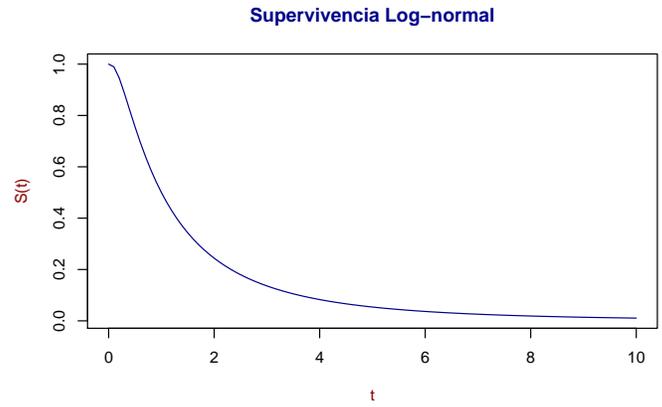
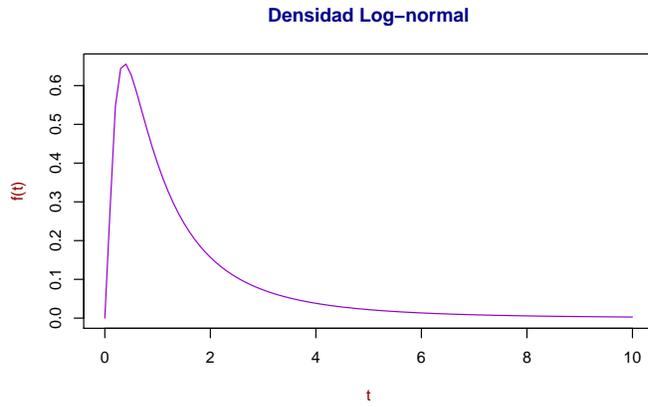
con $\Phi(\cdot)$, la función de distribución de una normal estándar.

Función de riesgo

$$h(t) = \frac{f(t)}{S(t)} = \frac{\frac{1}{\sqrt{2\pi\sigma^2 t}} \exp\left\{-\frac{(\log(t) - \mu)^2}{2\sigma^2}\right\}}{1 - \Phi\left(\frac{\log(t) - \mu}{\sigma}\right)}$$

Función de riesgo acumulado

$$H(t) = \int_0^t h(u) du$$



¿Qué tipo de riesgos modela?

Analicemos el comportamiento de $h(t)$

Para valores grandes de t

$$\lim_{t \rightarrow \infty} h(t) = \frac{\lim_{t \rightarrow \infty} \frac{1}{\sqrt{2\pi\sigma^2 t}} \exp\left\{-\frac{(\log(t) - \mu)^2}{2\sigma^2}\right\}}{\lim_{t \rightarrow \infty} 1 - \Phi\left(\frac{\log(t) - \mu}{\sigma}\right)} = \frac{0}{0}$$

utilizando *L'Hôpital*

$$f'(t) = -\frac{1}{\sqrt{2\pi\sigma^2 t^2}} \exp\left\{-\frac{(\log(t) - \mu)^2}{2\sigma^2}\right\} + \frac{1}{\sqrt{2\pi\sigma^2 t}} \exp\left\{-\frac{(\log(t) - \mu)^2}{2\sigma^2}\right\} \left(-\frac{1}{\sigma^2 t} (\log(t) - \mu)\right)$$

y

$$S'(t) = -\frac{1}{\sqrt{2\pi\sigma^2 t}} \exp\left\{-\frac{(\log(t) - \mu)^2}{2\sigma^2}\right\}$$

$$\Rightarrow \lim_{t \rightarrow \infty} h(t) = \lim_{t \rightarrow \infty} \frac{f'(t)}{S'(t)} = \lim_{t \rightarrow \infty} \left(\frac{1}{t} + \frac{\log(t) - \mu}{\sigma^2 t}\right) = \frac{\sigma^2 + \log(t) - \mu}{\sigma^2 t} = \frac{\infty}{\infty} \text{ !CHALE!}$$

“De nuez”, *L'Hôpital*

$$\lim_{t \rightarrow \infty} h(t) = \lim_{t \rightarrow \infty} \frac{\frac{1}{t}}{\sigma^2 t} = \frac{1}{\sigma^2 t} = 0$$

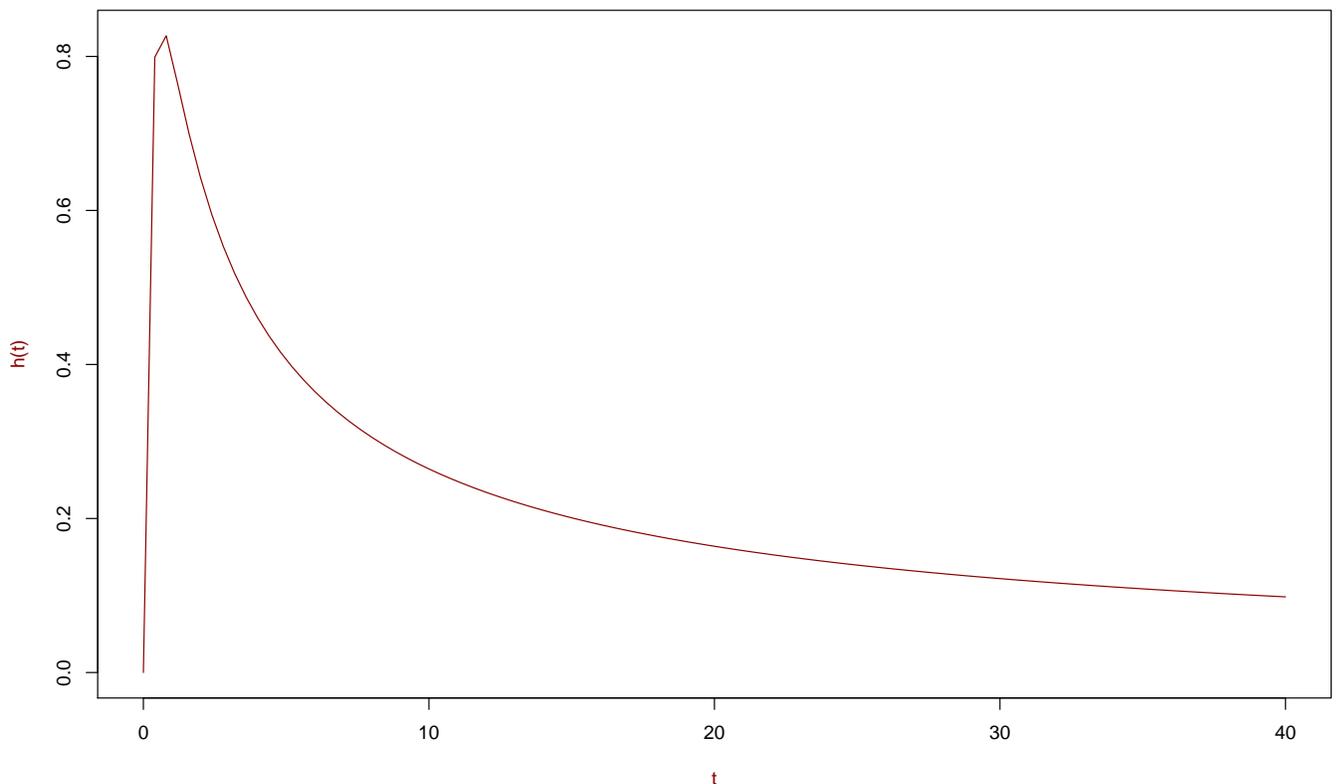
lo que implica que el riesgo desaparece cuando t es grande (?).

Y ¿qué pasa si t es pequeño?

$$\lim_{t \rightarrow 0} h(t) = \frac{0}{1} = 0, \quad \text{ya que } \lim_{t \rightarrow 0} S(t) = 1$$

Entonces, la lognormal modela riesgos que inician en cero, crecen hasta alcanzar un máximo y después decrecen a cero. Este último hecho la hace cuestionable como modelo de supervivencia. Sin embargo, puede utilizarse, sobre todo, si valores grandes del tiempo no son de interés.

Riesgo Log-normal



Características numéricas

Media y varianza

$$\mathbb{E}(T) = \exp\left\{\mu + \frac{\sigma^2}{2}\right\}, \quad \mathbb{V}(T) = \exp\{2\mu + 2\sigma^2\} - \exp\{2\mu + \sigma^2\}$$

Mediana

$$t = e^\mu$$

Distribución Gamma

Función de densidad

$$f(t) = \frac{\lambda^k t^{k-1} \exp\{-\lambda t\}}{\Gamma(k)}, \quad k > 0, \lambda > 0, t > 0$$

Función de supervivencia

$$S(t) = \int_t^\infty f(u) du = 1 - GI(k, \lambda t)$$

con $GI(k, x)$ la función gamma incompleta, definida como

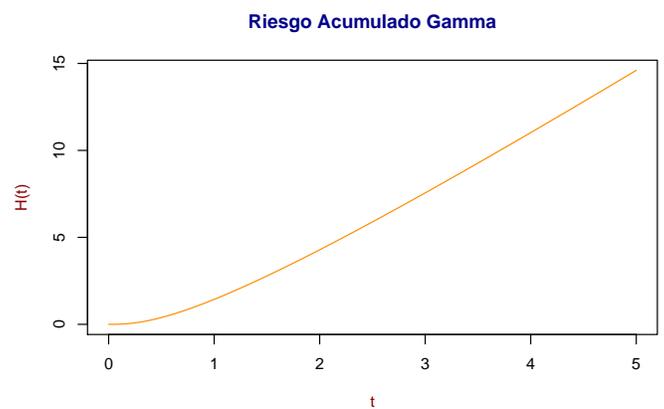
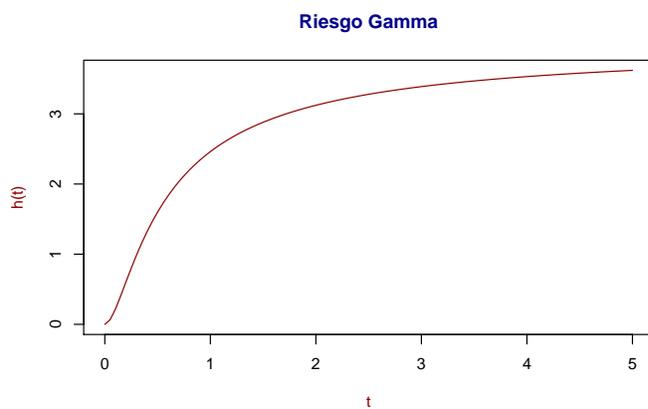
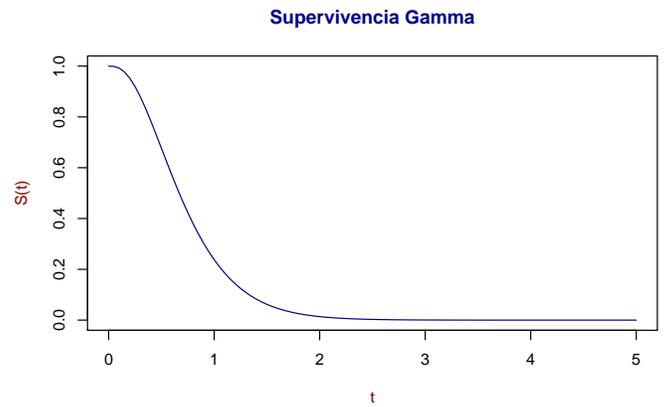
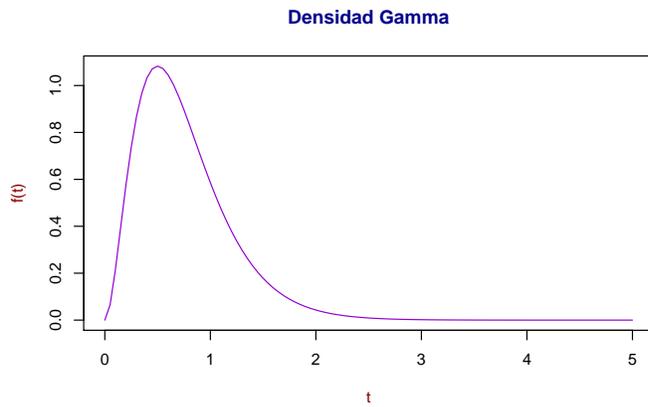
$$GI(k, x) = \frac{1}{\Gamma(k)} \int_0^x u^{k-1} e^{-u} du$$

Función de riesgo

$$h(t) = \frac{f(t)}{S(t)} = \frac{\lambda^k t^{k-1} \exp\{-\lambda t\}}{\Gamma(k) (1 - GI(k, \lambda t))}$$

Función de riesgo acumulado

$$H(t) = \int_0^t h(u) du$$



¿Qué tipo de riesgos modela?

Si t es grande

$$\lim_{t \rightarrow \infty} h(t) = \lim_{t \rightarrow \infty} \frac{t^{k-1} \exp\{-\lambda t\}}{\int_t^{\infty} u^{k-1} e^{-u} du} = \frac{0}{0}$$

L' Hôpital

$$\begin{aligned} \lim_{t \rightarrow \infty} h(t) &= \lim_{t \rightarrow \infty} \frac{f'(t)}{S'(t)} = \lim_{t \rightarrow \infty} \frac{(k-1)t^{k-2} \exp\{-\lambda t\} - \lambda t^{k-1} \exp\{-\lambda t\}}{-t^{k-1} \exp\{-\lambda t\}} \\ &= \lim_{t \rightarrow \infty} - \left[\frac{(k-1)}{t} - \lambda \right] = \lambda \end{aligned}$$

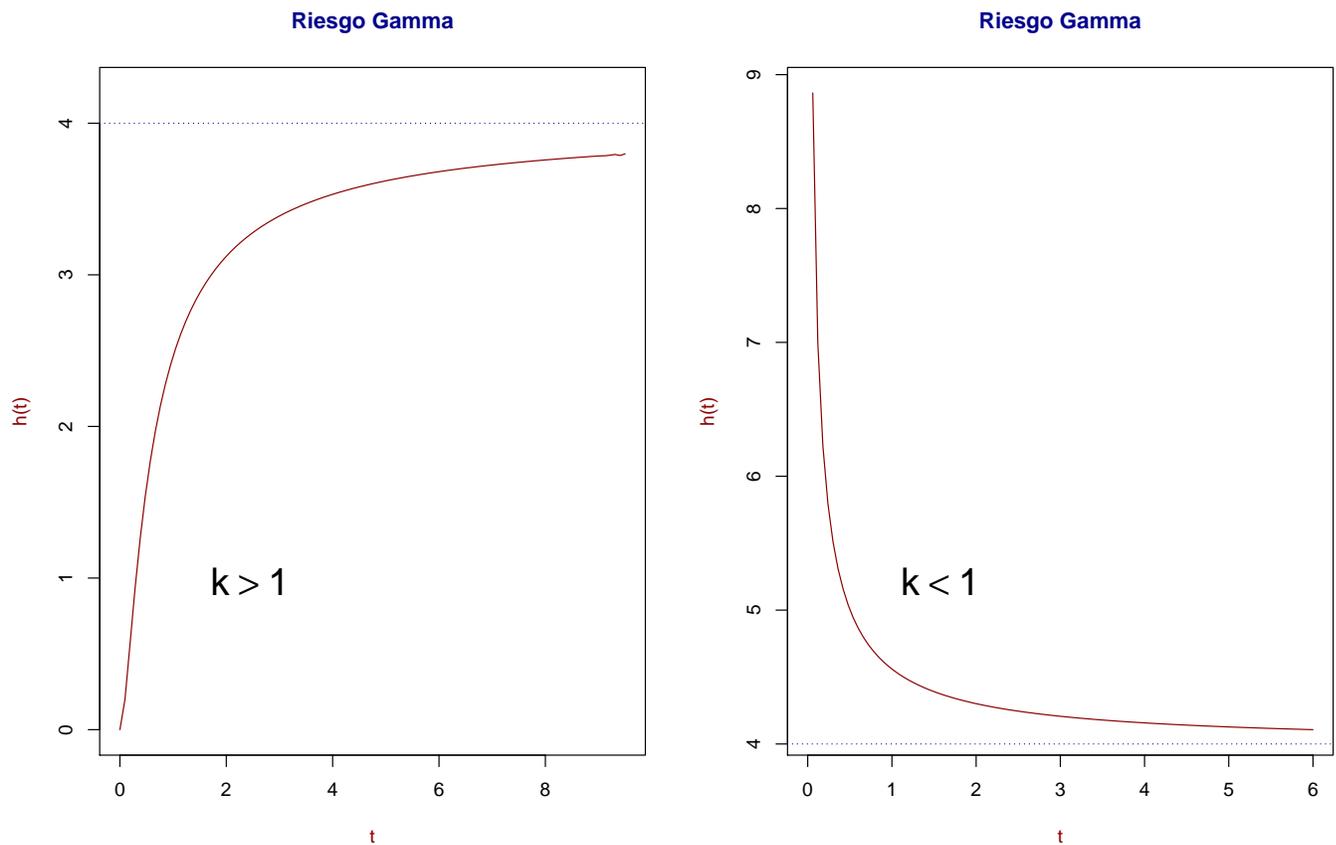
independiente del parámetro k .

¿Qué pasa para valores pequeños de t ?

$$\lim_{t \rightarrow 0} h(t) = \lim_{t \rightarrow 0} \frac{t^{k-1} \exp\{-\lambda t\}}{\int_t^\infty u^{k-1} e^{-u} du} = \lim_{t \rightarrow 0} \frac{t^{k-1} \exp\{-\lambda t\}}{1} = \lim_{t \rightarrow 0} t^{k-1} * \lim_{t \rightarrow 0} \exp\{-\lambda t\} = \lim_{t \rightarrow 0} t^{k-1}$$

$$\Rightarrow \lim_{t \rightarrow 0} t^{k-1} = \begin{cases} \infty, & \text{si } k < 1 \\ 0, & \text{si } k > 1 \end{cases}$$

Finalmente, si $k < 1$ la gamma modela riesgos que inicialmente son muy grandes y después decrecen hasta hacerse constantes (λ) para valores grandes de t . Y si $k > 1$, modela riesgos que son nulos en un inicio, y crecen hasta hacerse constantes (λ) para t grande. Si $k = 1$, la gamma es la exponencial y modela riesgos constantes (λ).



Características numéricas

Media y varianza

$$\mathbb{E}(T) = \frac{k}{\lambda}, \quad \mathbb{V}(T) = \frac{k}{\lambda^2}$$

Distribución Gompertz

Función de densidad

$$f(t) = \lambda \varphi^t \exp \left\{ -\frac{\lambda}{\log(\varphi)} (\varphi^t - 1) \right\}, \quad \lambda > 0, \varphi > 0, t > 0$$

Función de supervivencia

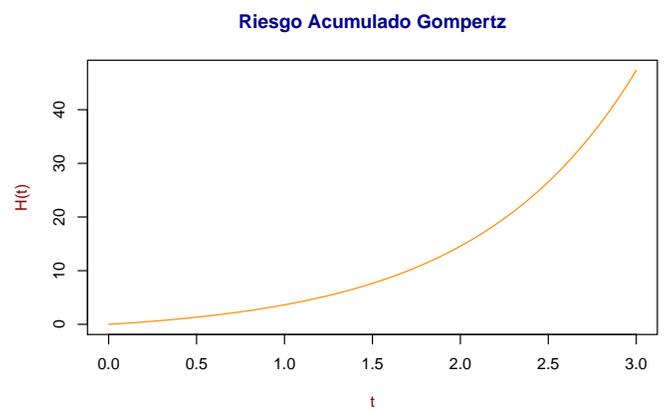
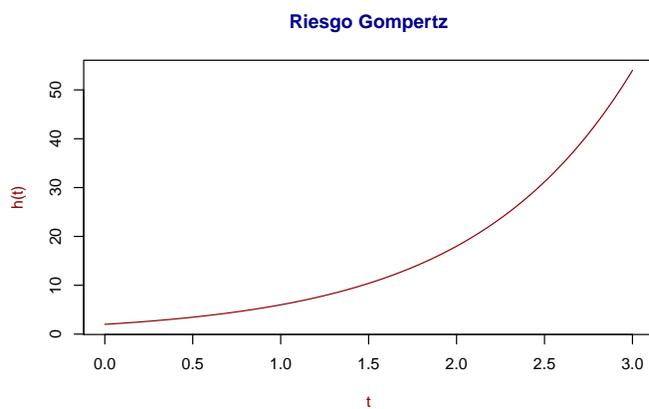
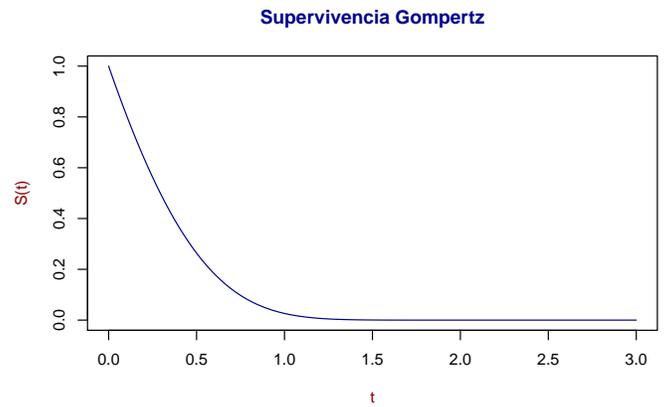
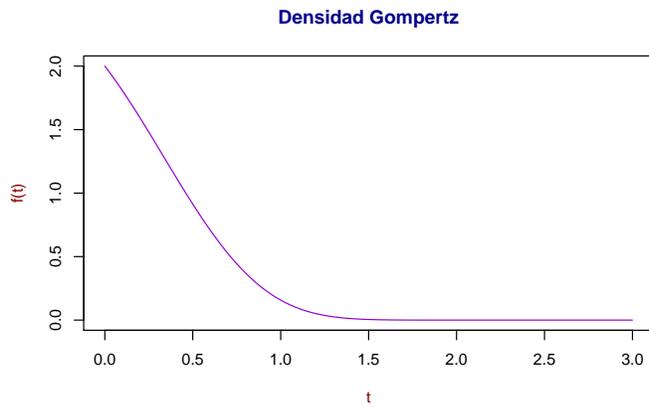
$$S(t) = \exp \left\{ -\frac{\lambda}{\log(\varphi)} (\varphi^t - 1) \right\}$$

Función de riesgo

$$h(t) = \frac{f(t)}{S(t)} = \lambda \varphi^t$$

Función de riesgo acumulado

$$H(t) = \frac{\lambda (\varphi^t - 1)}{\log(\varphi)}$$

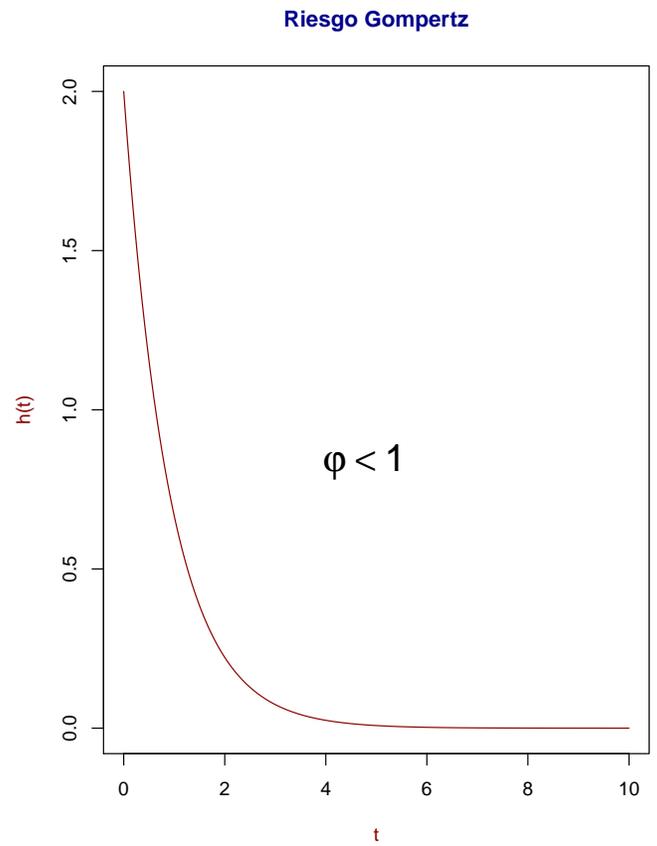
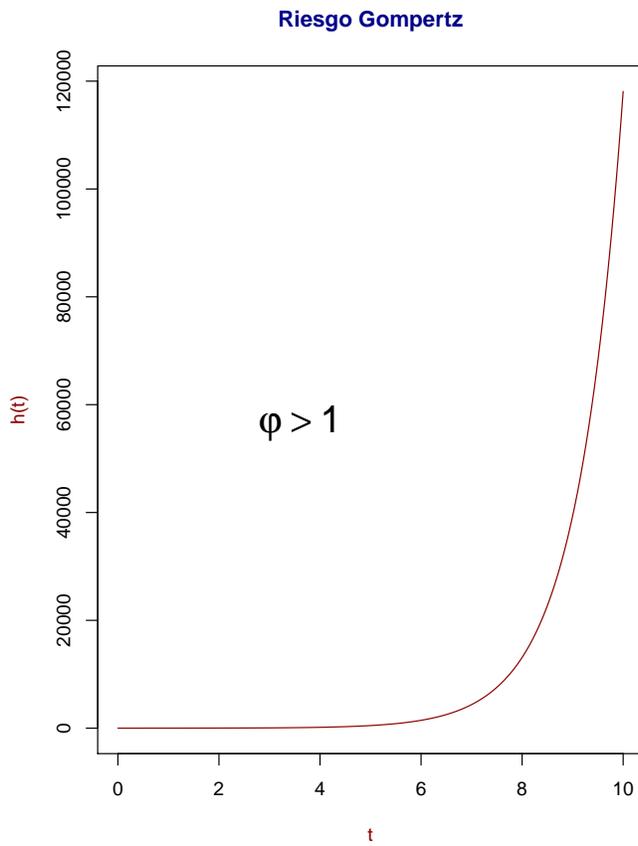


¿Qué tipo de riesgos modela?

Si $\varphi > 1 \Rightarrow \lim_{t \rightarrow \infty} h(t) = \infty$ y $h(0) = \lambda$

Si $\varphi < 1 \Rightarrow \lim_{t \rightarrow \infty} h(t) = 0$

Entonces, Si $\varphi > 1$ modela riesgos que son distintos de cero, (λ), al inicio del estudio y son infinitos, cuando t es grande. Si $\varphi < 1$, modela riesgos que inician en λ y eventualmente desaparecen.



Tarea 2

Dada la función de densidad de la Gompertz

$$f(t) = \lambda \varphi^t \exp \left\{ -\frac{\lambda}{\log(\varphi)} (\varphi^t - 1) \right\}, \quad \lambda > 0, \varphi > 0, t > 0$$

Demostrar que si $\varphi \rightarrow 1 \Rightarrow f(t) = \text{Exp}(\lambda)$.

Distribución log-logística

Una variable aleatoria, T , se dice que tiene una distribución log-logística si su logaritmo, $\log T$, tiene una distribución logística. Su función de densidad está dada por

$$f(t) = \frac{\gamma \lambda t^{\gamma-1}}{(1 + \lambda t^\gamma)^2}, \quad \gamma > 0, \lambda > 0, t > 0$$

Función de distribución

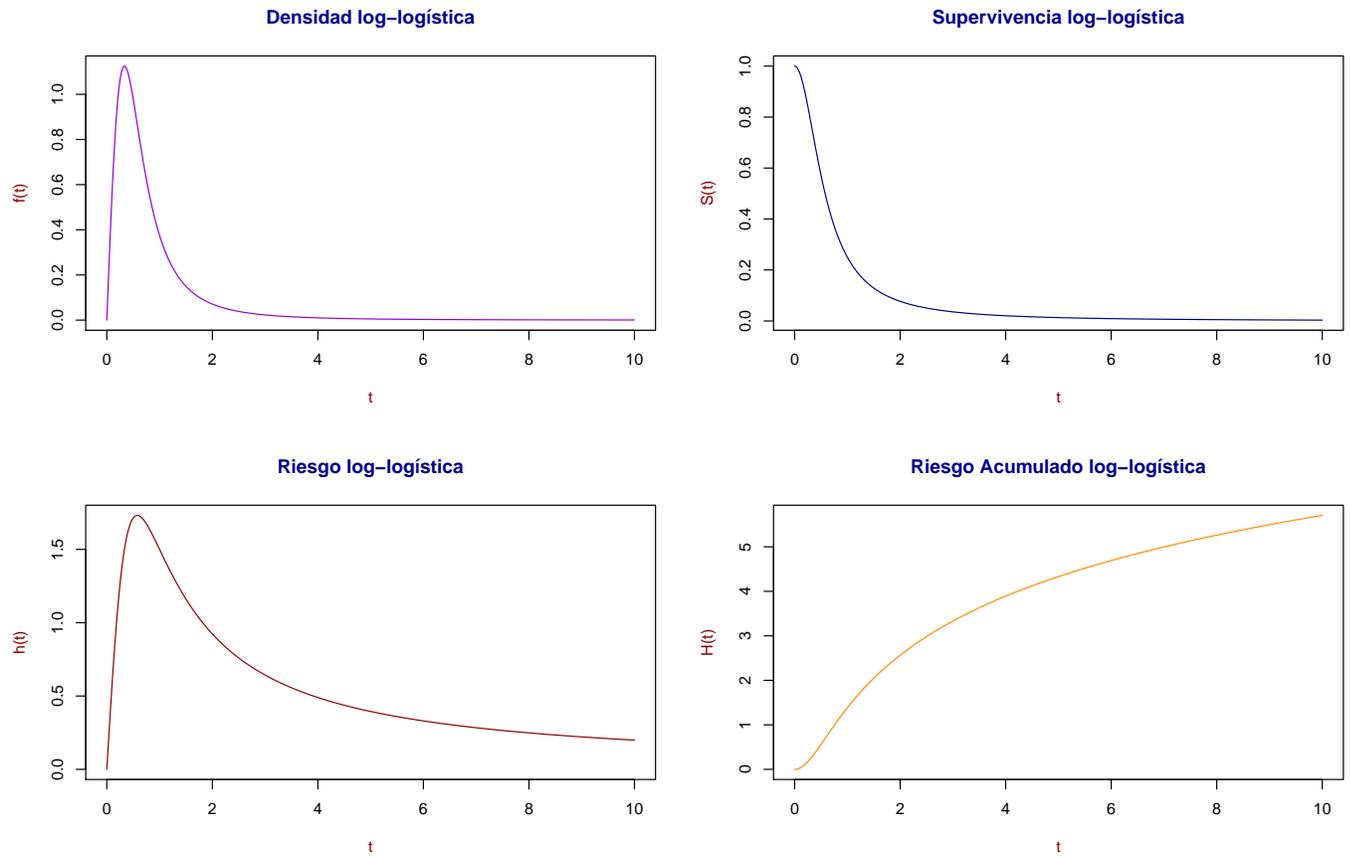
$$S(t) = \int_t^\infty \frac{\gamma \lambda u^{\gamma-1}}{(1 + \lambda u^\gamma)^2} du = - (1 + \lambda u^\gamma)^{-1} \Big|_t^\infty = \frac{1}{1 + \lambda t^\gamma}$$

Función de riesgo

$$h(t) = \frac{f(t)}{S(t)} = \frac{\gamma \lambda t^{\gamma-1}}{1 + \lambda t^\gamma}$$

Función de riesgo acumulado

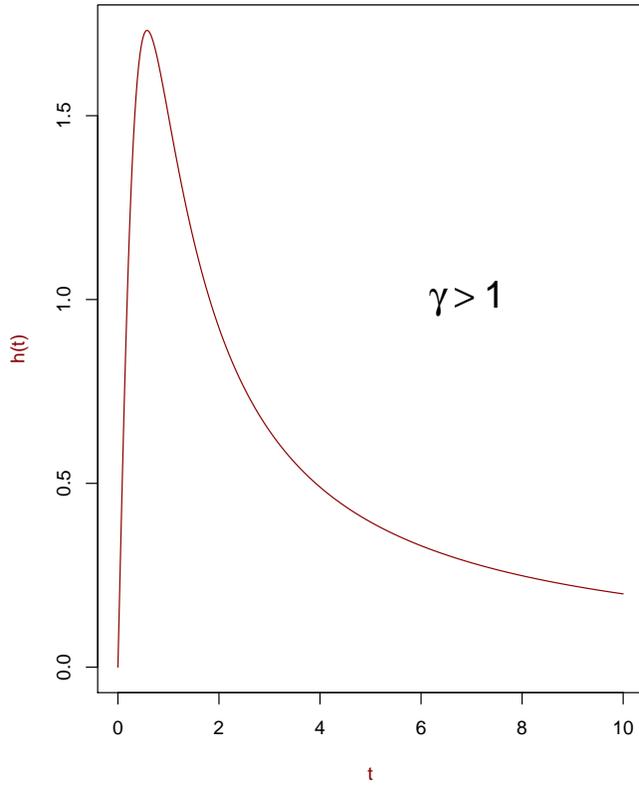
$$H(t) = \ln(1 + \lambda t^\gamma)$$



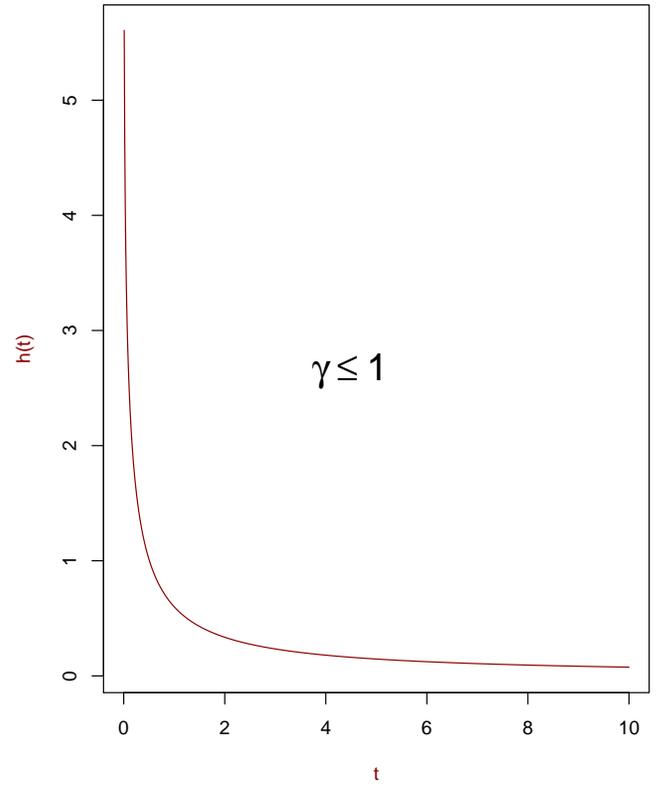
¿Qué riesgos modela?

Obsérvese que el numerador de la función de riesgo es el mismo que el de la función de riesgo de la Weibull, pero el denominador hace que esta función tenga características distintas. A saber, es monótona decreciente para $\gamma \leq 1$. Para $\gamma > 1$, crece hasta un máximo y decrece a cero para valores grandes de t .

Riesgo log-logística



Riesgo log-logística



Métodos no paramétricos para el análisis de supervivencia

Dada la importancia que tiene la función de supervivencia en esta área estadística, se han propuesto diversos estimadores de esta función.

Función de supervivencia empírica

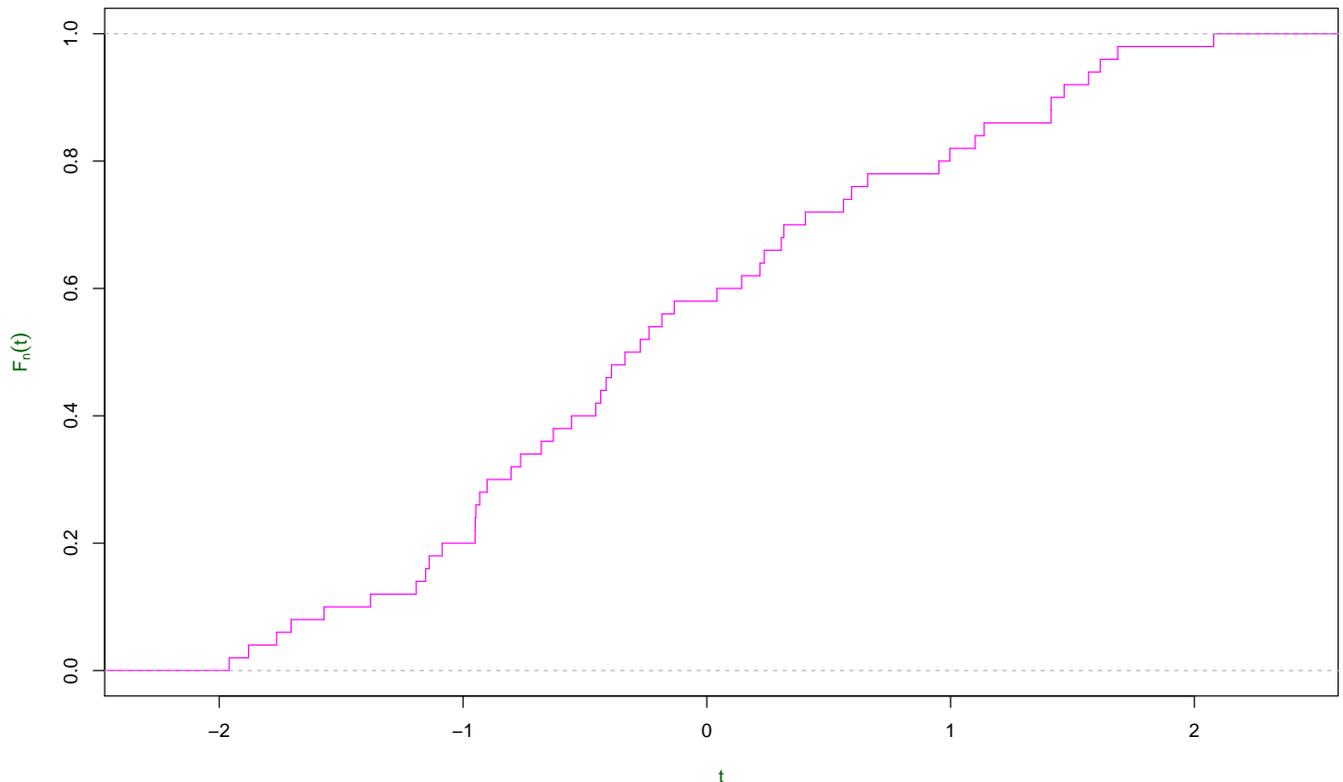
En estadística usual, la función de distribución empírica es un estimador “chido” de la función de distribución. De hecho, al resultado

$$\sup_{t \in \mathbb{R}} |F_n(t) - F(t)| \xrightarrow{c.s.} 0, \text{ cuando } n \rightarrow \infty$$

se le conoce como *el teorema fundamental de la estadística*. Con

$$F_n(t) = \frac{\#T'_s \leq t}{n}$$

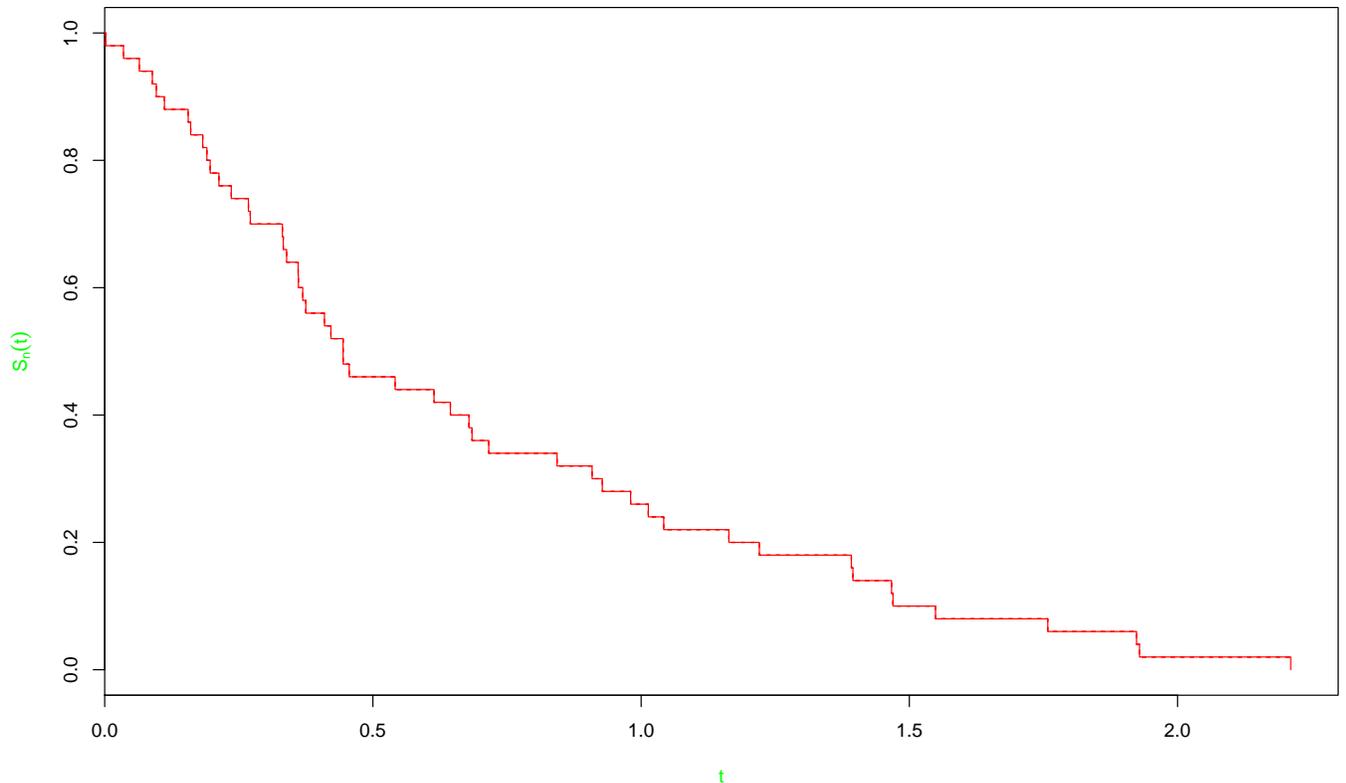
Función de distribución empírica



Entonces, podríamos pensar que la función de supervivencia empírica, definida como

$$S_n(t) = \frac{\#T'_s \geq t}{n}$$

Función de Supervivencia Empírica



podría ser un buen estimador de la función de supervivencia.

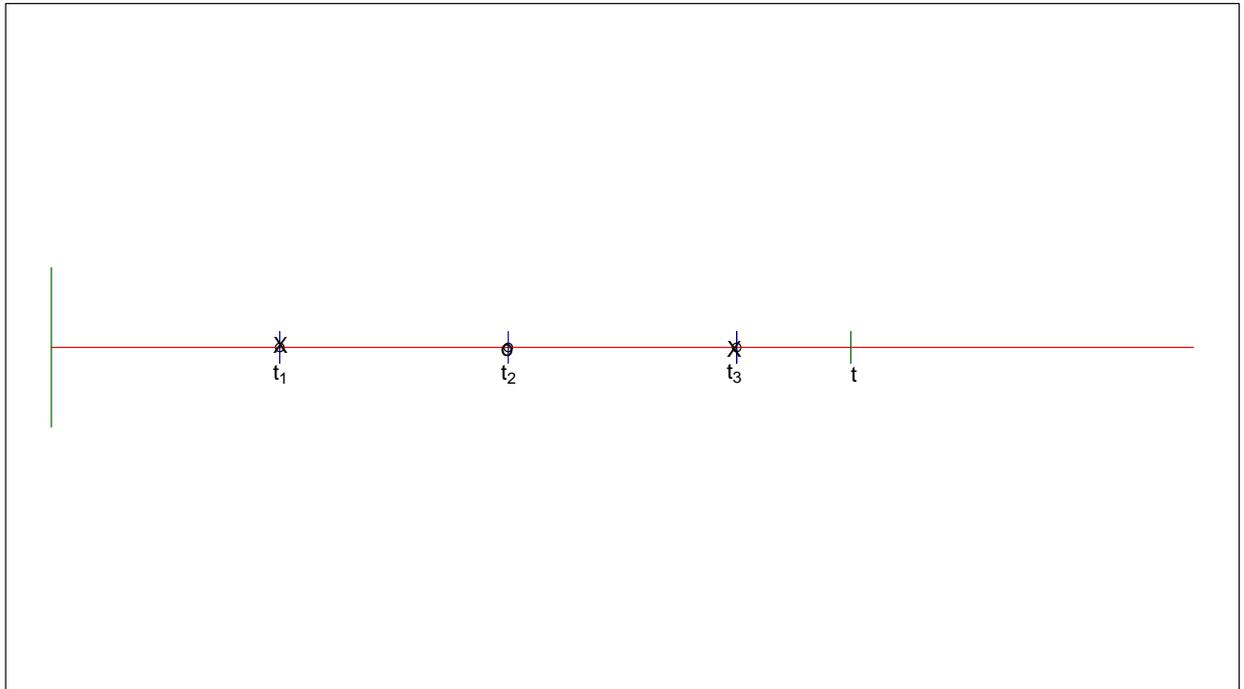
Características de S_n

- Función escalonada decreciente.
- Si no hay empates, saltos de tamaño $1/n$.
- Si hay empates (d): saltos de tamaño d/n .

Desventaja: Desperdicia información, ya que no toma en cuenta si t_i fue falla o censura.

Consideremos el siguiente ejemplo con n individuos, con los tres primeros tiempos de supervivencia: t_1, t_2 y t_3 , donde t_1 y t_3 son fallas y t_2 es una censura, y queremos calcular la

supervivencia al tiempo, t . Entonces tenemos



tiempo

$$S_n(t) = \frac{n-3}{n}$$

pero, en realidad no sabemos en qué momento falló el individuo que reporta el tiempo censurado t_2 . Lo que sí sabemos es que falló después de t_2 , por lo tanto, estamos *subestimando* la verdadera supervivencia. Entonces, la función de supervivencia empírica no es un buen estimador de la función de supervivencia. Si *no hay censura*, $S_n(t)$ sí es un buen estimador.

Tabla de vida

Un segundo estimador de la función de supervivencia es el que se obtiene a través de la llamada *Tabla de vida*, conocida también como *Tabla actuarial de vida*.

- Extensión de la tabla de frecuencias relativas.
- El tiempo de observación se divide en k intervalos.

$$0 = a_0 < a_1 < a_2 < \dots < a_k = \infty$$

Definamos

intervalo $I_j = [a_{j-1}, a_j)$ $j=1, \dots, k$.

d_j : # individuos que mueren en I_j

c_j : # individuos censurados en I_j

n_j : # individuos en riesgo al inicio de I_j (*i.e.* vivos y no censurados al tiempo a_{j-1})

¿Cómo construimos la función de supervivencia asociada a esta tabla?

Obsérvese que

$$S(a_{j-1}) = \mathbb{P}[T \geq a_{j-1}] = \mathbb{P}[T \geq a_0] * \mathbb{P}[T \geq a_1 | T \geq a_0] * \mathbb{P}[T \geq a_2 | T \geq a_1] * \dots * \mathbb{P}[T \geq a_{j-1} | T \geq a_{j-2}]$$

Ahora, obsérvese que

$$\underbrace{\mathbb{P}[T \geq a_{j-1} | T \geq a_{j-2}]}_{\text{probabilidad de sobrevivir al periodo } [a_{j-2}, a_{j-1})} = 1 - \underbrace{\mathbb{P}[T < a_{j-1} | T \geq a_{j-2}]}_{\text{probabilidad de morir en el periodo } [a_{j-2}, a_{j-1})}$$

¿Cómo calcular estas probabilidades?

Supondremos que las censuras se distribuyen uniformes en el intervalo (suena parecido a los histogramas, ¿no?). Entonces, supondremos que los individuos se censuran a mitad del

intervalo (marca de clase). Pero, cómo estimamos esas probabilidades?

Si no hubiera censura

$$\mathbb{P}[T < a_j | T \geq a_{j-1}] = \frac{d_j}{n_j}$$

el supuesto sobre la censura, nos permite considerar que los individuos censurados permanecieron expuestos al riesgo *la mitad del intervalo*. Por lo tanto

$$\mathbb{P}[T < a_j | T \geq a_{j-1}] = \frac{d_j}{n_j - c_j/2}$$

esta es la probabilidad estimada de morir en el intervalo $[a_{j-1}, a_j)$. La probabilidad estimada de sobrevivir a este intervalo es

$$\mathbb{P}[T \geq a_j | T \geq a_{j-1}] = 1 - \frac{d_j}{n_j - c_j/2}$$

¿Cómo definimos entonces la función de supervivencia?

$$\tilde{S}(a_{j-1}) = \prod_{i=1}^{j-1} \left(1 - \frac{d_i}{n_i - c_i/2} \right)$$

Probabilidad de que un individuo sobreviva (esté vivo) al inicio del intervalo I_j .

Y para un valor, t , en general

$$\tilde{S}(t) = \tilde{S}(a_{j-1}), \quad \text{si } t \in I_j, \quad j=1,2,\dots,k.$$

Función de riesgo

$$\tilde{\lambda}(a_{j-1}) = \frac{d_j}{(n_j - d_j/2)\tau_j}$$

con $n'_j = n_j - c_j/2$ y τ_j : tamaño del intervalo j . Esta función es propiamente, la función de riesgo por unidad de tiempo (longitud del intervalo). Obsérvese que para las muertes d_j , se hace un supuesto similar al de la censura, y se considera que ocurrieron a mitad del intervalo, por lo que estuvieron la mitad del tiempo en riesgo.

Obviamente, todas estas cantidades: $\hat{q}_j = \frac{d_j}{n'_j}$, $\hat{p}_j = 1 - \hat{q}_j$ y \tilde{S} , están sujetas a variación muestral. Sus varianzas son

$$\widehat{var}(\tilde{S}) = \tilde{S}^2 \sum_{i=1}^j \frac{\hat{q}_i}{n'_i \hat{p}_i}$$

$$\widehat{var}(\hat{p}_j) = \frac{\hat{p}_j \hat{q}_j}{n'_j}$$

$$\widehat{var}(\hat{q}_j) = \frac{\hat{q}_j - \hat{q}_j^2}{n'_j}$$

Las últimas dos expresiones se deducen del hecho que dado n_j , $d_j \sim Bin(n_j, 1 - p_j)$.

La construcción de la tabla de vida tiene, esencialmente, los mismos problemas que la construcción de un histograma, a saber

- Número de intervalos
- Tamaño de los intervalos

Finalmente, los resultados se resumen en una tabla cuyo formato es

Tabla de vida										
Int.	#muer.	#cen.	#ind. en rie.	n'_j	\hat{q}_j	\hat{p}_j	\tilde{S}	$\widehat{var}(\tilde{S})$	$\widehat{var}(\hat{p}_j)$	$\widehat{var}(\hat{q}_j)$
I_j	d_j	c_j	n_j	$n_j - c_j/2$	d_j/n'_j	$1 - \hat{q}_j$				

El bueno: El estimador Kaplan-Meier (K-M)

Sabemos que, en general, el método de máxima verosimilitud proporciona buenos estimadores de las características de interés. El estimador K-M es el estimador máximo verosímil de la función de supervivencia. Este estimador también es conocido como *estimador producto límite*, por la forma en que se construye.

Preámbulo

- Usualmente tenemos una muestra aleatoria, $t_1, t_2, \dots, t_n, t_i \sim f(\cdot; \underline{\theta})$. $f(\cdot)$ proporciona información sobre $\underline{\theta}$.
- $L(\underline{\theta}; \underline{t})$: función de verosimilitud (“junta” la información que aportan, sobre $\underline{\theta}$, los n individuos de la muestra)
- $L(\underline{\theta}; \underline{t})$, es función de $\underline{\theta}$ y de \underline{t} . Pero, para una muestra fija \underline{t} observada, es función sólo de $\underline{\theta}$.

Depende de $\underline{\theta}$!DESCONOCIDO!. La idea es encontrar el valor de $\underline{\theta}$ que maximice la probabilidad de haber observado la muestra, i.e., cuál es el valor de $\underline{\theta}$ que hace máxima la probabilidad de observar \underline{t} .

- Queremos estimar $S(t)$...!No es un parámetro!...es !una función!
- Espacio paramétrico: Espacio de funciones, y L es una funcional.

Construcción de la verosimilitud

Suponemos:

k tiempos de falla observados t_1, t_2, \dots, t_k

c_j censuras entre dos tiempos de falla consecutivos, $j = 0, 1, \dots, k$, i.e. suponemos

c_0 censuras entre t_0 y t_1

c_1 censuras entre t_1 y t_2

\vdots

c_{k-1} censuras entre t_{k-1} y t_k

c_k censuras después de $t_k \in [t_k, \infty)$

d_k empates en t_k , i.e., d_k tiempos de falla ocurridos en t_k , i.e. supondremos

d_0 empates en t_0

d_1 empates en t_1

\vdots

d_k empates en t_k

El K-M es el caso límite del proceso de tabla de vida: # de intervalos $\rightarrow \infty$, por consiguiente: tamaño del intervalo $\rightarrow 0$. Entonces, “los intervalos” deberían contener únicamente una falla o todas las fallas empatadas ahí.

$$I_j = [t_j, t_{j+1}), \quad j = 0, 1, \dots, k$$

En I_j hay c_j censuras en los tiempos $t_{j1}, t_{j2}, \dots, t_{jc_j}$

¿Cómo construir la verosimilitud, $L(S)$?

Los individuos contribuyen a la verosimilitud de dos maneras:

Individuo falló $\Rightarrow f(\cdot; \underline{\theta})$ es su contribución

Individuo se censuró $\Rightarrow S(\cdot; \underline{\theta})$ es su contribución, porque se supone que vivió al menos “un instante” después del valor de censura observado.

\Rightarrow si se censuró contribuye con $S(t_{rs})$ (Se censuró en el r -ésimo intervalo, y es la s -ésima censura de las c_r totales en ese intervalo.

Si murió, por ejemplo, en t_j , su contribución es: $S(t_j^-) - S(t_j)$. Donde $S(t_j^-) = \lim_{x \rightarrow 0^+} S(t_j - x)$.

Entonces, $S(t_j^-) - S(t_j) =$ “la probabilidad de morir en el instante t_j ” $= f(t_j)$.

Y ¿cómo queda la verosimilitud?

Contribución de las fallas

$$\prod_{i=1}^k [S(t_i^-) - S(t_i)]^{d_i} \quad (\text{independencia})$$

Contribución de las censuras

$$\prod_{i=1}^k \prod_{j=1}^{c_j} S(t_{ij}) \quad (\text{independencia})$$

Por lo tanto

$$L(S) = \prod_{i=1}^k \left\{ [S(t_i^-) - S(t_i)]^{d_i} \prod_{j=1}^{c_j} S(t_{ij}) \right\}$$

Características que debe tener la función $\hat{S}(t)$ que maximice esta verosimilitud

- Debe ser discontinua (salta) en los tiempos de falla
- Continua en los tiempos de censura
- Constante entre dos tiempos de falla consecutivos

Como podemos ver, la forma funcional de esta función de supervivencia que andamos buscando, es similar a la de la supervivencia para tiempos discretos, entonces

$$S(t_j^-) = \prod_{i=0}^{j-1} (1 - \lambda_i), \quad j = 1, 2, \dots, k \text{ y } \lambda_i: \text{ probabilidad de morir en el tiempo } t_i$$

y

$$S(t_j) = \prod_{i=0}^j (1 - \lambda_i), \quad j = 1, 2, \dots, k$$

$$\Rightarrow S(t_j^-) - S(t_j) = \prod_{i=0}^{j-1} (1 - \lambda_i) [1 - (1 - \lambda_j)] = \lambda_j \prod_{i=0}^{j-1} (1 - \lambda_i)$$

si consideramos la posibilidad de tener d_j empates en cada tiempo, t_j . Entonces

$$\Rightarrow [S(t_j^-) - S(t_j)]^{d_j} = \lambda_j^{d_j} \prod_{i=0}^{j-1} (1 - \lambda_i)^{d_j}$$

y

$$S(t_{ji}) = \prod_{\ell=0}^j (1 - \lambda_{\ell}) = (1 - \lambda_j) \prod_{\ell=0}^{j-1} (1 - \lambda_{\ell})$$

Considerando todas las censuras entre los tiempos t_j y t_{j+1} , tenemos

$$S(t_{ji}) = (1 - \lambda_j)^{c_j} \prod_{\ell=0}^{j-1} (1 - \lambda_{\ell})^{c_j}$$

Entonces, la verosimilitud queda como

$$L(S) = L(\lambda_0, \lambda_1, \dots, \lambda_k) = \prod_{j=0}^k \lambda_j^{d_j} (1 - \lambda_j)^{c_j} \left(\prod_{\ell=0}^{j-1} (1 - \lambda_{\ell})^{d_j + c_j} \right)$$

Desarrollando y agrupando para los distintos valores de j , tenemos

$$\begin{aligned} L(S) = & \lambda_0^{d_0} (1 - \lambda_0)^{c_0} * \lambda_1^{d_1} (1 - \lambda_1)^{c_1} (1 - \lambda_0)^{d_1 + c_1} * \lambda_2^{d_2} (1 - \lambda_2)^{c_2} (1 - \lambda_0)^{d_2 + c_2} (1 - \lambda_1)^{d_2 + c_2} * \\ & \lambda_3^{d_3} (1 - \lambda_3)^{c_3} (1 - \lambda_0)^{d_3 + c_3} (1 - \lambda_1)^{d_3 + c_3} (1 - \lambda_2)^{d_3 + c_3} * \dots * \lambda_k^{d_k} (1 - \lambda_k)^{c_k} (1 - \lambda_0)^{d_k + c_k} (1 - \lambda_1)^{d_k + c_k} * \\ & \dots * (1 - \lambda_{k-1})^{d_k + c_k} \end{aligned}$$

Recordemos que se había definido n_j : # individuos en riesgo en el intervalo I_j , aquí, al tiempo, t_j , como los vivos y no censurados

$$\Rightarrow n_j = \sum_{i=j}^k (d_i + c_j)$$

entonces

$$\begin{aligned} L(S) = & \lambda_0^{d_0} \lambda_1^{d_1} \lambda_2^{d_2} \dots \lambda_k^{d_k} (1 - \lambda_0)^{c_0 + d_1 + c_1 + d_2 + c_2 + d_3 + c_3 + \dots + d_k + c_k} (1 - \lambda_1)^{c_1 + d_2 + c_2 + d_3 + c_3 + \dots + d_k + c_k} \dots \\ & (1 - \lambda_{k-1})^{c_{k-1} + c_k + d_k} (1 - \lambda_k)^{c_k} = \lambda_0^{d_0} \lambda_1^{d_1} \lambda_2^{d_2} \dots \lambda_k^{d_k} (1 - \lambda_0)^{n_0 - d_0} (1 - \lambda_1)^{n_1 - d_1} \dots (1 - \lambda_k)^{n_k - d_k} \end{aligned}$$

Finalmente, la verosimilitud se puede escribir como

$$L(S) = L(\lambda_0, \lambda_1, \dots, \lambda_k) = \prod_{i=0}^k \lambda_i^{d_i} (1 - \lambda_i)^{n_i - d_i}$$

Obsérvese que esta verosimilitud tiene un “rostro bastante familiar”: Bernoulli(λ_i) en cada tiempo, t_i . Con $\mathbb{P}(\text{morir en } t_i) = \lambda_i$ y $\mathbb{P}(\text{no morir en } t_i) = 1 - \lambda_i$.

Ton's 'ora sí, lo de siempre

$$\log L(S) = \ell(S) = \sum_{i=0}^k d_i \log(\lambda_i) + (1 - d_i) \log(1 - \lambda_i)$$

$$\frac{\partial \ell(S)}{\partial \lambda_i} = \frac{d_i}{\lambda_i} - \frac{n_i - d_i}{1 - \lambda_i} \Rightarrow \hat{\lambda}_i = \frac{d_i}{n_i}$$

Será un máximo?

$$\frac{\partial^2 \ell(S)}{\partial \lambda_i^2} = -\frac{d_i}{\lambda_i^2} - \frac{n_i - d_i}{(1 - \lambda_i)^2}, \quad y \quad \frac{\partial^2 \ell(S)}{\partial \lambda_i \partial \lambda_j} = 0$$

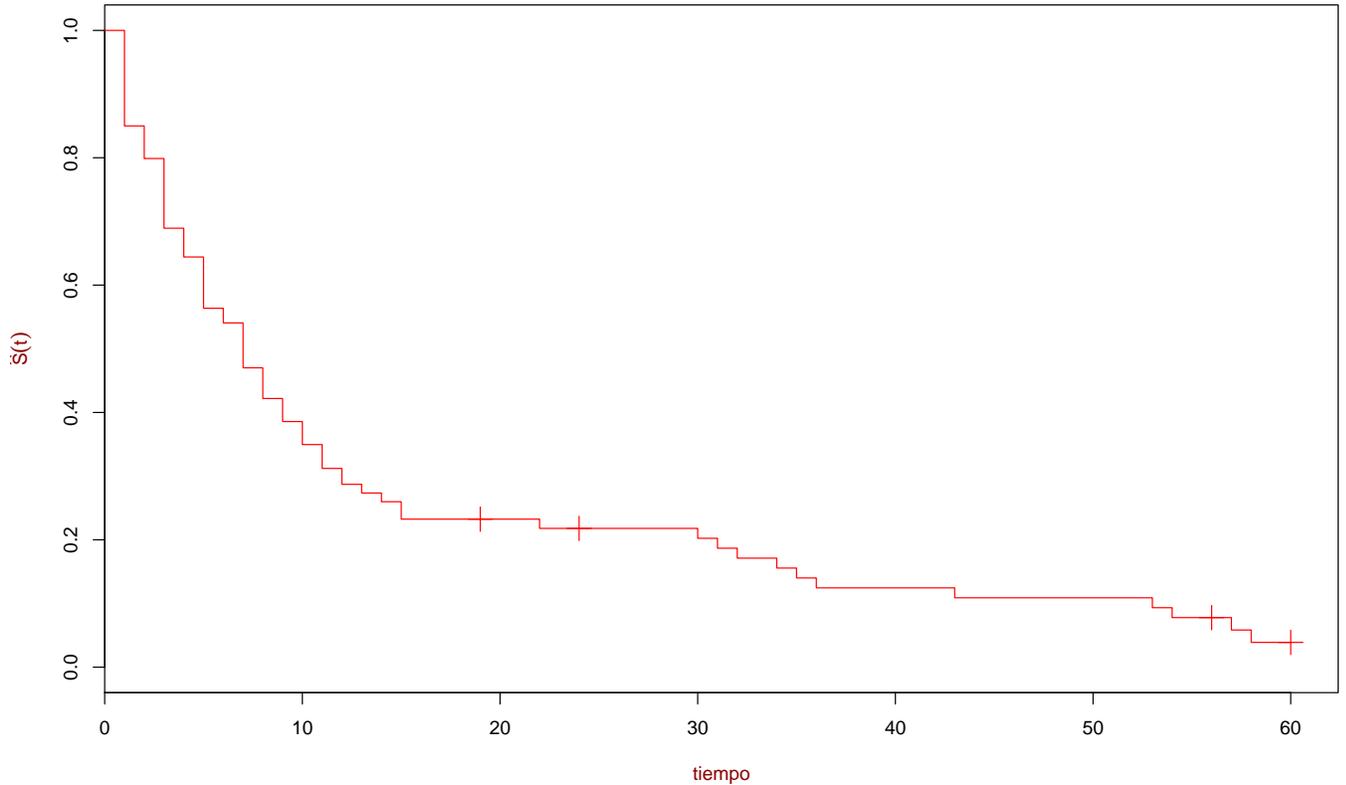
Entonces, *la matriz Hessiana* es negativa definida, por lo tanto, $\hat{\lambda}_i$ $i=1,2,\dots,k$, es un máximo.

Por lo que el estimador máximo verosímil de $S(t)$ es

$$\hat{S}(t) = \prod_{i|t_i < t} \left(1 - \frac{d_i}{n_i}\right)$$

Este es el estimador Kaplan-Meier o estimador producto límite.

Estimador Kaplan–Meier de la función de supervivencia



Varianza del estimador K-M

- $\widehat{\text{var}}(\hat{\lambda}_j) = \frac{\hat{\lambda}_j(1 - \hat{\lambda}_j)}{n_j}$ (Bernoulli)

- $\widehat{\text{var}}(\hat{p}_j) = \frac{\hat{p}_j(1 - \hat{p}_j)}{n_j}$ (Bernoulli)

- Método delta. Sea \mathbf{X} una v.a. con media θ y varianza σ^2 , y sea g una función diferenciable, entonces

$$\mathbb{E}[g(X)] \approx g(\theta),$$

$$\text{Var}[g(X)] \approx g'(\theta)^2 \text{Var}(X) = g'(\theta)^2 \sigma^2$$

Deducción. Primero observemos que, para $t_k \leq t < t_{k+1}$

$$\log(\hat{S}(t)) = \sum_{j=1}^k \log\left(1 - \frac{d_j}{n_j}\right) = \sum_{j=1}^k \log(1 - \hat{q}_j) = \sum_{j=1}^k \log(\hat{p}_j)$$

Entonces

$$\text{Var} \left[\log(\hat{S}(t)) \right] = \text{Var} \left(\sum_{j=1}^k \log(\hat{p}_j) \right) = \sum_{j=1}^k \text{Var}(\log(\hat{p}_j))$$

Aquí se asume que las \hat{p}_j son independientes, y, de hecho, no lo son. Aplicando el método Delta, tenemos

$$\text{Var}(\log(\hat{p}_j)) \approx \left(\frac{1}{\hat{p}_j} \right)^2 \frac{\hat{p}_j(1 - \hat{p}_j)}{n_j} = \frac{1 - \hat{p}_j}{n_j \hat{p}_j}$$

Entonces

$$\text{Var} \left[\log(\hat{S}(t)) \right] \approx \sum_{j=1}^k \frac{1 - \hat{p}_j}{n_j \hat{p}_j}$$

Pero necesitamos la varianza de $S(t)$, no de su logaritmo. Aplicando “de nuez” el método Delta, ahora para $\log(S(t))$

$$\text{Var} [\log(S(t))] = \left(\frac{1}{S(t)} \right)^2 * \text{Var} (S(t))$$

Entonces, despejando $\text{Var} (S(t))$, tenemos

$$\text{Var}(S(t)) \approx S(t)^2 \text{Var} \left[\log(\hat{S}(t)) \right]$$

Ahora, sustituyendo por los estimadores de estas cantidades, tenemos

$$\widehat{\text{Var}}(\hat{S}(t)) \approx \hat{S}(t)^2 \sum_{j=1}^k \frac{1 - \hat{p}_j}{n_j \hat{p}_j}$$

Si le restituimos su valor a \hat{p}_j , obtenemos la fórmula de Greenwood para esta varianza

$$\widehat{\text{Var}}(\hat{S}(t)) \approx \hat{S}(t)^2 \sum_{j=1}^k \frac{d_j}{n_j(n_j - d_j)}$$

Y para cualquier t

$$\widehat{\text{Var}}(\hat{S}(t)) \approx \hat{S}(t)^2 \sum_{j|t_j < t} \frac{d_j}{n_j(n_j - d_j)}$$

Propiedades del estimador K-M

Qué propiedades deseables debería tener un estimador? Insesgamiento, eficiencia, consistencia. Distribución?

- *Insesgamiento*. Si t^* es un punto fijo en el tiempo

$$\mathbb{E}(\hat{S}(t^*)) \rightarrow S(t^*) \quad \text{cuando } n \rightarrow \infty$$

Puntualmente es asintóticamente insesgado. Independiente del patrón de censura observado.

- *Consistencia*

$$\text{Var}(\hat{S}(t^*)) \rightarrow 0, \quad \text{cuando } n \rightarrow \infty$$

Es asintóticamente consistente

- *Eficiencia*

$$\overline{\widehat{\text{Var}}(\hat{S}(t))} \leq \overline{\widehat{\text{Var}}(\tilde{S}(t))}$$

En promedio es asintóticamente eficiente.

Como proceso estocástico: $\{S(t)\}_{t \geq 0}$ es una Martingala, y

$$\hat{S}(t^*) \approx N(S(t^*), \widehat{\text{Var}}[\hat{S}(t^*)]) \quad \text{cuando } n \rightarrow \infty$$

Intervalos de confianza

Se construyen intervalos para cada valor de t donde se evalúe el $K-M$. La forma general es

$$S(t) \in \left(\hat{S}(t) \pm Z_{1-\alpha/2} \sqrt{\widehat{\text{Var}}(\hat{S}(t))} \right)$$

Estos intervalos tienen algunas “broncas” que debemos analizar.

Supongamos que $n = 10$ y $d_0 = 0$ y en t_1 $d_1 = 1$ y tomemos $t_1 < t < t_2$. Entonces

$$\hat{S}(t) = \prod_{i|t_i < t} \left(1 - \frac{d_i}{n_i} \right) = \left(1 - \frac{d_0}{n_0} \right) \left(1 - \frac{d_1}{n_1} \right) = 1 * \left(1 - \frac{1}{10} \right) = \frac{9}{10}$$

calculemos su correspondiente varianza

$$\begin{aligned} \widehat{\text{Var}}(\hat{S}(t)) &\approx \hat{S}(t)^2 \sum_{j|t_j < t} \frac{d_j}{n_j(n_j - d_j)} = \left(\frac{9}{10} \right)^2 \left(\frac{0}{10(10-0)} + \frac{1}{10(10-1)} \right) \approx \left(\frac{9}{10} \right)^2 \frac{1}{100} \\ &= \frac{9^2}{100^2} \Rightarrow \sqrt{\widehat{\text{Var}}(\hat{S}(t))} = \frac{9}{100} \end{aligned}$$

Si construimos el intervalo al 95 % de confianza, entonces $Z_{1-\alpha/2} = Z_{0.975} \approx 2$, y tenemos

$$S(t) \in \left(\cdot, \frac{9}{10} + 2 * \frac{9}{100} = \frac{54}{50} > 1!!! \right)$$

Una transformación que nos puede “echar la mano” es

$$\log(\hat{S}(t)) \approx N \left(\log(S(t)), \sum_{j=1}^k \frac{d_j}{n_j(n_j - d_j)} \right)$$

entonces

$$\log(S(t)) \in \left(\log(\hat{S}(t)) \pm Z_{1-\alpha/2} \sqrt{\sum_{j=1}^k \frac{d_j}{n_j(n_j - d_j)}} \right)$$

y

$$S(t) \in \left(\exp \left\{ \log(\hat{S}(t)) \pm Z_{1-\alpha/2} \sqrt{\sum_{j=1}^k \frac{d_j}{n_j(n_j - d_j)}} \right\} \right)$$

Este intervalo ya no tiene las “broncas” del anterior.

Ya que \hat{S} no es simétrica, no conviene tomar la media como medida de tendencia central, en su lugar, hay que reportar la *mediana*. Definida como el valor de $t = t^*$, tal que

$$\hat{S}(t^*) = 0.5$$

cuya interpretación es: *Tiempo medio de supervivencia en el estudio*.

Uso del K-M para proponer alguna familia exponencial

Modelo exponencial

$$S(t) = e^{-\lambda t} \Rightarrow \log(S(t)) = -\lambda t$$

Entonces, si graficamos $\log(\hat{S}(t))$ vs. t y observamos aproximadamente una línea recta con pendiente negativa, que pasa por el origen, tendremos evidencia empírica para ajustar el modelo exponencial.

Modelo Weibull

$$S(t) = e^{-\lambda t^\gamma} \Rightarrow \log(S(t)) = -\lambda t^\gamma \Rightarrow \log(-\log(S(t))) = \log(\lambda) + \gamma \log(t)$$

Si graficamos $\log(-\log(\hat{S}(t)))$ vs. $\log(t)$ y obtenemos una línea recta con pendiente positiva (estimador de γ) y ordenada al origen, que es un estimador de $\log(\lambda)$, tendremos evidencia empírica de que los tiempos de supervivencia se distribuyen Weibull.

Modelo Log-normal

$$S(t) = 1 - \Phi\left(\frac{\log(t) - \mu}{\sigma}\right) \Rightarrow 1 - S(t) = \Phi\left(\frac{\log(t) - \mu}{\sigma}\right)$$

Si graficamos $1 - \hat{S}(t)$ vs. $\log(t)$ en *papel normal*, y obtenemos una línea recta con pendiente positiva (estimador de $1/\sigma$) y ordenada al origen estimador de $-\mu/\sigma$, habrá evidencia empírica de que el modelo subyacente a nuestros datos es log-normal.

Estimador Nelson-Aalen de la función de riesgo acumulada

Para estimar la función de riesgo acumulada $\Lambda(t)$, una alternativa simple es hacerlo mediante

$$\hat{\Lambda}(t) = -\log(\hat{S}(t))$$

Otra forma, es utilizar directamente el estimador *Nelson-Aalen*

$$\hat{\Lambda}(t) = \sum_{j|t_j < t} \frac{d_j}{n_j}$$

La varianza de este estimador, puede aproximarse a través de

$$\widehat{Var}(-\log(\hat{S}(t)))$$

Breslow sugiere estimar la función de supervivencia por medio de

$$\hat{S}(t) = \exp\{-\hat{\Lambda}(t)\}$$

La construcción de este estimador se fundamenta a través de la teoría de *procesos de conteo*.

Esperemos que nos de tiempo pa' ver esta construcción.

Ejemplo: Kaplan-Meier a mano

Mostremos cómo se construye el estimador K-M a mano. Los datos son dos grupos de pacientes con leucemia. La siguiente tabla muestra los tiempos de supervivencia (en semanas) de estos dos grupos.

Grupos de pacientes con leucemia	
Grupo 1 (n=21) tratamiento	Grupo 2 (n=21) placebo
6	1
6	1
6	2
7	2
10	3
13	4
16	4
22	5
23	5
6+	8
9+	8
10+	8
11+	8
17+	11
19+	11
20+	12
25+	12
32+	15
32+	17
34+	22
35+	23

+ indica una observación censurada

Las tablas ordenadas por el tiempo de supervivencia son

Tablas tiempos de falla ordenados							
tratamiento				placebo			
$t_{(j)}$	n_j	d_j	c_j	$t_{(j)}$	n_j	d_j	c_j
0	21	0	0	0	21	0	0
6	21	3	1	1	21	2	0
7	17	1	1	2	19	2	0
10	15	1	2	3	17	1	0
13	12	1	0	4	16	2	0
16	11	1	3	5	14	2	0
22	7	1	0	8	12	4	0
23	6	1	5	11	8	2	0
				12	6	2	0
				15	4	1	0
				17	3	1	0
				22	2	1	0
				23	1	1	0

Obsérvese que el grupo placebo no presentó censuras, por lo que su estimador K-M coincide con la función de supervivencia empírica. La tabla correspondiente es

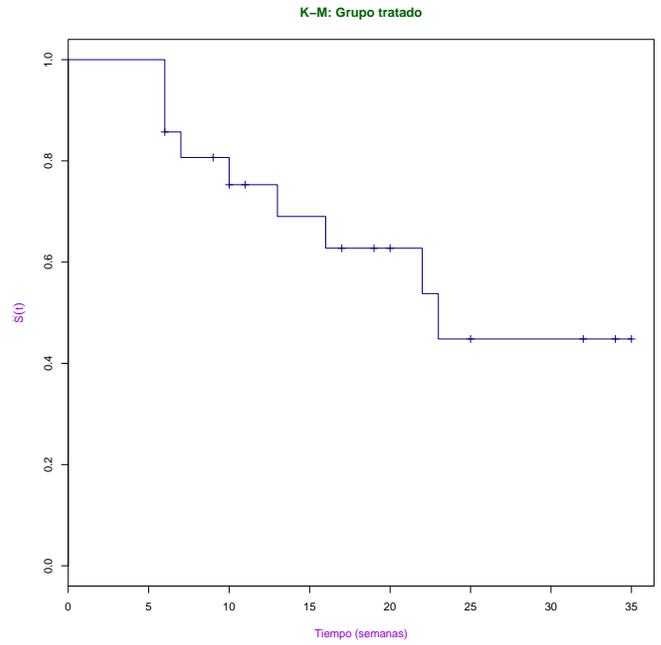
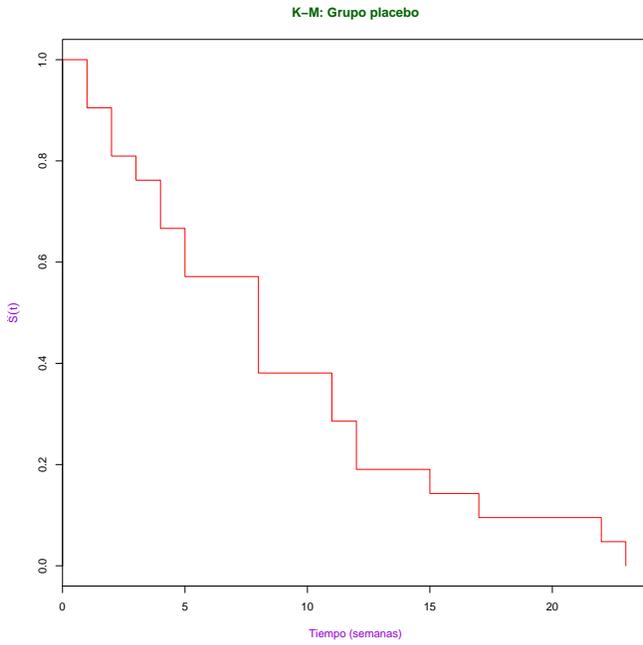
Función de supervivencia empírica				
Grupo placebo				
$t_{(j)}$	n_j	d_j	c_j	$\hat{S}(t_{(j)})$
0	21	0	0	1
1	21	2	0	$19/21 = 0.90$
2	19	2	0	$17/21 = 0.81$
3	17	1	0	$16/21 = 0.76$
4	16	2	0	$14/21 = 0.67$
5	14	2	0	$12/21 = 0.57$
8	12	4	0	$8/21 = 0.38$
11	8	2	0	$6/21 = 0.29$
12	6	2	0	$4/21 = 0.19$
15	4	1	0	$3/21 = 0.14$
17	3	1	0	$2/21 = 0.10$
22	2	1	0	$1/21 = 0.05$
23	1	1	0	$0/21 = 0.00$

Para el grupo tratado, como sí se tienen observaciones censuradas, la función de supervivencia es propiamente el K-M. Que se calcula de la siguiente manera

Estimador Kaplan-Meier

tratamiento

$t_{(j)}$	n_j	d_j	c_j	$\hat{S}(t_{(j)})$
0	21	0	0	1
6	21	3	1	$1 \times (1 - 3/21) = 0.857$
7	17	1	1	$1 \times (1 - 3/21) \times (1 - 1/17) = 0.857 * 0.941 = 0.806$
10	15	1	2	$0.806 \times (1 - 1/15) = 0.806 * 0.933 = 0.752$
13	12	1	0	$0.752 \times (1 - 1 - 12) = 0.752 * 0.917 = 0.690$
16	11	1	3	$0.690 \times (1 - 1/11) = 0.690 * 0.909 = 0.627$
22	7	1	0	$0.627 \times (1 - 1/7) = 0.627 * 0.857 = 0.537$
23	6	1	5	$0.537 \times (1 - 1/6) = 0.537 * 0.833 = 0.447$



Comparación de poblaciones mediante técnicas no paramétricas

En la estadística usual, uno de los problemas más importantes es la comparación de poblaciones. Por ejemplo, en un estudio de supervivencia se puede estar interesado en estudiar si la aplicación de algún tratamiento genera una diferencia de supervivencia en la población; o si el nivel educativo de un sujeto, reduce el tiempo requerido para encontrar trabajo, o prolonga el tiempo en el que está empleado. Una manera común de hacer las comparaciones entre estas poblaciones, es a través del planteamiento de una prueba paramétrica, que requiere el cumplimiento de varios supuestos para su implementación. En supervivencia, dado que las observaciones pueden estar censuradas, es más difícil construir pruebas paramétricas para realizar esta comparación. Una alternativa más conveniente es realizar una prueba *no paramétrica* para este fin.

Prueba log-rank, Mantel-Haenszel, Cox-Mantel

Consideremos primero el desarrollo de esta prueba para dos poblaciones únicamente, Grupo I y Grupo II. Supongamos que hay k tiempos de falla distintos en la *muestra total* donde se combinan los dos grupos: $t_{(1)} < t_{(2)} < \dots < t_{(k)}$, y que al tiempo $t_{(j)}$, d_{1j} individuos del grupo I, y d_{2j} del grupo II, fallan, $j=1,2,\dots,k$. Suponga también que hay n_{1j} y n_{2j} individuos en riesgo del grupo I y II, respectivamente, justo antes del tiempo $t_{(j)}$. Por lo tanto, hay $d_j = d_{1j} + d_{2j}$ fallas totales, de $n_j = n_{1j} + n_{2j}$ individuos en riesgo, en $t_{(j)}$, $j=1,2,\dots,k$.

Queremos probar la hipótesis de igualdad de supervivencia entre las poblaciones, i.e.,

$$H_0 : S_1(t) = S_2(t), \forall t > 0 \quad \text{vs.} \quad H_a : S_1(t) \neq S_2(t) \text{ p.a. } t > 0$$

Una manera de hacer esta prueba, es considerar la diferencia entre el número de fallas observadas en los dos grupos, que ocurren en cada uno de los tiempos de falla, y el número esperado bajo la hipótesis nula (igualdad de experiencia de supervivencia).

Coloquemos esta información en una *tabla de contingencia*, para cada tiempo de falla distinto, con los renglones dados por el grupo de pertenencia, y las columnas por el *status* de supervivencia.

Si los marginales totales en la tabla ($d_j, n_j - d_j$) se toman como fijos, y la hipótesis nula es cierta, las entradas de la tabla quedan determinadas conociendo únicamente d_{1j} , el número

Tabla de contingencia

Grupo	#fallas en t_j	# de sobrevivientes después de t_j	#ind. en ries. justo antes de t_j
I	d_{1j}	$n_{1j} - d_{1j}$	n_{1j}
II	d_{2j}	$n_{2j} - d_{2j}$	n_{2j}
Total	d_j	$n_j - d_j$	n_j

de fallas en t_j del grupo I. Por lo tanto, d_{1j} es una v.a. con distribución *hipergeométrica*. Concretamente

$$d_{1j} \sim \text{hipergeométrica } (n_j, d_j, n_{1j})$$

Entonces

$$Pr(D_{1j} = d_{1j}) = \frac{\binom{d_j}{d_{1j}} \binom{n_j - d_j}{n_{1j} - d_{1j}}}{\binom{n_j}{n_{1j}}}$$

El hecho de que las poblaciones tengan experiencias de vida iguales, se puede pensar como si existiera un factor que suponemos las hace diferentes, pero este factor *no tiene impacto en las supervivencias de los grupos*. Bajo esta argumentación, la prueba log-rank se puede plantear como una prueba de *independencia entre las supervivencia y este factor*. Como mencionamos, la tabla queda totalmente especificada conociendo d_{1j} , por lo que realizaremos la prueba utilizando esta variable. Dado que d_{1j} tiene distribución hipergeométrica, tenemos:

$$e_{1j} = \mathbb{E}(D_{1j}) = n_{1j} \frac{d_j}{n_j} \text{ y } V_{1j} = \mathbb{V}(D_{1j}) = \frac{n_{1j} n_{2j} d_j (n_j - d_j)}{n_j^2 (n_j - 1)}$$

Entonces, debemos de considerar esta comparación para cada tiempo t_j , $j=1,2,\dots,k$. Por lo tanto, la estadística de prueba es:

$$U_L = \sum_{j=1}^k (d_{1j} - e_{1j})$$

Observemos que $\mathbb{E}(U_L) = 0$ y $\mathbb{V}(U_L) = \sum_{j=1}^k V_{1j}$. Entonces, si las supervivencias en las dos poblaciones son iguales

$$U_L = \frac{\sum_{j=1}^k (d_{1j} - e_{1j})}{\sqrt{\sum_{j=1}^k V_{1j}}} \stackrel{a}{\sim} N(0, 1)$$

Equivalentemente

$$U_L^2 \stackrel{a}{\sim} \chi_{(1)}^2$$

Comparación de m poblaciones

Ahora queremos realizar la prueba

$$H_0 : S_1(t) = S_2(t) = \dots = S_m(t) \text{ vs. } H_a : S_r \neq S_s \text{ p.a. } r \neq s, r, s=1, \dots, m$$

Para realizar esta prueba, denotemos por \underline{d}_j el vector de fallas al tiempo t_j , $j=1, \dots, m$, con vector de medias $E(\underline{d}_j)$ y matriz de varianza y covarianza, cuya entrada (i, k) , está dada por

$$\begin{aligned} \mathbb{V}(d_{ij}) &= \frac{n_{1j}n_{2j}d_j(n_j - d_j)}{n_j^2(n_j - 1)}, y \\ \text{Cov}(d_{ji}, d_{jk}) &= -\frac{n_{ji}n_{jk}d_j(n_j - d_j)}{n_j^2(n_j - 1)}, i \neq k \end{aligned}$$

Si sumamos sobre todos los tiempos de falla, obtenemos

$$D = \sum_{j=1}^k \{\underline{d}_j - \mathbb{E}(\underline{d}_j)\}, \quad y \quad V = \sum_{j=1}^k \mathbb{V}(\underline{d}_j)$$

Mantel-Haenszel proponen probar la hipótesis de k supervivencias iguales, a través de la forma cuadrática

$$Q = D'V^{-1}D$$

que, bajo H_0 , se distribuye asintóticamente como una $\chi^2_{(m-1)}$. Con V^{-1} la inversa generalizada de V .

Prueba de Wilcoxon

Esta prueba es muy similar a la log-rank, la diferencia es que la Wilcoxon pondera las diferencias de la log-rank a través del número de sujetos en riesgo en cada tiempo t_j observado. Ya que por lo general la población en riesgo va decreciendo conforme el tiempo en el estudio crece, esta estadística proporciona *menos peso en las colas de la distribución* de la que tiene la log-rank. La estadística es

$$U_W = \sum_{j=1}^k n_j (d_{1j} - e_{1j})$$

Con $\mathbb{E}(U_W) = 0$ y $V_W = \mathbb{V}(U_W) = \sum_{j=1}^k n_j^2 V_{1j}$. Entonces, si las supervivencias son iguales

$$W_W = \frac{U_W^2}{V_W} \stackrel{a}{\sim} \chi^2_{(1)}$$

Se recomienda usar *log-rank* si se tiene evidencia empírica de *riesgos proporcionales* en las poblaciones. La extensión de esta prueba para $m > 2$ poblaciones, es similar a la log-rank.

Estimación para los modelos paramétricos

Introducción

- $T \sim f(t; \theta)$
- θ puede ser un vector, i.e., $\theta = (\theta_1, \theta_2, \dots, \theta_k)'$
- Se conoce la forma funcional de f pero se desconoce θ
- Inferencias sobre θ basadas en una muestra aleatoria
- Estimación puntual y por intervalos
- Pruebas de hipótesis
- Elemento fundamental: *MÁXIMA VEROSIMILITUD*

La verosimilitud para los distintos tipos de censura

De manera general, supongamos que $L_i, i=1,2,\dots,n$ son los tiempos máximos de observación para los sujetos, y T_i los tiempos de falla observados. Entonces, la información observada es

$$\begin{cases} T_i, & \text{si } T_i \leq L_i \\ L_i, & \text{si } T_i > L_i \end{cases}$$

i.e. observamos $Y_i = \min(T_i, L_i)$, y

$$\mathbb{P}(Y_i = L_i) = \mathbb{P}(T_i > L_i) = S(L_i), \quad f_Y(t) = f_T(t), \quad \text{si } t < L_i$$

Entonces

$$f_Y(t_i) \begin{cases} f(t_i), & \text{si } t_i \leq L_i \\ S(L_i), & \text{si } t_i > L_i \end{cases}$$

Entonces

$$f_Y(t_i) = f(t_i)^{\delta_i} S(L_i)^{1-\delta_i}, \quad \delta_i = \begin{cases} 1, & \text{si } t_i \leq L_i \\ 0, & \text{si } t_i > L_i \end{cases}$$

La verosimilitud (la “vero”pa’ los cuates), queda como

$$L(\theta, \underline{t}) = \prod_{i=1}^n f_Y(t_i) = \prod_{i=1}^n f(t_i)^{\delta_i} S(t_i)^{1-\delta_i}$$

por supuesto, en la supervivencia $t_i = L_i$.

Censura tipo I

Recordemos que en este caso se fija un tiempo máximo de observación $L = t_c$, y todas las observaciones que no presentaron la falla antes de este tiempo, resultan observaciones censuradas. Por lo tanto, la verosimilitud es

$$\begin{aligned}
 L(\theta, \underline{t}) &= \prod_{i=1}^n f(t_i)^{\delta_i} S(t_c)^{1-\delta_i} \\
 &= \prod_{i=1}^n f(t_i)^{\delta_i} S(t_c)^{n-\sum \delta_i} \\
 &= \prod_{i=1}^n f(t_i)^{\delta_i} S(t_c)^{n-k}
 \end{aligned}$$

con $\sum_{i=1}^n \delta_i = k$.

Censura tipo II

Para esta censura observamos k fallas de n posibles. Observamos $T_{(1)} < T_{(2)} < \dots < T_{(k)}$ tiempos de falla, y para los siguientes $n - k$ individuos registramos $T_{(k)}$ como su tiempo de censura. Dada la información de las primeras k fallas, las $n - k$ restantes son fijas (no aleatorias). Por lo que la verosimilitud está dada por la densidad conjunta de las estadísticas de orden

$$\begin{aligned}
 L(\theta, \underline{t}) &= f_{1,2,\dots,k}(t_1, t_2, \dots, t_k) = \frac{n!}{(n-k)!k!} \prod_{i=1}^k f(t_i) [1 - F(t_k)]^{n-k} \\
 &= \frac{n!}{(n-k)!k!} \prod_{i=1}^k f(t_i) [S(t_k)]^{n-k} \\
 &= \frac{n!}{(n-k)!k!} \prod_{i=1}^k f(t_i)^{\delta_{(i)}} [S(t_k)]^{1-\delta_{(i)}}
 \end{aligned}$$

$\delta_{(1)} = \dots = \delta_{(k)} = 1$, $\delta_{(k+1)} = \dots = \delta_{(n)} = 0$.

Y cómo se ve esta verosimilitud cuando se tiene censura aleatoria?

Censura aleatoria

Recordemos que en este tipo de censura el investigador no tiene ningún control sobre la misma. Definamos las variables aleatorias

T_i : Tiempo de falla, y C_i : Tiempo de censura. Observamos $Y_i = \min(T_i, C_i)$. Supondremos $T_i \perp C_i$, lo que implica que la censura es *no informativa*. La muestra observada es

$$(Y_i, \delta_i) \quad \delta_i = \begin{cases} 1, & \text{si } Y_i = T_i \\ 0, & \text{si } Y_i = C_i \end{cases}$$

Supongamos que f y S son las funciones de densidad y supervivencia asociadas a T , y h y H las correspondientes asociadas a C . Entonces encontremos las funciones de densidad de la muestra observada, Y . Para hacerlo, debemos de considerar los dos casos: $\delta = 0$ y $\delta = 1$.

- Caso $\delta = 0$.

Utilizando probabilidad condicional, tenemos

$$\begin{aligned} f_Y(y) = f_{(Y,\delta)}(y, \delta) &= f_{(Y,\delta)}(y, \delta = 0) \\ &= f_{(Y|\delta)}(y|\delta = 0) \mathbb{P}(\delta = 0) \\ &= h(t) \mathbb{P}(T > C) \\ &= h(t)S(t) \end{aligned}$$

- Caso $\delta = 1$.

$$\begin{aligned} f_Y(y) = f_{(Y,\delta)}(y, \delta) &= f_{(Y,\delta)}(y, \delta = 1) \\ &= f_{(Y|\delta)}(y|\delta = 1) \mathbb{P}(\delta = 1) \\ &= f(t) \mathbb{P}(C > T) \\ &= f(t)H(t) \end{aligned}$$

Juntando estas dos expresiones, podemos escribir esta densidad como

$$\begin{aligned} f_{(Y,\delta)}(y, \delta) &= (f(t)H(t))^\delta (S(t)h(t))^{1-\delta} \\ &= f(t)^\delta S(t)^{1-\delta} h(t)^{1-\delta} H(t)^\delta \end{aligned}$$

Por lo que la verosimilitud queda como

$$L(\theta, \underline{t}) = \prod_{i=1}^n f(t_i)^{\delta_i} S(t_i)^{1-\delta_i} \prod_{i=1}^n h(t_i)^{1-\delta_i} H(t_i)^{\delta_i}$$

Dado el supuesto de independenciam entre los tiempos de falla y la censura, asumimos que el segundo término, que sólo depende de las funciones para la censura, no involucra parámetros de interés para *el tiempo de falla* que es nuestro objetivo de estudio. Por lo tanto, la verosimilitud es proporcional

$$\prod_{i=1}^n f(t_i)^{\delta_i} S(t_i)^{1-\delta_i}$$

Si observamos detenidamente estas tres verosimilitudes, notaremos que su *kernel* es igual, esto es, no importa qué censura esté presente en la muestra, la verosimilitud es esencialmente ¡la misma!.

Estimación paramétrica para la distribución exponencial (θ)

$$f(t) = \frac{1}{\theta}e^{-t/\theta}, \quad S(t) = e^{-t/\theta}.$$

$$\begin{aligned} L(\theta, \underline{t}) &= \prod_{i=1}^n f(t_i)^{\delta_i} S(t_i)^{1-\delta_i} \\ &= \prod_{i=1}^n \left(\frac{1}{\theta} e^{-t_i/\theta} \right)^{\delta_i} (e^{-t_i/\theta})^{1-\delta_i} \\ &= \prod_{i=1}^n \left(\frac{1}{\theta} \right)^{\delta_i} e^{-t_i/\theta} \\ &= \theta^{-\sum_{i=1}^n \delta_i} e^{-1/\theta \sum_{i=1}^n t_i} \end{aligned}$$

De donde obtenemos la log-verosimilitud

$$\log L(\theta, \underline{t}) = -\sum_{i=1}^n \delta_i \log(\theta) - \frac{1}{\theta} \sum_{i=1}^n t_i$$

realizando el usual proceso para encontrar el estimador de θ , tenemos

$$\hat{\theta} = \frac{\sum_{i=1}^n t_i}{\sum_{i=1}^n \delta_i}$$

Ojo: $\sum_{i=1}^n t_i$ es el total de los tiempos observados, sean fallas o censuras; mientras que $\sum_{i=1}^n \delta_i$ es

el total de fallas observadas. Sin censura, esta suma es n y $\hat{\theta} = \bar{t}$ usual.

Y el proceso para hacer inferencias es igual para estos tres tipos de censura?

Veamos primero que para el caso de censura tipo *II*, en número de fallas $\sum \delta_i$ es fijo, ya que fue fijado por el investigador. Para los tipos restantes (*I* y aleatoria) este número es *aleatorio*, este hecho es importante ya que implica que sólo es posible hacer *inferencias exactas* para

la censura tipo II. En los otros dos casos, habrá que hacerlas aproximadas o asintóticas.

Inferencia censura tipo II

Tenemos $\sum t_i$ aleatorio y $\sum \delta_i = k$ fijo, y

$$\sum_{i=1}^n t_i = \sum_{i=1}^k t_i + (n - k)t_{(k)} \leftarrow \text{no aleatorio}$$

Sabemos que

$$\sum_{i=1}^k t_i \sim \text{Gamma}(k, 1/\theta) \Rightarrow 2\theta \sum_{i=1}^k t_i \sim \chi_{(2k)}^2$$

Esta estadística es una pivotal para θ , con ella podemos construir intervalos de confianza y realizar pruebas de hipótesis sobre este parámetro.

Inferencias para censura tipo I y aleatoria

En estos dos casos: $\sum t_i$ y $\sum \delta_i$ son ambos aleatorios, así que haremos inferencias asintóticas.

De las propiedades del estimador máximo verosímil, tenemos

$$\hat{\theta}_{M.V.} \stackrel{a}{\sim} N(\theta, \widehat{\text{var}}(\hat{\theta}))$$

con

$$\widehat{\text{var}}(\hat{\theta}) = \frac{1}{-E \left\{ \left(\frac{\partial^2}{\partial^2 \theta} \log L(\theta, \underline{t}) \right) \right\} \Big|_{\theta=\hat{\theta}}} : \text{Información observada de Fisher}$$

Entonces

$$\frac{\hat{\theta}_{M.V.} - \theta}{\sqrt{\widehat{\text{var}}(\hat{\theta})}} \approx N(0, 1)$$

- Pivotal asintótica para construir intervalos de confianza.
- Estadística asintótica para realizar pruebas de hipótesis.

Si suponemos $\sum \delta_i = k$, para el caso de la exponencial

$$\widehat{var}(\hat{\theta}) = \frac{\hat{\theta}^2}{k}$$

Algoritmos numéricos de estimación

Mostraremos la forma de estimar el parámetro de una exponencial en presencia de censura, a través del algoritmo **EM**.

Algoritmo EM (Dempster, Laird, and Rubin (1977))

El algoritmo EM es un eficiente proceso iterativo para calcular estimadores máximo verosímiles (ML), en presencia de datos faltantes o desconocidos. En el proceso de estimación ML, deseamos estimar los parámetros del modelo para los que los datos observados sean más creíbles o probables.

Cada iteración del algoritmo EM consiste de dos pasos:

i) El *paso-E* (Esperanza) En este paso, los datos faltantes se estiman dados los datos observados y el estimador actual de los parámetros del modelo. Esto se realiza utilizando la esperanza condicional, que justifica la elección del nombre del algoritmo. Una vez realizado el paso-E, esta función de verosimilitud se maximiza bajo el supuesto que los datos faltantes son conocidos. Los datos estimados faltantes generados por el paso-E, se utilizan en lugar de los datos faltantes actuales. La convergencia está garantizada ya que el algoritmo incrementa la verosimilitud en cada iteración.

Situaciones de Datos Incompletos:

Explícitas

- Datos Faltantes
- Datos truncados y/o censurados
- Datos agrupados

Implícitas

- Mezclas
- Efectos Aleatorios
- Clases Latentes
- Estructuras de Variables Latentes

El algoritmo EM puede explicarse de una manera fácil en términos de una muestra que

consiste de dos componentes, uno observado y uno no observado o faltante.

Muchas de las aplicaciones comunes del algoritmo EM involucran problemas con datos faltantes, pero no es necesario que siempre ocurra de esta manera. Algunas veces un algoritmo EM puede construirse con base a una “variable aleatoria artificial” para suplementar los datos observados.

Descripción del algoritmo

Sea $\mathbf{Y} = (\mathbf{U}, \mathbf{V})$ y supongamos que tenemos observaciones de \mathbf{U} pero no de \mathbf{V} . Entonces, existen *dos verosimilitudes*: una basada en la muestra completa (pero desconocida) y la otra basada sólo en la muestra observada. Queremos estimar el parámetro θ , inmerso en la distribución de ambos componentes de \mathbf{Y} .

Ya que no tenemos todas las observaciones, no podemos escribir exactamente la verosimilitud. Entonces, podemos utilizar *provisionalmente* el valor $\theta^{(k)}$ (la k-ésima iteración del algoritmo) para aproximar esta verosimilitud completa a través del valor esperado de \mathbf{V} . La aproximación será, generalmente, una función de θ y $\theta^{(k)}$. Entonces, maximizaremos esta aproximación con respecto a θ para obtener el siguiente valor $\theta^{(k+1)}$ del proceso iterativo.

Denotemos por $L_c(\theta; u, v)$ y ℓ_{L_c} la verosimilitud y la log-verosimilitud de la muestra completa. Entonces, la verosimilitud para la muestra observada \mathbf{U} es

$$L(\theta; u) = \int L_c(\theta; u, v) dv$$

y, por supuesto, $\ell_L(\theta; u) = \log L(\theta; u)$. Esta es la función que debemos maximizar, si es que el máximo existe. El problema, sin embargo, *es determinar esta función*; es decir, promediar sobre \mathbf{V} . (En otras palabras, la integral está dada, pero no sabemos qué integrar). Obviamente, el promedio sobre \mathbf{V} es el valor esperado respecto a la distribución marginal de \mathbf{V} . Este es un problema estándar, y estimaremos esta esperanza utilizando los valores observados \mathbf{U} y un valor provisional de θ .

Iniciamos con un valor provisional de θ , llamémoslo $\theta^{(0)}$. Dado cualquier valor provisional $\theta^{(k)}$, calculamos el siguiente valor provisional $\theta^{(k+1)}$ con la condición de que no decremente

la verosimilitud.

Entonces, el algoritmo EM para aproximar la maximización de $L(\theta; u)$, tiene dos pasos alternados. Los pasos se iteran hasta lograr la convergencia.

- Paso-E: calcular $Q^{(k)}(u, \theta^{(k)}) = \mathbb{E}_{\mathbf{V}|u, \theta^{(k)}} (\ell_{L_c}(\theta; u, \mathbf{V}))$
- Paso-M: Determinar el valor de $\theta = \theta^{(k+1)}$ que maximice $Q^{(k)}(u, \theta^{(k)})$.

Entonces, la sucesión $\theta^{(1)}, \theta^{(2)}, \theta^{(3)}, \dots$ converge a un máximo local de la verosimilitud observada $L(\theta; u)$, bajo condiciones generales (incluida, por supuesto, la no existencia de un máximo local alrededor de θ). El algoritmo EM puede ser muy lento para converger.

Exponencial censurada

Comentamos que uno de los casos en los que los datos incompletos se presentan de manera explícita, es cuando se tienen datos censurados. Hagamos el desarrollo de el algoritmo EM para datos exponenciales censurados. El contexto del problema es como sigue

Se registran tiempos de vida útil de n bulbos. Se supone que este tiempo se distribuye $exp(\theta)$. Supongamos que tenemos k tiempos completos de falla y $n - k$ tiempos censurados. Desarrollar el algoritmo EM para este caso. Utilizaremos la parametrización,

$$f(t|\theta) = \frac{1}{\theta} e^{-t/\theta}, \quad t > 0, \theta > 0$$

Entonces, es claro que en este caso los datos completos \mathbf{Y} son los k tiempos de falla observados y los “faltantes” son los $n - k$ datos censurados. Supongamos, sin pérdida de generalidad, que los primeros k tiempos corresponden a los datos observados y los siguientes $n - k$ a los censurados. Entonces, el vector de datos observados tiene la forma:

$$\mathbf{X} = (\mathbf{T}, \mathbf{Z}) = (t_1, \dots, t_k, z_1 = t_{k+1}, \dots, z_{n-k} = t_n)$$

Entonces, la verosimilitud completa es

$$\begin{aligned}
L_c(\theta|\mathbf{T}, \mathbf{Z}) &= \prod_{i=1}^k f(t_i|\theta) \prod_{i=k+1}^n f(z_i|\theta) \\
&= \prod_{i=1}^k \frac{1}{\theta} e^{-t_i/\theta} \prod_{i=k+1}^n \frac{1}{\theta} e^{-z_i/\theta}
\end{aligned}$$

con respectiva log-verosimilitud dada por

$$\begin{aligned}
\ell_c(\theta|\mathbf{T}, \mathbf{Z}) &= \sum_{i=1}^k \left(-\log(\theta) - \frac{t_i}{\theta} \right) + \sum_{i=k+1}^n \left(-\log(\theta) - \frac{z_i}{\theta} \right) \\
&= -n\log(\theta) - \frac{1}{\theta} \left(\sum_{i=1}^k t_i + \sum_{i=k+1}^n z_i \right)
\end{aligned}$$

Entonces, el paso E del algoritmo, consiste en remplazar los valores desconocidos, z_i , por sus correspondiente esperanzas condicionales, dados los datos: (\mathbf{T}, θ^k) . El problema se reduce entonces a, si c es el valor censurado de una exponencial, encontrar

$$\mathbb{E}(c|\mathbf{T}, \theta^{(k)})$$

Para calcular esta esperanza, recurriremos a algunos resultados sobre la distribución exponencial.

- Propiedad de pérdida de memoria de la exponencial. Para $T \sim \exp(\theta)$ y $0 < t_0 < t_1$

$$\mathbb{P}[T > t_1 | T > t_0] = \frac{\mathbb{P}[T > t_1 \wedge T > t_0]}{\Pr[T > t_0]} = \frac{\mathbb{P}[T > t_1]}{\Pr[T > t_0]} = \frac{e^{-t_1/\theta}}{e^{-t_0/\theta}} = e^{-(t_1-t_0)/\theta} \sim \exp(\theta)$$

'Ora sí. Sea \mathbf{c} el tiempo de censura observado en el i -ésimo individuo. Queremos calcular: $\mathbb{E}(T|T \geq \mathbf{c})$, con $T \sim \exp(\theta)$. Primeramente $\mathbb{P}[T|T \geq c] = e^{-\theta(T-c)}$

$$\begin{aligned}
\Rightarrow \mathbb{E}(T|T \geq c) &= \int_c^\infty \frac{t}{\theta} e^{-(t-c)/\theta} dt \\
&= \frac{e^{c/\theta}}{\theta} \int_c^\infty t e^{-t/\theta}, \text{ (Integrando por partes)} \\
&= \frac{e^{c/\theta}}{\theta} [\theta e^{-c/\theta} (c + \theta)] \\
&= c + \theta
\end{aligned}$$

Entonces, $\mathbb{E}(c|\mathbf{T}, \theta^{(k)}) = c + \theta^{(k)}$. Observemos que los $n - k$ valores de censura c , son los valores censurados observados t_{k+1}, \dots, t_n . Por lo que, la función a maximizar es

$$\begin{aligned}
Q^{(k)}(\theta, \theta^{(k)}) &= -n \log(\theta) - \frac{1}{\theta} \left(\sum_{i=1}^k t_i - \sum_{i=k+1}^n t_i + \theta^{(k)} \right) \\
&= -n \log(\theta) - \frac{1}{\theta} \left(\sum_{i=1}^n t_i + (n - k) \theta^{(k)} \right)
\end{aligned}$$

El paso M que consiste en resolver la igualdad

$$\frac{\partial Q^{(k)}(\theta, \theta^{(k)})}{\partial \theta} = 0$$

tiene la solución analítica

$$\theta^{(k+1)} = \frac{\sum_{i=1}^n t_i + (n - k) \theta^{(k)}}{n}$$

Finalmente, iteramos estos dos pasos hasta obtener convergencia.

Estimación paramétrica para la distribución Weibull(γ, λ)

$f(t) = \gamma \lambda t^{\gamma-1} \exp\{-\lambda t^\gamma\}$, $S(t) = \exp\{-\lambda t^\gamma\}$. Entonces

$$L(\gamma, \lambda, \underline{t}) = \prod_{i=1}^n f(t_i)^{\delta_i} S(t_i)^{1-\delta_i} \text{ y}$$

$$\begin{aligned} \log L(\gamma, \lambda, \underline{t}) &= \sum_{i=1}^n \delta_i \log(f(t_i)) + (1 - \delta_i) \log(S(t_i)) \\ &= \sum_{i=1}^n \delta_i (\log(\gamma) + \log(\lambda) + (\gamma - 1) \log(t_i) - \lambda t_i^\gamma) - (1 - \delta_i) \lambda t_i^\gamma \\ &= \log(\gamma) \sum_{i=1}^n \delta_i + \log(\lambda) \sum_{i=1}^n \delta_i + (\gamma - 1) \sum_{i=1}^n \delta_i \log(t_i) - \lambda \sum_{i=1}^n t_i^\gamma \end{aligned}$$

Las derivadas parciales son

$$\begin{aligned} \frac{\partial \log L(\gamma, \lambda, \underline{t})}{\partial \gamma} &= \frac{\sum_{i=1}^n \delta_i}{\gamma} + \sum_{i=1}^n \delta_i \log(t_i) - \lambda \sum_{i=1}^n t_i^\gamma \log(t_i) = 0 \\ \frac{\partial \log L(\gamma, \lambda, \underline{t})}{\partial \lambda} &= \frac{\sum_{i=1}^n \delta_i}{\lambda} - \sum_{i=1}^n t_i^\gamma = 0 \end{aligned}$$

De la parcial con respecto a λ , tenemos

$$\hat{\lambda} = \frac{\sum_{i=1}^n \delta_i}{\sum_{i=1}^n t_i^\gamma} \quad ! \text{ Depende de } \gamma !$$

y substituyendo en la parcial con respecto a γ

$$\frac{\sum_{i=1}^n \delta_i}{\gamma} + \sum_{i=1}^n \delta_i \log(t_i) - \frac{\sum_{i=1}^n \delta_i}{\sum_{i=1}^n t_i^\gamma} \sum_{i=1}^n t_i^\gamma \log(t_i) = 0$$

Esta es una ecuación que sólo depende de γ . Para obtener $\hat{\gamma}$, hay que resolverla por métodos numéricos (Newton-Raphson), ya que es una ecuación no lineal. Después

$$\hat{\lambda} = \frac{\sum_{i=1}^n \delta_i}{\sum_{i=1}^n t_i^{\hat{\gamma}}}$$

aplicando la propiedad de invarianza del estimador máximo verosímil. Obtener las varianzas exactas de estos estimadores está “ en chino”, así que “ nos conformaremos” con sus varianzas asintóticas. La matriz de información observada de Fisher es

$$I(\gamma, \lambda) = \begin{bmatrix} \frac{\partial^2 \log L}{\partial \gamma^2} & \frac{\partial^2 \log L}{\partial \gamma \lambda} \\ \frac{\partial^2 \log L}{\partial \lambda \gamma} & \frac{\partial^2 \log L}{\partial \lambda^2} \end{bmatrix}_{(\gamma=\hat{\gamma}, \lambda=\hat{\lambda})}$$

Los elementos de la diagonal de la inversa de esta matriz, corresponden a las varianzas asintóticas de estos estimadores.

Para las distribuciones log-normal y gamma, dado que sus funciones de supervivencia no tienen expresiones cerradas, sólo es posible maximizar la correspondiente verosimilitud a través de métodos numéricos, para realizar inferencias asintóticas.

Algoritmo Newton-Raphson para la distribución Weibull con censura

Uno de los métodos más utilizados para la optimización en situaciones estadísticas, es el método de Newton-Raphson. Este método se basa en la aproximación cuadrática de la función que queremos optimizar. El óptimo de la aproximación (que es fácil de calcular) proporciona una aproximación del óptimo de la función. Si esta aproximación no está suficientemente cerca a este óptimo, se calcula una nueva aproximación y se repite el proceso sucesivamente hasta obtener convergencia. En este caso, lo que debemos aproximar a través de este algoritmo, es la solución de la *log-verosimilitud*, $\ell(\theta)$, para estimar los parámetros de la Weibull con censura. La t -ésima iteración del algoritmo es

$$\theta^{(t+1)} = \theta^{(t)} - \frac{\ell'(\theta^{(t)})}{\ell''(\theta^{(t)})}$$

Utilizaremos una parametrización diferente a la que presentamos antes en este modelo Weibull (esta es la parametrización que utiliza \mathbf{R}). La densidad es ahora

$$f(t; \alpha, \beta) = \frac{\beta}{\alpha} \left(\frac{t}{\alpha}\right)^{\beta-1} \exp\left\{-\left(\frac{t}{\alpha}\right)^\beta\right\}$$

Cuya log-verosimilitud (nuestra función a maximizar) incorporando datos censurados es

$$\ell(\alpha, \beta, \underline{t}) = \log(\alpha) \sum_{i=1}^n \delta_i - \log(\beta) \sum_{i=1}^n \delta_i + (\beta - 1) \sum_{i=1}^n \delta_i \log(t_i) - \sum_{i=1}^n \left(\frac{t_i}{\alpha}\right)^\beta$$

Las derivadas necesarias para esta implementación son

$$\begin{aligned}
\frac{\partial \ell(\alpha, \beta, \underline{t})}{\partial \alpha} &= \frac{-\beta \sum_{i=1}^n \delta_i}{\alpha} + \frac{\beta}{\alpha} \sum_{i=1}^n \left(\frac{t_i}{\alpha}\right)^\beta \\
\frac{\partial \ell(\alpha, \beta, \underline{t})}{\partial \beta} &= \frac{\sum_{i=1}^n \delta_i}{\beta} - \log(\alpha) \sum_{i=1}^n \delta_i + \sum_{i=1}^n \delta_i \log(t_i) - \sum_{i=1}^n \left(\frac{t_i}{\alpha}\right)^\beta \log\left(\frac{t_i}{\alpha}\right) \\
-\frac{\partial^2 \ell(\alpha, \beta, \underline{t})}{\partial \alpha^2} &= \frac{-\beta \sum_{i=1}^n \delta_i}{\alpha^2} + \frac{\beta(\beta+1)}{\alpha^2} \sum_{i=1}^n \left(\frac{t_i}{\alpha}\right)^\beta \\
-\frac{\partial^2 \ell(\alpha, \beta, \underline{t})}{\partial \alpha \partial \beta} &= \frac{\sum_{i=1}^n \delta_i}{\alpha} - \frac{1}{\alpha} \sum_{i=1}^n \left(\frac{t_i}{\alpha}\right)^\beta - \frac{\beta}{\alpha} \sum_{i=1}^n \left(\frac{t_i}{\alpha}\right)^\beta \log\left(\frac{t_i}{\alpha}\right) \\
-\frac{\partial^2 \ell(\alpha, \beta, \underline{t})}{\partial \beta^2} &= \frac{\sum_{i=1}^n \delta_i}{\beta^2} + \sum_{i=1}^n \left(\frac{t_i}{\alpha}\right)^\beta \log^2\left(\frac{t_i}{\alpha}\right)
\end{aligned}$$

Pruebas de hipótesis: Procedimientos asintóticos

Hacer inferencias exactas en supervivencia es la excepción más que la regla, por lo que juegan un papel importantísimo los procesos de inferencia asintóticos. Vamos a presentar tres pruebas estadísticas derivadas de la verosimilitud, para realizar pruebas asintóticas de hipótesis .

Caso: θ uniparamétrico

Queremos probar

$$H_0 : \theta = \theta_0 \quad vs. \quad H_a : \theta \neq \theta_0$$

Por lo general, es de interés probar $\theta_0 = 0$.

- *Cociente generalizado de verosimilitudes*

$$\Lambda = \frac{\sup_{\theta \in \Theta_0} L(\theta, \underline{t})}{\sup_{\theta \in \Theta} L(\theta, \underline{t})} = \frac{L(\theta_0, \underline{t})}{L(\hat{\theta}, \underline{t})} = \frac{L(0)}{L(\hat{\theta})}$$

Por teoría asintótica, se sabe que

$$-2\log\Lambda = -2\log\left(\frac{L(0)}{L(\hat{\theta})}\right) = 2\log L(\hat{\theta}) - 2\log L(0) \stackrel{a}{\sim} \chi_1^2$$

- *Prueba de Wald (“Disney”)*

Como $\hat{\theta}$ es máximo verosímil, tenemos

$$\begin{aligned} \hat{\theta} &\stackrel{a}{\sim} N(\theta, \hat{\text{var}}(\hat{\theta})), \quad \text{con} \quad \hat{\text{var}}(\hat{\theta}) = I(\hat{\theta})^{-1} \\ \Rightarrow \text{bajo } H_0 : \theta = 0, \quad \frac{\hat{\theta}}{\sqrt{I(\hat{\theta})^{-1}}} &\stackrel{a}{\sim} N(0, 1) \Rightarrow \frac{\hat{\theta}^2}{I(\hat{\theta})^{-1}} \stackrel{a}{\sim} \chi_1^2 \end{aligned}$$

- *Prueba de puntajes (Score)*

La función de puntajes es: $U = \frac{\partial}{\partial \theta} \log L(\theta, \underline{t}) = \sum_{i=1}^n \frac{\partial}{\partial \theta} \log f(t_i; \theta)$. Bajo $H_0 : \theta = 0$

$$\frac{U(\underline{0})}{\widehat{\text{var}}(U(\underline{0}))} = \frac{U(\underline{0})}{I(\underline{0})} \stackrel{a}{\sim} \chi_1^2$$

Note que la estadística de Wald es equivalente a la estadística

$$Z = \frac{\hat{\theta}}{\sqrt{\widehat{\text{var}}(\hat{\theta})}} \stackrel{a}{\sim} \sim N(0, 1)$$

Caso: $\underline{\theta}$ multi paramétrico

- Cociente de verosimilitudes

$$H_0 : \underline{\theta} = \underline{0} \quad \text{vs.} \quad H_a : \underline{\theta} \neq \underline{0}$$

$$\Lambda = \frac{\sup_{\underline{\theta} \in \Theta_0} L(\underline{\theta}, \underline{t})}{\sup_{\underline{\theta} \in \Theta} L(\underline{\theta}, \underline{t})} = \frac{L(\underline{0})}{L(\hat{\underline{\theta}})}$$

Entonces

$$-2 \log \Lambda = -2 \log \left(\frac{L(\underline{0})}{L(\hat{\underline{\theta}})} \right) = 2 \log L(\hat{\underline{\theta}}) - 2 \log L(\underline{0}) \stackrel{a}{\sim} \chi_{(\dim(\underline{\theta}))}^2$$

Wald

Si $\underline{\theta} = \underline{0}$, entonces

$$\hat{\underline{\theta}}' I^{-1}(\hat{\underline{\theta}}) \hat{\underline{\theta}} \stackrel{a}{\sim} \chi_{(\dim(\underline{\theta}))}^2$$

Puntajes

Si $\underline{\theta} = \underline{0}$, entonces

$$U'(\underline{0}) I^{-1}(\underline{0}) U(\underline{0}) \stackrel{a}{\sim} \chi_{(\dim(\underline{\theta}))}^2$$

Estas tres pruebas son asintóticamente equivalentes; sin embargo, por lo general, dan valores distintos. No es claro cuál de ellas es la “más mejor”. La más común es la de cociente de verosimilitudes.

Pruebas de hipótesis asintóticas (Teoría general)

Supongamos que deseamos realizar el contraste de hipótesis

$$\mathbb{H}_0 : \theta \in \Theta_0 \quad vs. \quad \mathbb{H}_a : \theta \in \Theta_1, \quad \Theta_0 \cap \Theta_1 = \emptyset, \quad \Theta_0 \cup \Theta_1 = \Theta$$

Supongamos que θ_0 es el supremo en el espacio que determina la hipótesis nula. Denotemos por $\ell(\theta)$ la log-verosimilitud y $\hat{\theta}_n$ el estimador consistente de θ , raíz de la ecuación de verosimilitud, $\ell'(\theta) = 0$. Intuitivamente, mientras más lejos esté θ_0 de $\hat{\theta}_n$ tendremos evidencia suficiente en contra de la hipótesis nula. Pero, ¿qué tanto es “suficiente”? Notemos que si θ_0 está cercano a $\hat{\theta}_n$, entonces $\ell(\theta_0)$ debe estarlo de $\ell(\hat{\theta}_n)$ y $\ell'(\theta_0)$ de $\ell'(\hat{\theta}_n)$

Existen tres pruebas asintóticas para realizar este contraste de hipótesis

- La prueba de **Wald** que se basa en el valor de $(\hat{\theta}_n - \theta_0)$
- La prueba de **el cociente de verosimilitudes**, basada en $\ell(\hat{\theta}_n) - \ell(\theta_0)$, y
- La prueba de **puntajes (score)** que se basa en $\ell'(\theta_0)$

Algunos resultados que serán de utilidad para realizar los desarrollos de estas pruebas, son los que obtuvimos en la demostración de las características asintóticas del estimador máximo verosímil (MLE). Éstos son

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N\left(0, \frac{1}{\mathbf{I}(\theta)}\right) \quad (1)$$

$$\frac{1}{\sqrt{n}}\ell'(\theta_0) \xrightarrow{d} N(0, \mathbf{I}(\theta)) \quad (2)$$

$$-\frac{1}{\sqrt{n}}\ell''(\theta_0) \xrightarrow{\mathbb{P}} \mathbf{I}(\theta) \quad (3)$$

Cociente de verosimilitudes

Dado que esta es, tal vez, la más popular de estas tres pruebas asintóticas, la desarrollaremos primero. La hipótesis a contrastar son las enunciadas anteriormente

$$\mathbb{H}_0 : \theta \in \Theta_0 \quad vs. \quad \mathbb{H}_a : \theta \in \Theta_1, \quad \Theta_0 \cap \Theta_1 = \emptyset, \quad \Theta_0 \cup \Theta_1 = \Theta$$

Como mencionamos, esta prueba se basa en $\ell(\hat{\theta}_n) - \ell(\theta_0)$. Los hechos que mencionamos en la parte intuitiva, esencialmente implican que suponemos suavidad y continuidad de la función de log-verosimilitud. Entonces, bajo estos supuestos, desarrollaremos $\ell(\theta_0)$ en series de Taylor alrededor de $\hat{\theta}_n$, el estimador máximo verosímil de θ . Entonces

$$\ell(\theta_0) = \ell(\hat{\theta}_n) + \ell'(\hat{\theta}_n)(\theta_0 - \hat{\theta}_n) + \frac{\ell''(\tilde{\theta}_n)}{2}(\theta_0 - \hat{\theta}_n)^2 + \dots \quad (\tilde{\theta}_n \in (\theta_0, \hat{\theta}_n))$$

$$-(\ell(\theta_0) - \ell(\hat{\theta}_n)) \approx -\frac{\ell''(\tilde{\theta}_n)}{2}(\theta_0 - \hat{\theta}_n)^2 \quad (\text{ya que } \ell'(\hat{\theta}_n) = 0)$$

$$-2\log(\Lambda_n) = 2(\ell(\hat{\theta}_n) - \ell(\theta_0)) \approx -\frac{\ell''(\tilde{\theta}_n)}{n} \left(\sqrt{n}(\hat{\theta}_n - \theta_0) \right)^2$$

Obsérvese que, como vimos en la demostración correspondiente al MLE, el hecho de que $\tilde{\theta}_n \in (\theta_0, \hat{\theta}_n)$, implica que $\tilde{\theta}_n \xrightarrow{\mathbb{P}} \theta_0$. Entonces, utilizando los resultados desprendidos de la demostración de las características asintóticas del MLE, tenemos que cuando $n \rightarrow \infty$

$$\begin{aligned} \sqrt{n}(\hat{\theta}_n - \theta_0) &\xrightarrow{d} N\left(0, \frac{1}{\mathbf{I}(\theta)}\right), y \\ -\frac{1}{\sqrt{n}}\ell''(\tilde{\theta}_n) &\xrightarrow{\mathbb{P}} \mathbf{I}(\theta) \end{aligned}$$

En este caso que asumimos que, bajo la hipótesis nula, el verdadero valor de θ es θ_0 , y utilizando el teorema de Slutsky tenemos que

$$-2\log(\Lambda_n) \rightarrow \mathbf{I}(\theta_0)N\left(0, \frac{1}{\mathbf{I}(\theta_0)}\right)^2 = \chi_{(1)}^2$$

que era lo que se quería probar.

Prueba de Wald

Ahora veamos una prueba asintótica alternativa, para realizar el mismo contraste. Es decir, nuevamente queremos hacer la prueba de hipótesis

$$\mathbb{H}_0 : \theta \in \Theta_0 \quad vs. \quad \mathbb{H}_a : \theta \in \Theta_1, \quad \Theta_0 \cap \Theta_1 = \emptyset, \quad \Theta_0 \cup \Theta_1 = \Theta$$

Obsérvese que de los tres resultados desprendidos del desarrollo de las características asintóticas del **MLE**, el primero de ellos (1) sugiere que, bajo el supuesto de que la hipótesis nula es cierta, i.e., suponiendo que el verdadero valor de θ es θ_0 , si definimos

$$\mathbf{W}_n = \sqrt{n\mathbf{I}(\theta_0)} (\hat{\theta}_n - \theta_0)$$

entonces, \mathbf{W}_n , llamada la **estadística de Wald**, debe converger en distribución a una normal estándar ($N(0,1)$), bajo \mathbb{H}_0 , ya que es simplemente la estandarización correspondiente. Obsérvese también que este resultado debe ser cierto si definimos

$$\begin{aligned} \mathbf{W}_n &= \sqrt{n\hat{\mathbf{I}}(\theta_0)} (\hat{\theta}_n - \theta_0) \\ &= \frac{\hat{\theta}_n - \theta_0}{\sqrt{\frac{\hat{\mathbf{I}}^{-1}(\theta_0)}{n}}} \stackrel{a}{\sim} N(0, 1) \\ \mathbf{W}_n^2 &= n\mathbf{I}(\theta_0) (\hat{\theta}_n - \theta_0)^2 \stackrel{a}{\sim} \chi_{(1)}^2 \end{aligned}$$

con $\hat{\mathbf{I}}(\theta_0)$ un estimador consistente de $\mathbf{I}(\theta_0)$. Por ejemplo, $\hat{\mathbf{I}}(\theta_0) = \mathbf{I}(\hat{\theta}_n)$. Entonces

Definición: \mathbf{W}_n es la **prueba de Wald** para rechazar $\mathbb{H}_0 : \theta \in \Theta_0$ en favor de $\mathbb{H}_a : \theta \in \Theta_1$.

Prueba de score (puntajes) o la estadística score de Rao

Esta prueba está basada en las características de la función score o de puntajes que utilizamos para construir la cota inferior de Cramér-Rao. Recordemos que la función de puntajes era

$$\mathbf{U}(\theta) = \ell'(\theta) = \frac{\partial \ell}{\partial \theta} = \sum_{i=1}^n \log f(x_i; \theta)$$

Recordemos además que

$$\mathbb{E}_\theta(\mathbf{U}) = 0, \text{ y}$$

$$\mathbb{V}_\theta(\mathbf{U}) = \mathbb{E}_\theta(\mathbf{U}^2) = n\mathbb{E}_\theta \left\{ \left(\frac{\partial}{\partial \theta} \log f(x_i; \theta) \right)^2 \right\}$$

Estos valores de las esperanzas, se calculan bajo el verdadero valor de θ . Entonces, retomando nuestro objetivo de realizar la prueba de hipótesis ya mencionada, si \mathbb{H}_0 es cierta, entonces $\theta = \theta_0$, y las esperanzas se calcularían ahora bajo el verdadero valor de θ que es θ_0 . Entonces, utilizando el resultado (2) de nuestra lista de resultados sobre las características asintóticas del **MLE**, tenemos que

$$\mathbf{R}_n = \frac{\mathbf{U}_n}{\sqrt{n\mathbf{I}(\theta_0)}} \stackrel{a}{\sim} N(0, 1)$$

Podemos remplazar $\mathbf{I}(\theta_0)$ por un estimador consistente $\hat{\mathbf{I}}(\theta_0)$, pero usualmente no se hace. Una de las ventajas de esta prueba es que no es necesario calcular $\hat{\theta}_n$, y si usamos $\mathbf{I}(\hat{\theta}_n)$ en lugar de $\mathbf{I}(\theta_0)$, perderíamos esta ventaja. Entonces tenemos la siguiente definición

Definición: \mathbf{R}_n es la **prueba score o score de Rao** para rechazar $\mathbb{H}_0 : \theta \in \Theta_0$ en favor de $\mathbb{H}_a : \theta \in \Theta_1$.

Generalización al caso multiparamétrico

La generalización de estas tres estadísticas para el caso multiparamétrico es inmediata. Los argumentos serán esencialmente los mismos, pero las aproximaciones de Taylor involucradas en los desarrollos, corresponderían a las respectivas versiones vectoriales de este desarrollo. Las hipótesis en este caso son

$$\mathbb{H}_0 : \underline{\theta} \in \Theta_0 \quad \text{vs.} \quad \mathbb{H}_a : \underline{\theta} \in \Theta_1, \quad \underline{\theta} \in \mathbb{R}^k$$

Supongamos que, $\underline{\theta}_0$ es el supremo en el espacio determinado por la hipótesis nula. Entonces,

$$\mathbf{W}_n = n \left(\hat{\underline{\theta}} - \underline{\theta}_0 \right)' \mathbf{I}(\underline{\theta}_0) \left(\hat{\underline{\theta}} - \underline{\theta}_0 \right) \stackrel{a}{\sim} \chi_{(k)}^2$$

$$\mathbf{R}_n = \frac{1}{n} \nabla \ell(\underline{\theta}_0)' \mathbf{I}^{-1}(\underline{\theta}_0) \nabla \ell(\underline{\theta}_0) \stackrel{a}{\sim} \chi_{(k)}^2$$

$$\Lambda_n = 2 \left(\ell(\hat{\underline{\theta}}) - \ell(\underline{\theta}_0) \right) \stackrel{a}{\sim} \chi_{(k)}^2$$

Modelos con covariables

Hasta ahora hemos supuesto $\{T_1, T_2, \dots, T_n\}$ v.s.a.s.i.i.d.. El supuesto de *idénticamente distribuidas* se refiere a encontrarse a individuos *idénticos* que, en poblaciones humanas, es prácticamente imposible. En animales (una camada) o máquinas es más fácil controlar la heterogeneidad.

Factores que pueden afectar los tiempos de falla (factores de heterogeneidad)

- Tratamientos
- Características propias de los individuos
- Medio ambiente (variables exógenas)

Tratamientos: Individuos que son alterados a voluntad del investigador

Características propias: Características intrínsecas de los sujetos: sexo, edad, peso, etc. no se pueden controlar de la misma forma como se hace con un tratamiento. Se puede controlar la distribución de estas características en la muestra.

Medio ambiente: Pueden hacer que los individuos no sean comparables. No se nace con esta característica: lugar de residencia, clínica de atención médica. El tratamiento puede cambiar a lo largo del tiempo si se detecta otro mejor. O puede aumentarse o disminuirse la dosis, de acuerdo a la evolución del individuo. Algunas características del individuo pueden ir cambiando, o modificarse como resultado del tratamiento: Dieta puede afectar, peso, presión arterial, etc.

¿Cómo incorporar estos factores al análisis del tiempo de falla?

- Un vector de covariables \mathbf{Z}_i para cada individuo (igualito que en regresión lineal) que van a afectar el tiempo de falla
- $\mathbf{Z}_{p \times 1}$ vector de covariables (tratamiento, características del individuo, variables exógenas)
- \mathbf{Z} constante o puede depender del tiempo \mathbf{Z}_t

Si tuviéramos dos individuos con el mismo vector de covariables, los consideraríamos homogéneos y ahí se valdría todo lo que hemos hecho. ¿Cómo modelar?

1.- Definir \underline{Z} tal que $\underline{Z} = \underline{0}$ sea una población de referencia o básica

2.- Describir la forma como cambia T al pasar de un individuo con covariables $\underline{Z} = \underline{0}$ a otro con covariables $\underline{Z} \neq \underline{0}$

Las formas de cambio propuestas dan lugar a dos grandes familias de modelos, las más usuales pero no las únicas.

Familia de riesgos proporcionales

Se basa en el cambio de la función de riesgo de un individuo con covariables \underline{Z} con respecto a la función de riesgo de la población básica $\underline{Z} = \underline{0}$

Modelo de vida acelerada

Se basa en modelar el cambio de la función de supervivencia, entre un individuo con covariables \underline{Z} y un individuo en la población básica $\underline{Z} = \underline{0}$

Procedimiento de análisis

1.- Modelo para la población básica (cualquiera de los ya discutidos)

2.- Determinar el cambio en el modelo anterior cuando $\underline{Z} \neq \underline{0}$. Naturalmente, este cambio se especifica de forma paramétrica.

Modelo de riesgos proporcionales (D.R. Cox (1972))

- Población básica $\underline{\mathbf{Z}} = \underline{\mathbf{0}}$, con función de riesgo $h_0(t)$
- Supondremos que existe $\rho > 0$, tal que

$$h(t|\underline{\mathbf{Z}}) = \rho h_0(t)$$

$\rho > 1$ la población con covariables $\underline{\mathbf{Z}}$, tiene mayor riesgo de presentar la falla

$0 < \rho < 1$ la población básica tiene mayor riesgo de presentar esta falla

$$\frac{h(t|\underline{\mathbf{Z}})}{h_0(t)} = \rho : \text{RIESGOS PROPORCIONALES}$$

ρ no depende de t . El cociente es el mismo $\forall t$

Modelo general

$$h(t|\underline{\mathbf{Z}}) = \rho(\underline{\mathbf{Z}})h_0(t)$$

Características de $\rho(\underline{\mathbf{Z}})$

- $\rho(\underline{\mathbf{Z}}) > 0$
- $\rho(\underline{\mathbf{0}}) = 1$

forma paramétrica de $\rho(\underline{\mathbf{Z}})$

$$\rho(\underline{\mathbf{Z}}) = \exp\{\underline{\beta}'\underline{\mathbf{Z}}\}$$

$$\underline{\beta}' = (\beta_1, \beta_2, \dots, \beta_p)$$

- No depende de t
- $\rho(\underline{\mathbf{Z}}) > 0$
- $\rho(\underline{\mathbf{0}}) = 1$

Entonces, el modelo de riesgos proporcionales es

$$h(t|\underline{\mathbf{Z}}) = h_0(t) \exp \left\{ \underline{\beta}' \underline{\mathbf{Z}} \right\}$$

y

$$\frac{h(t|\underline{\mathbf{Z}})}{h_0(t)} = \exp \left\{ \underline{\beta}' \underline{\mathbf{Z}} \right\} \text{ constante } \forall t$$

Para comparar poblaciones con covariables $\underline{\mathbf{Z}}_1$ y $\underline{\mathbf{Z}}_2$

$$\frac{h(t|\underline{\mathbf{Z}}_1)}{h(t|\underline{\mathbf{Z}}_2)} = \exp \left\{ \underline{\beta}' (\underline{\mathbf{Z}}_1 - \underline{\mathbf{Z}}_2) \right\}$$

!!! Independiente de $h_0(t)$!!! e independiente de t .

Observaciones

1.-

$$\log \left(\frac{h(t|\underline{\mathbf{Z}})}{h_0(t)} \right) = \underline{\beta}' \underline{\mathbf{Z}} = \beta_1 z_1 + \beta_2 z_2 + \dots + \beta_p z_p \text{ Modelo lineal}$$

De hecho, la relación lineal se logra con el logaritmo del cociente, así que propiamente el modelo es log-lineal.

2.-

$$\log \left(\frac{h(t|\underline{\mathbf{Z}})}{h_0(t)} \right) = \log (h(t|\underline{\mathbf{Z}})) - \log (h_0(t)) = \underline{\beta}' \underline{\mathbf{Z}} \text{ constante}$$

entonces, $\log (h(t|\underline{\mathbf{Z}}))$ y $\log (h_0(t))$ son *paralelas* $\forall t$.

3.-

$$S(t|\underline{\mathbf{Z}}) = \exp \left\{ - \int_0^t \rho(\underline{\mathbf{Z}}) h_0(u) du \right\} = \exp \left\{ \left[- \int_0^t h_0(u) du \right] \rho(\underline{\mathbf{Z}}) \right\} = (S_0(t))^{\rho(\underline{\mathbf{Z}})} = (S_0(t))^{\exp \left\{ \underline{\beta}' \underline{\mathbf{Z}} \right\}}$$

Interpretación de los parámetros

Como ya vimos, el modelo de riesgos proporcionales lo podemos escribir como un modelo log-lineal

$$\log \left(\frac{h(t|\mathbf{Z})}{h_0(t)} \right) = \underline{\beta}' \mathbf{Z} = \beta_1 z_1 + \beta_2 z_2 + \dots + \beta_p z_p$$

Consideremos el modelo con un solo regresor, Z . Si el regresor es continuo y tomamos z y $z + 1$ dos valores consecutivos de esta covariable, tenemos:

$$Z = z + 1 \Rightarrow \log \left(\frac{h(t|z+1)}{h_0(t)} \right) = \beta_1 * (z + 1) = \beta_1 * z + \beta_1, \text{ y para}$$

$$Z = z \Rightarrow \log \left(\frac{h(t|z)}{h_0(t)} \right) = \beta_1 * z, \text{ entonces}$$

$$\log \left(\frac{h(t|z+1)}{h_0(t)} \right) - \log \left(\frac{h(t|z)}{h_0(t)} \right) = \log \left(\frac{h(t|z+1)}{h(t|z)} \right) = \beta_1 * z + \beta_1 - \beta_1 * z = \beta_1$$

entonces β_1 se interpreta como el cambio en el logaritmo del cociente de riesgos por unidad de cambio en esta covariable. Si tomamos la exponencial, entonces $\exp(\beta_1)$ se interpreta como el cambio en el cociente de riesgos por un incremento de una unidad en esta covariable, manteniendo constante el resto de las variables. Cuando estos parámetros se estiman, se interpretan igual agregando la frase: “cambio promedio o cambio esperado, manteniendo constante...” por unidad de cambio de la covariable. Si el incremento entre los dos valores de la covariable es r , entonces el cambio entre los riesgos es $\exp(r\beta_1)$ y se interpreta de manera semejante.

Si la variable explicativa es dicotómica, $\exp(\beta_1)$ es el cambio en el cociente de riesgos entre un sujeto con valor de la covariable “1” y otro con valor “0”. Es importante notar que el cambio en el cociente de riesgos que se registra por unidad de cambio en una variable continua, es el mismo, independientemente del valor de la covariable, situación no plausible en muchos estudios. Esta dificultad puede salvarse categorizando la variable continua. En este caso, habrá tantos parámetros como categorías menos uno tenga la variable, y $\exp(\beta_j)$ será el cambio en el cociente de riesgos entre un individuo en la categoría j de esta variable, contra uno en el nivel basal, generalmente “0”. Cuando el análisis se realiza con un

paquete estadístico, debe verificarse cuál es la categoría de referencia que utiliza este paquete.

Estimación y pruebas de hipótesis en el modelo de riesgos proporcionales (PH)

Construcción de la verosimilitud

La información de la que disponemos para construir la verosimilitud del **PH** es: $(T_i, \delta_i, \underline{Z}_i)$, $i=1,2,\dots,n$. Las covariables en \underline{Z}_i , pueden ser continuas o discretas, y, por el momento, no dependientes del tiempo. Supongamos que

- Hay k distintos tiempos de falla t_1, t_2, \dots, t_k . Por ahora, supongamos que *no hay empates* en estos tiempos.
- R_i : el i -ésimo conjunto en riesgo. El conjunto de los n_i individuos vivos y no censurados al tiempo t_i

R_0 : Todos los individuos (n)

R_1 : Individuos vivos y no censurados antes de la primera falla (t_1)

⋮

R_k : Individuos vivos y no censurados antes de la k -ésima falla (t_k)

¿Cómo construir la verosimilitud en este caso?

Consideremos la probabilidad de que un individuo muera en algún tiempo t_j . Este individuo tiene covariables \underline{Z}_j

$\mathbb{P}[\text{un individuo con covariables } \underline{Z}_j \text{ muera en } t_j \mid \text{Estuvo en riesgo, i.e., que pertenece}$

$$\text{al conjunto en riesgo } R_j] = \frac{\mathbb{P}[\text{un individuo con covariables } \underline{Z}_j \text{ muera en } t_j]}{\mathbb{P}[\text{Conjunto en riesgo } R_j]}$$

$$= \frac{h(t_j | \underline{Z}_j)}{\sum \text{Riesgo de todos los individuos del conjunto } R_j} = \frac{h(t_j | \underline{Z}_j)}{\sum_{\ell \in R_j} h(t_\ell | \underline{Z}_\ell)}$$

bajo el supuesto de riesgos proporcionales,

$$\frac{h(t_j|\underline{\mathbf{Z}}_j)}{\sum_{\ell \in R_j} h(t_\ell|\underline{\mathbf{Z}}_\ell)} = \frac{h_0(t_j) \exp\{\underline{\beta}' \underline{\mathbf{Z}}_j\}}{\sum_{\ell \in R_j} h_0(t_\ell) \exp\{\underline{\beta}' \underline{\mathbf{Z}}_\ell\}} = \frac{\exp\{\underline{\beta}' \underline{\mathbf{Z}}_j\}}{\sum_{\ell \in R_j} \exp\{\underline{\beta}' \underline{\mathbf{Z}}_\ell\}}$$

¿y la verosimilitud? Hay que tomar el producto sobre todos los tiempos de falla

$$L(\underline{\beta}, \underline{t}, \underline{\delta}, \mathbf{Z}) = \prod_{j=1}^k \frac{\exp\{\underline{\beta}' \underline{\mathbf{Z}}_j\}}{\sum_{\ell \in R_j} \exp\{\underline{\beta}' \underline{\mathbf{Z}}_\ell\}}$$

Observaciones

Esta “cosa” no es una verosimilitud en el sentido usual, porque

- El número de términos no es el número de individuos en la muestra, sino el de fallas.
- Si observamos un individuo más, podemos provocar que cambien los conjuntos en riesgo (la nueva verosimilitud, no sería proporcional a la anterior)
- Si quitamos un individuo, puede ocurrir lo mismo que en el punto anterior.
- Obsérvese que los individuos censurados no contribuyen al numerador de esta verosimilitud, pero sí al denominador, ya que pertenecen a los conjuntos en riesgo, R_j
- La verosimilitud depende únicamente del rango de los tiempos de falla, ya que éstos determinan el conjunto en riesgo de cada tiempo de falla, lo que hace este procedimiento *semi paramétrico*, ya que se deben estimar los parámetros asociados a las covariables, pero no se asume distribución sobre los tiempos de supervivencia.

Finalmente, podemos reescribir esta verosimilitud como

$$L(\underline{\beta}, \underline{t}, \underline{\delta}, \mathbf{Z}) = \prod_{j=1}^n \left[\frac{\exp\{\underline{\beta}' \underline{\mathbf{Z}}_j\}}{\sum_{\ell \in R_j} \exp\{\underline{\beta}' \underline{\mathbf{Z}}_\ell\}} \right]^{\delta_j}$$

y su log-verosimilitud

$$L(\underline{\beta}, \underline{t}, \underline{\delta}, \mathbf{Z}) = \sum_{j=1}^n \delta_j \left[\underline{\beta}' \underline{\mathbf{Z}}_j - \log \left(\sum_{\ell \in R_j} \exp\{\underline{\beta}' \underline{\mathbf{Z}}_\ell\} \right) \right]$$

Dada la log-verosimilitud, la estimación de los parámetros se realiza por los procesos usuales. La derivada parcial de la verosimilitud con respecto a β_k , $k = 1, 2, \dots, p$ es

$$U_k(\underline{\beta}) = \frac{\partial \log L(\underline{\beta})}{\partial \beta_k} = \sum_{j=1}^n \delta_j \left[z_{kj} - \frac{\sum_{\ell \in R_j} z_{k\ell} \exp \{ \underline{\beta}' \mathbf{Z}_\ell \}}{\sum_{\ell \in R_j} \exp \{ \underline{\beta}' \mathbf{Z}_\ell \}} \right]$$

que podemos ver, de manera intuitiva, como una suma de valores “observados” menos valores “esperados”

$$U_k(\underline{\beta}) = \frac{\partial \log L(\underline{\beta})}{\partial \beta_k} = \sum_{j=1}^n \delta_j (z_{kj} - \bar{z}_{kj})$$

donde \bar{z}_{kj} es el “promedio ponderado” de la covariable z_k , sobre todos los individuos en riesgo al tiempo t_j .

Los estimadores de β_k , $k = 1, 2, \dots, p$ se obtienen igualando a cero estas ecuaciones de verosimilitud, mismas que se resuelven por métodos numéricos.

Las pruebas de hipótesis se realizan mediante los procedimientos asintóticos ya descritos. Estas pruebas se hacen para verificar la significancia global del modelo, y para probar la significancia de cada covariable dentro de este modelo. Esta última es muy importante, ya que sirve para determinar el impacto que tiene una característica particular de los sujetos, en el riesgo de presentar la falla. En símbolos se escriben

$$H_0 : \underline{\beta} = \mathbf{0} \Rightarrow \beta_1 = \beta_2 = \dots = \beta_p = 0 \quad vs \quad H_a : \underline{\beta} \neq \mathbf{0} \Rightarrow \beta_j \neq 0 \quad p.a. \quad j = 1, 2, \dots, p \quad (\text{Global})$$

$$H_0 : \beta_j = 0 \quad vs. \quad H_a : \beta_j \neq 0 \quad j = 1, 2, \dots, p$$

Estimación no paramétrica de la función de riesgo básica

Como observamos, la verosimilitud del modelo **PH** cancela la función de riesgo básica. No obstante, es importante contar con una estimación no paramétrica de esta función, porque es necesaria en algunos procedimientos posteriores.

Recordemos que la verosimilitud para datos de supervivencia, se escribe de manera general como

$$L(\theta, \underline{t}) = \prod_{i=1}^n f(t_i)^{\delta_i} S(t_i)^{1-\delta_i} = \prod_{i=1}^n [h(t_i)S(t_i)]^{\delta_i} S(t_i)^{1-\delta_i} = \prod_{i=1}^n h(t_i)^{\delta_i} S(t_i) = \prod_{i=1}^n h(t_i)^{\delta_i} \exp\{-H(t_i)\}$$

Bajo el modelo **PH**, esta verosimilitud queda como

$$L(\underline{\beta}, h_0(t), \underline{t}, \underline{\delta}, \mathbf{Z}) = \prod_{j=1}^n h_0(t_j)^{\delta_j} \exp\{\underline{\beta}' \mathbf{Z}_j\}^{\delta_j} \exp\left(-H_0(t_j) \exp\{\underline{\beta}' \mathbf{Z}_j\}\right)$$

Analicemos los componentes de esta verosimilitud.

$\exp\left(-H_0(t_j) \exp\{\underline{\beta}' \mathbf{Z}_j\}\right)$ es máxima cuando $h_0(t) = 0$, pero $h_0(t_j)^{\delta_j}$ debería de contribuir de manera importante a la verosimilitud. Entonces, debe suceder que

$\hat{h}_0(t) = 0, \forall t \notin \{t_1, t_2, \dots, t_k\}$, y hay que encontrar los valores $\hat{h}_0(t_j)$, que maximicen esta verosimilitud.

Ahora, la verosimilitud depende de los valores de los parámetros en $\underline{\beta}$. La forma de lidiar con este problema es reemplazarlos por sus valores estimados $\hat{\underline{\beta}}$, y trabajar con la llamada *pseudo verosimilitud*, para estimar $h_0(t_i)$. Hemos supuesto que sólo tenemos k falla, entonces, la log-verosimilitud queda como

$$\begin{aligned}
\log L(h_0(t_j)) &= \sum_{j=1}^k \left[\log(h_0(t_j)) - \hat{\beta}' \mathbf{Z}_j - H_0(t_j) \exp \left\{ \hat{\beta}' \mathbf{Z}_j \right\} \right] \\
&= \sum_{j=1}^k \left[\log(h_0(t_j)) - \hat{\beta}' \mathbf{Z}_j - \sum_{t_i \leq t_j} h_0(t_i) \exp \left\{ \hat{\beta}' \mathbf{Z}_j \right\} \right] \\
&= \sum_{j=1}^k \left[\log(h_0(t_j)) - \hat{\beta}' \mathbf{Z}_j - h_0(t_j) \sum_{\ell \in R_j} \exp \left\{ \hat{\beta}' \mathbf{Z}_\ell \right\} \right]
\end{aligned}$$

Finalmente, derivando con respecto a $h_0(t_j)$, se tiene que el estimador no paramétrico de la función de riesgo básica está dado por

$$\hat{h}_0(t_j) = \frac{1}{\sum_{\ell \in R_j} \exp \left\{ \hat{\beta}' \mathbf{Z}_\ell \right\}}, \quad j = 1, 2, \dots, k$$

Este estimador fue desarrollado por Breslow (1972). El estimador no paramétrico de la función de supervivencia básica es

$$\hat{S}_0(t) = \prod_{j|t_j < t} \left[1 - \frac{1}{\sum_{\ell \in R_j} \exp \left\{ \hat{\beta}' \mathbf{Z}_\ell \right\}} \right]$$

Bondad de ajuste del PH

Una vez que se ha estimado el modelo, es indispensable juzgar qué tan bien se ajusta a nuestros datos. Además del ajuste, hay que recordar que el modelo se construye bajo el supuesto de *riesgos proporcionales*, mismo que debe verificarse para que nuestras inferencias sean confiables. Como en el modelo de regresión lineal, para verificar estos supuestos un elemento fundamental son los residuos del modelo, sólo que, para este modelo, contamos con varios tipos de residuos. Los residuos en supervivencia son ligeramente diferentes a los de otro tipo de modelos, debido a la presencia de la censura.

Residuos Cox-Snell

Sea T_i el tiempo de falla, con función de supervivencia asociada $S_i(t)$. Entonces, la variable aleatoria $S_i(T_i)$, se distribuye como una uniforme $[0, 1]$. Demostración, basta probar que $U = F_i(T_i)$ se distribuye uniforme $[0, 1]$. Supondremos que T_i es continua

$$\begin{aligned} Pr[U \leq u] &= Pr[F_i(T_i) \leq u] = Pr[T_i \leq F_i^{-1}(u)] = F_i(F_i^{-1}(u)) = u \\ &\Rightarrow S_i(T_i) = 1 - U \Rightarrow S_i(T_i) \sim U[0, 1] \end{aligned}$$

Los resultados anteriores implican que

$$H_i(t_i) = -\log(S_i(T_i)) \sim \exp(1)$$

Entonces, si el modelo es correcto, la función de riesgo acumulada estimada de cada individuo, en cada tiempo de falla o cada tiempo de censura, debe ser una *muestra censurada de una distribución exponencial con parámetro uno*. Esta cantidad es conocida como el *residuo Cox-Snell*

Si ajustamos el modelo **PH** a los datos, para verificar que ajusta bien, tendremos

$$\hat{S}(t|\mathbf{Z}_j) = \left[\hat{S}_0(t) \right]^{\exp\{\hat{\beta}'\mathbf{Z}_j\}}$$

y los residuos Cox-Snell son

$$r_{C_i} = \hat{H}(t_j) = -\log \left[\hat{S}(t|\underline{\mathbf{Z}}_j) \right]$$

Procedimiento gráfico para verificar el ajuste del modelo PH

Sabemos que la funciones de supervivencia y de riesgo acumulado de una exponencial son $S(t) = e^{-\lambda t}$ y $H(t) = \lambda t$, respectivamente. Entonces, una manera gráfica de verificar que el ajuste del modelo **PH** es adecuado, es que

- Al graficar $-\log[\hat{S}(t|\underline{\mathbf{Z}}_j)]$ vs. t observemos una línea recta, que pasa por el origen y tiene pendiente menos uno
- Al graficar $\log(-\log[\hat{S}(t|\underline{\mathbf{Z}}_j)])$ vs. $\log(t)$ observemos una línea recta que pasa por el origen y tiene pendiente uno

Residuos modificados de Cox-Snell

Los *residuos modificados de Cox-Snell*, intentan corregir los residuos originales de Cox-Snell para las observaciones censuradas. La propuesta es

$$r'_{C_i} = \begin{cases} r_{C_i}, & \text{para las fallas} \\ r_{C_i} + 1, & \text{para las censuras} \end{cases}$$

Argumentos de la propuesta

- El verdadero residuo (no observado) debe ser mayor que el residuo observado (censurado)
- Este “exceso de residuo” tiene también una distribución exponencial (1)
- $\mathbb{E}(\exp(1)) = 1$. Entonces, al valor del residuo r_{C_i} , se le suma el valor esperado del “exceso de residuo”.

Utilizando la variable indicadora δ_i , podemos escribir estos residuos como

$$r'_{C_i} = 1 - \delta_i + r_{C_i}$$

Otra versión de estos residuos modificados

$$r''_{C_i} = \begin{cases} r_{C_i}, & \text{para las fallas} \\ r_{C_i} + 0.693, & \text{para las censuras} \end{cases}$$

Argumentos

Evidencia empírica de que sumar “1” infla demasiado el residuo, en su lugar se “debe usar” la mediana (en lugar de la media) del exceso de residuo.

$$S(t) = e^{-t} \Rightarrow S(t_{50}) = e^{-t_{50}} = 0.5 \Rightarrow t_{50} = \log(2) = 0.693$$

Justificación formal del “aumento de uno” a los residuos Cox-Snell censurados.

- Propiedad de pérdida de memoria de la exponencial. Para $T \sim \exp(\lambda)$ y $0 < t_0 < t_1$

$$\mathbb{P}[T > t_1 | T > t_0] = \frac{\mathbb{P}[T > t_1 \wedge T > t_0]}{\Pr[T > t_0]} = \frac{\mathbb{P}[T > t_1]}{\Pr[T > t_0]} = \frac{e^{-\lambda t_1}}{e^{-\lambda t_0}} = e^{-\lambda(t_1 - t_0)} \sim \exp(\lambda)$$

’Ora sí. Sea \mathbf{c} el tiempo de censura observado en el i -ésimo individuo. Queremos calcular: $\mathbb{E}(T | T \geq \mathbf{c})$, con $T \sim \exp(1)$. Primeramente $\mathbb{P}[T | T \geq c] = e^{-(T-c)}$

$$\Rightarrow \mathbb{E}(T | T \geq c) = \int_c^\infty t e^{-(t-c)} dt = e^c \int_c^\infty t e^{-t} dt = e^c [-e^{-t}(t+1)]_c^\infty = e^c [e^{-c}(c+1)] = c+1$$

Residuos de martingalas

Se define el *residuo de martingala* para el i -ésimo individuo como

$$r_i = \delta_i - \hat{\Lambda}(t_i), \quad i = 1, 2, \dots, n$$

Interpretación: Los residuos de martingalas pueden verse como la diferencia entre el número observado de muertes (0 ó 1) por cada sujeto i , entre el tiempo 0 y t_i , y el número esperado basado en el modelo.

Propiedades

- Los r'_i s tienen media cero

- Son aproximadamente *no correlacionados* (para muestras grandes)

Los residuos de martingalas sirven para verificar que las covariables ingresan de forma lineal al modelo. Si una covariable realmente tiene una relación lineal con el riesgo, los puntos en la gráfica: residuos de martingalas vs. valores de la covariable, deben verse distribuidos completamente al azar (*ruido blanco*). Este proceso debe realizarse para cada variable en el modelo. Si la variable es dicotómica o politómica, estas gráficas resultan poco útiles.

Residuos de devianza

Una desventaja que tienen los residuos de martingalas, es que tienden a ser asimétricos con sesgo hacia la derecha, una solución a este problema es usar los *residuos de devianza*. Para cada observación i , los residuos de devianza se definen como una transformación de los de martingalas, dada por

$$\hat{D}_i = \text{signo}(\hat{r}_i) \sqrt{-2[\hat{r}_i + \delta_i \log(\delta_i - \hat{r}_i)]}$$

Estos residuos tienen características similares a los residuos de regresión lineal (media: 0 y varianza: 1). Son negativos para observaciones con tiempo de supervivencia menores que el esperado. Se pueden utilizar para graficarlos contra los valores de cada covariable (como en los de martingalas), o para graficarlos vs. el logaritmo de los valores ajustados, y verificar que las observaciones tienen varianza constante, media cero y detectar observaciones con residuos grandes (mayores que 2 o 3 en valor absoluto).

Residuos de Schoenfeld

Se definen para cada tiempo de falla observado, como

$$r_{ij}^s = Z_{ij}(t_i) - \bar{Z}_j(t_i), \quad j = 1, \dots, p \quad i = 1, 2, \dots, k$$

Características

- Representa la diferencia entre la covariable observada y el promedio sobre el conjunto en riesgo en cada tiempo de falla
- Se calculan para cada covariable

- No están definidos para tiempos de supervivencia censurados
- Útiles para verificar tendencia en el tiempo o falta de proporcionalidad, basada en gráficas contra el tiempo. Además, sirven como base para pruebas formales sobre si, tanto el modelo como cada una de las covariables, cumplen el supuesto de riesgos proporcionales
- Suman cero; tienen valor esperado cero, y son no correlacionados (en muestras grandes)

Residuos escalados de Schoenfeld

Estos residuos tienen más uso que los de Schoenfeld, principalmente porque son más parecidos a los de mínimos cuadrados ordinarios. Se definen como

$$r_{ij}^* = k \text{Var}(\underline{\hat{\beta}}) r_{ij}^s$$

con k el número de fallas y $\text{Var}(\underline{\hat{\beta}})$ la matriz $p \times p$ de varianza-covarianza del vector de parámetros estimados al ajustar el modelo **PH**. Grambsch y Therneau (1994) demostraron que si $\hat{\beta}_j$ es el coeficiente estimado al ajustar el modelo **PH**, sin covariables dependientes del tiempo, entonces

$$\mathbb{E}(r_{ij}^*) \approx \beta_j(t_i) - \hat{\beta}_j \quad \text{o bien} \quad E(r_{ij}^*) + \hat{\beta}_j \approx \beta_j(t_i)$$

Este hecho sugiere que graficar $r_{ij}^* + \hat{\beta}_j$ vs. el tiempo o alguna función $g(t)$ del tiempo, puede ser adecuada para visualizar la naturaleza y “tamaño” de la no proporcionalidad de los riesgos. Puede ajustarse una línea a la gráfica, que debe verse aproximadamente horizontal; además de una prueba para verificar que la pendiente es *cero*. Una pendiente *distinta de cero*, proporciona evidencia contra riesgos proporcionales.

Estructura de la prueba

$$\begin{aligned} \beta_j(t) &= \beta_j + \theta_j(g_j(t) - \bar{g}_j), \quad j = 1, 2, \dots, p \\ \mathbb{H}_0 : \theta_j &= 0 \quad \text{vs.} \quad \mathbb{H}_a : \theta_j \neq 0 \end{aligned}$$

Diagnóstico de las observaciones en el modelo PH

Las pruebas anteriores proporcionan indicios del ajuste global del modelo a los datos, lo que significa que *globalmente* el modelo ajusta; sin embargo, es conveniente analizar el impacto que tiene sobre las características del modelo, *cada una de las observaciones que lo componen*, i.e., hay que realizar un *diagnóstico* de ellas. Para hacer esta tarea, utilizaremos los residuos score o de puntaje.

Residuos Score o de puntajes

Este residuo se deriva de las funciones score o de puntajes para cada estimador del coeficiente asociado a una covariable, en esencia, para cada covariable

$$U_k(\underline{\beta}) = \frac{\partial \log L(\underline{\beta})}{\partial \beta_k} = \sum_{j=1}^n \delta_j (z_{kj} - \bar{z}_{kj})$$

donde \bar{z}_{kj} es el “promedio ponderado” de la covariable z_k , sobre todos los individuos en riesgo al tiempo t_j . Nótese que el término dentro de la sumatoria son los residuos de Schoenfeld. Si suponemos que se observaron k tiempos de falla, estos residuos determinan una matriz, Δ , de $p \times k$.

Cómo y pa’ qué se usan estos residuos?

Influencia de las observaciones en

- En todo el vector de parámetros: $\hat{\underline{\beta}}$ (Similar a la distancia de Cook de regresión lineal)
- En cada parámetro $\hat{\beta}_j$ (Como los *dfbetas* de regresión lineal), $j=1,2,\dots,p$

La manera usual de detectar la influencia de la i -ésima observación es calculando el valor

$$d_i = \hat{\underline{\beta}} - \hat{\underline{\beta}}_{(i)} \quad \text{o} \quad d_{ij} = \hat{\beta}_j - \hat{\beta}_{j(i)}, \quad i = 1, \dots, n \quad j = 1, \dots, p$$

que es la diferencia entre los estimadores con y sin la i -ésima observación. Realizar “a patín” estos procesos para cada observación, esta “ca...”. La forma eficiente de hacerlo es la siguiente

Dada Δ nuestra matriz, $p \times k$, de residuos score, consideremos el producto

$$\Delta' Var(\hat{\underline{\beta}})$$

Con $Var(\hat{\underline{\beta}})$ como se definió antes. Se puede demostrar que

$$\Delta_i \hat{\beta}_j \approx \hat{\beta}_j - \hat{\beta}_{j(i)}, \quad \text{con } \Delta_i = (r_{i1}^*, r_{i2}^*, \dots, r_{ip}^*)'$$

Una medida de influencia global de la observación i , se calcula a través de

$$B = \Delta' Var(\hat{\underline{\beta}}) \Delta$$

Un argumento de álgebra lineal muestra que los valores absolutos de los elementos del vector asociado al eigenvalor más grande de la matriz \mathbf{B} , estandarizados para que este vector tenga *norma uno*, es una medida de la bondad de ajuste del modelo para cada una de las observaciones. Ya que el vector es de norma uno, la suma del cuadrado de sus elementos debe ser uno, entonces, una observación será influyente en $\hat{\underline{\beta}}$, si su contribución a esta suma es “considerable”.

En ninguno de los dos casos presentados, se tiene manera de saber si el valor de esta estadística es “grande”, para tener evidencia de que la observación en cuestión es influyente en el vector de parámetros o en algún parámetro, según sea el caso que se esté tratando, por lo que es recomendable observar el impacto en la característica en cuestión, al ajustar el modelo con y sin la observación.

Extensiones al modelo PH: covariables dependientes del tiempo

Hasta ahora, hemos considerado que las covariables en el modelo **PH** no dependen del tiempo. Asumimos que se miden al inicio del estudio y sus valores permanecen fijos a lo largo del estudio. Como es común en estudios de supervivencia y, en general, en estudios longitudinales, los individuos son monitoreados a lo largo del mismo, y otras covariables, cuyos valores pueden cambiar en el transcurso del estudio, se recolectan. Estas variables que cambian a través del tiempo se denominan *variables dependientes del tiempo*. Una variable que toma un valor de “0” hasta que un determinado evento sucede (tener novio (a), cambio de tratamiento, trasplante de un órgano), y después toma el valor “1”, es un ejemplo de una variable discreta dependiente del tiempo. Por supuesto, se pueden considerar covariables esencialmente continuas, de las que se registren una serie de valores a lo largo del estudio. Algunos ejemplos son: presión arterial, niveles de colesterol, índice de masa corporal, tamaño de un tumor, dosis de un medicamento, etc.

Tipos de variables dependientes del tiempo

Es común considerar dos tipos de variables que cambian a través del tiempo, que se denominan: *variables internas* y *variables externas*. Las variables internas denotan una característica particular del individuo, y pueden medirse únicamente mientras está vivo: presión arterial, niveles de colesterol, índice de masa corporal, tamaño de un tumor, trasplante de un órgano, etc. En cada caso, estas variables reflejan la condición del sujeto y sus valores pueden asociarse con la supervivencia del mismo.

Las variables externas, son variables dependientes del tiempo que no necesariamente requieren que el sujeto esté vivo para existir, tales como la edad, la dosis de una droga de la que se ha determinado, de manera previa, cómo debe variar a lo largo del estudio, y, variables cuya existencia es totalmente independiente de sujeto, como condiciones climáticas de presión atmosférica, niveles de algún contaminante, temperatura, etc.

En este caso, el modelo ya no cumple el supuesto de riesgos proporcionales para esas covariables dependientes del tiempo. El modelo de Cox para esta situación se escribe como

$$h(t|\mathbf{Z}) = h_0(t) \exp \left\{ \underline{\beta}'_1 \mathbf{Z}_1 + \underline{\beta}'_2 \mathbf{Z}_2(t) \right\}$$

con \mathbf{Z}_1 las matriz de variables que no dependen del tiempo y \mathbf{Z}_2 las que sí dependen del

tiempo. Los procesos de estimación son similares a los del modelo de riesgos proporcionales. La introducción de covariables dependientes del tiempo, no genera problemas técnicos en el ajuste del modelo, los problemas se presentan en la manera como hay que adecuar la base de datos para que un proceso automático estime el modelo con estas covariables.

La verosimilitud del modelo extendido de Cox

Recordemos que la verosimilitud de este modelo, en el caso de que no haya variables dependientes del tiempo es

$$L(\underline{\beta}, \underline{t}, \underline{\delta}, \mathbf{Z}) = \prod_{j=1}^n \left[\frac{\exp \{ \underline{\beta}' \underline{\mathbf{Z}}_j \}}{\sum_{\ell \in R_j} \exp \{ \underline{\beta}' \underline{\mathbf{Z}}_\ell \}} \right]^{\delta_j}$$

Esta verosimilitud compara el riesgo de un sujeto con covariables $\underline{\mathbf{Z}}_j$ contra los sujetos con covariables $\underline{\mathbf{Z}}_\ell$ en el conjunto en riesgo R_j al tiempo $t_{(j)}$. No obstante, no existe ninguna razón para suponer que las covariables $\underline{\mathbf{Z}}_i$ del i -ésimo sujeto, sean las mismas para todos los tiempos $t_{(j)}$. Por lo que, la verosimilitud del modelo **PH** con covariables dependientes del tiempo es

$$L(\underline{\beta}, \underline{t}, \underline{\delta}, \mathbf{Z}) = \prod_{j=1}^n \left[\frac{\exp \{ \underline{\beta}' \underline{\mathbf{Z}}_j(t_{(j)}) \}}{\sum_{\ell \in R_j} \exp \{ \underline{\beta}' \underline{\mathbf{Z}}_\ell(t_{(j)}) \}} \right]^{\delta_j}$$

Esta verosimilitud compara ahora, el riesgo de un sujeto con covariables $\underline{\mathbf{Z}}_j(t_{(j)})$ *al tiempo* $t_{(j)}$, contra el riesgo de los sujetos con covariables $\underline{\mathbf{Z}}_\ell(t_{(j)})$ *al tiempo* $t_{(j)}$ en el conjunto en riesgo $R_{(j)}$.

Ejemplo con una covariable dicotómica dependiente del tiempo

Supongamos que tenemos una variable categórica que denota el momento en que un individuo recibe el trasplante de un órgano (corazón, riñón, médula ósea, etc.). Entonces, podemos definir la variable

$$TR(t) = \begin{cases} 1, & \text{si } t \geq t_0 \\ 0, & \text{si } t < t_0 \end{cases}$$

con t_0 el momento en que el individuo recibe el trasplante. Esta variable se incorpora al modelo, que, obviamente, ya no cumple el supuesto de riesgos proporcionales.

Una variable “definida” dependiente del tiempo

Una variable cuyos valores no cambian en el transcurso del tiempo, pero que no cumple el supuesto de riesgos proporcionales, puede definirse dependiente del tiempo multiplicándola (i.e., generando una interacción) con el tiempo o alguna función de él. Por ejemplo, si la variable *SEXO*, no cumple el supuesto de riesgos proporcionales, se puede considerar la inclusión de $SEXO * g(t)$ dentro del modelo, para alguna función simple de t .

Un uso adicional de la metodología de covariables dependientes del tiempo, es probar el supuesto de riesgos proporcionales. Para probar si una covariable no dependiente del tiempo, Z_j , cumple o no este supuesto, podemos crear una variable dependiente del tiempo, $Z_j^*(t) = Z_j g(t)$, introducimos este término al modelo de Cox y probamos la hipótesis de que su parámetro asociado, digamos β_j^* , es cero

$$H_0 : \beta_j^* = 0 \quad \text{vs.} \quad H_a : \beta_j^* \neq 0$$

Si se rechaza la hipótesis nula, se tendrá evidencia de que Z_j es una variable dependiente del tiempo.

Cómo proponer $g(t)$?

Funciones comunes:

$$g(t) = \begin{cases} t \\ \log(t) \\ \exp(t) \\ \sqrt{t} \end{cases}$$

Modelo estratificado de riesgos proporcionales

Supongamos que al revisar el supuesto de riesgos proporcionales, una de nuestras covariables, por ejemplo sexo (0: hombre, 1: mujer), no cumple este supuesto, y el resto sí. Una estrategia para solventar este problema es *estratificar* nuestro análisis por la variable sexo. En este caso el modelo de riesgos proporcionales queda como

$$h_i(t|\underline{\mathbf{Z}}) = h_{0_i}(t) \exp \left\{ \underline{\beta}' \underline{\mathbf{Z}} \right\}, \quad i = 0, 1$$

donde $\underline{\mathbf{Z}}$ contiene a todas las covariables *menos sexo*, que es la variable de estratificación. El supuesto fundamental para utilizar el modelo de riesgos proporcionales estratificado, es que las covariables que sí cumplen el supuesto, tienen el mismo efecto en los estratos, en nuestro ejemplo, en hombre y mujeres. Esto implica que los $\underline{\hat{\beta}}$'s son iguales para los modelos de hombres y mujeres. Dado que la variable sexo no aparece en ninguno de los modelos generados por el proceso de estratificación, no es posible cuantificar el efecto de ella en el riesgo de los sujetos. Este proceso de estratificación se puede generalizar para el caso en que más de una covariable no cumple el supuesto de riesgos proporcionales. Si alguna de estas covariables es continua, es conveniente discretizarla para generar los estratos.

Modelo totalmente paramétrico

Si en el modelo de riesgos proporcionales se propone un modelo paramétrico para $h_0(t)$, entonces el modelo es totalmente paramétrico. El proceso de estimación es similar al usual, pero ahora se deben estimar los parámetros asociados a la función de riesgo básica.

Ejemplo

Supongamos que la función de riesgo básica es *Weibull*(α, λ), entonces el modelo asociado es

$$h(t|\underline{\mathbf{Z}}) = \lambda \alpha t^{\alpha-1} \exp \left\{ \underline{\beta}' \underline{\mathbf{Z}} \right\}$$

La manera natural de proponer la forma de $h_0(t)$, es realizando un análisis exploratorio con los elementos que pertenezcan a la población básica, $\underline{\mathbf{Z}} = \underline{\mathbf{0}}$. Las gráficas de una transformación del estimador Kaplan-Meier vs. alguna función del tiempo, pueden ser útiles para hacer

esta propuesta.

Tópicos no considerados

Algunos puntos interesantes, no considerados en este documento son

- Análisis con tiempos de supervivencia discreta
- Modelos con *frailty's* (Efectos aleatorios)
- Modelos multi-estado
- Modelos para riesgos competitivos
- Modelación Bayesiana
- Análisis a través del enfoque de procesos de conteo
- Etc., etc., etc., etc,...

FIN